BMJ Open

A quantitative investigation of inappropriate use of multivariate analysis and the importance of medical statistics experts in observational medical research: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021129
Article Type:	Research
Date Submitted by the Author:	14-Dec-2017
Complete List of Authors:	Nojima, Masanori; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Tokunaga, Mutsumi; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Nagamura, Fumitaka; the Institute of Medical Science, the University of Tokyo, Center for Translational Research
Keywords:	multivariate analysis, medical statistics, biostatistics, Epidemiology < TROPICAL MEDICINE, clinical research, observational research

A quantitative investigation of inappropriate use of multivariate analysis and the importance of medical statistics experts in observational medical research: a cross-sectional study

Masanori Nojima[1,2], Mutsumi Tokunaga[1,3], Fumitaka Nagamura[1,2]

1. Center for Translational Research, the Institute of Medical Science, the University of Tokyo

2. The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

3. Department of Health and Social Behavior, the University of Tokyo, School of Public Health

Corresponding Author

Masanori Nojima

Center for Translational Research / The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

TEL: +81-6409-2340

Email: nojima@ims.u-tokyo.ac.jp

Abstract

Objective: To investigate under what circumstances inappropriate use of multivariate analysis is likely to occur and to identify the population that needs more support with medical statistics.

Study Design and Settings: The frequency of the inappropriate use of multivariate analysis and related-factors were investigated in observational medical research publications.

Results: Using only variables that were significant in univariate analysis, an inappropriate algorithm was estimated to occur at 6.4% (95%CI: 4.8-8.5%). This was observed in 1.1% of the publications with a medical statistics expert (hereinafter "expert") as the first author, 3.5% if an expert was included as co-author, and in 12.2% if experts were not involved. In the publications where the number of cases was 50 or less and the study did not include experts, inappropriate algorithm usage was observed with a high proportion of 20.2%. The odds ratio of the involvement of experts for this outcome was 0.28 (95%CI: 0.15-0.53). The involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis (R = -0.652).

Conclusion: Based on the results of this study, the benefit of participation of medical statistics experts is obvious. Experts should be involved for proper confounding adjustment and interpretation of statistical models.

Keywords

multivariate analysis; medical statistics; biostatistics; epidemiology; clinical research; observational research

Strengths and limitations of this study

Strengths

- In studies where the number of events is small and medical statistics experts do not participate as co-authors, inappropriate multivariate analysis is often used, and sensitivity analysis by creating multiple models has not been conducted. Also in the country level investigation, the association between absence of experts and inappropriate multivariate analysis was remarkable. Even with various confounding factors adjustments, participation of experts was inversely correlated with inappropriate use of multivariate analysis.

- This is a unique research that quantitatively investigated the frequency and the

factors leading to inappropriate use of algorithms in variable selection of multivariate analysis. We also evaluated the quantitative efficacy of the involvement of medical statistics expert. As a result, the importance of experts' participation in medical research became clear.

- It is desirable to establish a statistical support system for researchers who have limited or no access to medical statistics experts.

Limitations

- Since the definition of outcome is complicated, there are many possibilities of misclassification. Therefore, the reliability may be higher in the examination of the relative difference rather than absolute values. In addition, the number of factors related to the quality of multivariate analysis are far more than those examined in this study.

- Even papers classified under the undesirable outcome this time may not always be inappropriate as multivariate analysis. For example, when the purpose of multivariate analysis is to construct a predictive model, there is no problem if a model with high predictive power is finally created. Our outcomes should be considered as potential inappropriate/desirable use of multivariate analysis.

1. Introduction

In the medical research field, "multivariate analysis" (some claim that it should be called "multivariable analysis"), typified by logistic regression or Cox regression, is widely used as a means of controlling confounding in observational research and creating a prognostic prediction model [1]. As statistical analysis software became widely used, multivariate analysis also became familiar to many medical researchers and clinicians. Although multivariate analysis is easily executed using software, understanding the statistical assumptions that constitute the premise of multivariate analysis and interpretation of the statistical model are very difficult for researchers who do not specialize in biostatistics. Consequently, it is concerning that multivariate analysis could become part of the "black box of statistics." Moreover, common misconceptions have been formed among medical researchers who are not specialized in statistics, which can interfere with correct understanding and interpretation of the results.

An American medical journal, "Annals of Internal Medicine"

(http://annals.org/aim/pages/AuthorInformationStatisticsOnly) describes its representative example as general statistical guidance on their website.

"Approaches that select factors for inclusion in a multivariable model only if the factors are 'statistically significant' in 'bivariate screening' are not optimal. A factor can be a confounder even if it is not statistically significant by itself because it changes the effect of the exposure of interest when it is included in the model, or because it is a confounder only when included with other covariates. … Better strategies than P value driven approaches for selecting variables are those that use external clinical judgment."

The problem with the algorithm in the first sentence of previous quotation has already been pointed out many times [1-3]. In Kenneth J. Rothman's "Epidemiology: An Introduction" [4], the author said, "The two primary ones (purpose) being to make predictions and to control for confounding." This algorithm ignores the true associated factor whose apparent association is weakened by confounding in univariate analysis, which is not reasonable for any purpose. However, although it is just personal experience as statistical consultant, we receive many questions like, "Only variables that were significant in univariate analysis are included in multivariate analysis, right?"

Knowing in what situations such inappropriate analysis is being done should lead to improvement in the quality of statistical analysis in medical research. However, there are no reports that summarize how multivariate analysis is carried out, including whether medical statistical experts are involved or not.

Based on the above situation, we decided to investigate under what circumstances inappropriate use is likely to occur and to identify the population that needs more support. Since inappropriate use of multivariate analysis (particularly in variable selection) is found even in published papers, we investigated its frequency and related factors in publications. Considering the feasibility, time constraints, and difficulty in the survey, we examined the following items as outcomes: 1) using only variables that were significant in univariate analysis, 2) using too many explanatory variables for few events. Additionally, as a desirable multivariate analysis method, we also investigated whether multiple models were created for the same outcome / factor relation as an outcome.

Many other things should be considered in multivariate analysis such as association of events with variables, premises on distribution of variables, and correlation between explanatory variables. Therefore, knowledge of both medical science and biostatistics is necessary to enable appropriate understanding of statistical models. We therefore assessed the association between medical statistics expert involvement (such as biostatistician and epidemiologist) and the outcomes. Based on this research, we found a high-risk population in the implementation of multivariate analysis and suggest improvement measures.

2. Materials and methods

2.1. Selection of applicable journals and publications

This study was conducted as a cross-sectional study. Here, target publications in this study are about medical research undertaking multivariate analysis. To target publications with various qualities and properties, a multistep sampling method was applied as described below. Briefly, we first selected scientific journals dealing with clinical medicine and epidemiology and then we sampled individual publications. Also, for "multivariate analysis," we chose logistic regression and Cox regression which are frequently performed in medical research. Details are as follows:

1) Journals were selected from the journals listed in Thomson Reuter's Journal Citation Report. We first selected 45 medical research fields including 609 journals from the list in the website in 2014 ("JCR year" was 2013). Selected research fields were listed in Supplementary Table 1.
2) With simple sampling, many journals with a small number of citations could be selected. Therefore, sampling was stratified by the impact factor which is an indicator directly reflecting citation frequency. The journals were classified into the following four layers according to the impact factor: "<2 (less than 2)," "2-<4 (two to less than 4)," "4-<6 (four to less than 6),"and "6< (more than 6)."
3) Subsequently, we selected journals whose number of articles exceeds 200 / year to avoid journals with few articles and extracted all journals with impact factor of 6 or more (71 journals). The sampling rates of other strata were set to extract the same number (71 × 4 = 284 journals, listed in Supplementary Table 2). Sampling rates according to impact factor were: over 6: 100%, 4-6: < 55.5%, 2-4: < 27.8%, and under 2: 45.8%. Journals selected for the investigation in this study were listed with this information in Supplementary Table 2.
4) We searched for publications in which logistic regression / Cox regression was performed from selected journal in PubMed (within the past 5 years: 2011-2015).

The search terms were "logistic + XXXX (journal name)" for logistic regression, and "hazard + XXXX (journal name)" for Cox regression, respectively. A publication database with 4086 (for logistic) and 11726 (for Cox) publications was constructed through the previously described process. Clinical trials were excluded when the word "random" or "trial" was included in the title or abstract. Meta-analysis was also excluded when the word "meta-analysis" was included in the title or abstract. All publications were from journals contracted with the University of Tokyo or open access articles.

5) To set the 95% confidence interval to the range of ± 3%, the target number of publications was 1200. To limit selection bias to choose journals with many publications with multivariate analysis, the sampling rate was calculated by applying a power function with an exponent < 1 to the number of publications (for logistic regression: $0.34*N^{0.644}$, for Cox regression: $0.54/N^{0.644}$, N: the number of publications in each journal).
6) Ineligible publications that could not be excluded by the above steps were excluded afterwards, and 571 papers (for logistic) and 541 (for Cox) were selected as the research subject. This number satisfies the target confidence interval set above.

2.2. Surveillance

The following information was collected from sampled publications by research assistants with knowledge of statistical analysis: affiliation of authors, country of the first author, method of variable selection for multivariate analysis (the primary outcome described below), number of the events (for multivariate analysis, categorized as: -20, 21-50, 51-100, and 101-), number of the covariates (categorized as: -2, 3-5, 6-10, 11-), etc. We decided whether authors or co-authors have expertise in biostatistics or epidemiology based on their affiliation. When the affiliation includes the following terms or related terms: epidemiology, public health, prevention, nutrition, social health, community health, occupational health, environmental health, population, global health, nutrition, biostatistics, statistics, mathematics, and clinical research, the author was considered a medical statistics expert (hereinafter, sometimes simply referred to as “expert”) in this research. Affiliation and the outcomes were independently collected by different assistants to avoid affecting determination of their association. For outcome-specific (not research-specific) information such as the number of events and the number of covariates, basically the information on the primary endpoint was collected, and if not applicable, information on the multivariate analysis first appearing in the abstracts or results was collected.

Since it was suggested that there are more problems in studies with few events (the number of events was 100 or less at the preliminary review), validation of the outcomes by the expert (the first author) was carefully done. In addition, the outcome of "Creating multiple models for the same outcome / factor relation" was surveyed by the first author. In this surveillance, for the studies where the number of events exceeds 100, because the number is extremely large, validation was carried out by 30% sampling.

2.3. Outcomes

All outcomes were defined as surrogates for the quality of multivariate analysis. These should be considered as inappropriate/desirable algorithms.

1. "Using only variables that were significant in univariate analysis" is the primary event for this study, which means that all variables screened with statistical significance in univariate analysis were automatically entered without manual selection of variables and without consideration for the relevance of variables. This includes cases when it is written as such in method section or it is obvious that it was implemented as such from expression of the tables. It is excluded from the event when variables were manually added or removed due to relevance to outcomes (such as a factor of interest or an established risk factor) or statistical consideration (such as multiple collinearity) after the screening in univariate analysis. However, it is not excluded when the stepwise method such as backward elimination method is only applied algorithmically for *post hoc* variable selection.
2. "Using too many explanatory variables for few events" is one of the secondary outcomes. This outcome was investigated only when the number of events for individual publication was equal to 50 or less and if the number of covariates was over 11 when the number of events was equal to 50 or less or the number of covariates was over 5 when the number of events was equal to 20 or less. The criteria was basically based on the study from Peduzzi et al. [5, 6], but because defining the exact number of events and covariates is sometimes very difficult, we relaxed that criterion; outcomes were taken only when the number of events is less than 50 and the number of covariates exceeds 20% of the number of events.
3. "Creating multiple models for the same outcome / factor relation" was determined as a desirable outcome for multivariate analysis. It was defined as the event only if tables were included for multiple models (because of screening efficiency). A representative example of this outcome was a fixed outcome and factors of interest related to various adjustment of covariates such as "adjustment for age," "age + sex,"

"age + sex + other important factors," etc. Subgroup analysis and analysis on different outcomes are not included in this outcome.

Of course, there are many other points to be considered in multivariate analysis, such as multiple collinearity and use of intermediate variables, but these were not included at this time because it is difficult to gather information from publications from various research areas.

2.4. Statistical analyses

Statistical analyses for binomial outcomes were performed using weighted generalized estimated equation (distribution = binomial, link = logit) with robust variance. Weight was basically defined as the inverse of the following formula: sampling rate stratified by impact factor * sampling rate based on the number of each journal (investigated / published). The correlation coefficient weighted by the number of publications was calculated using a general linear model. All statistical analyses were performed using SPSS 23 (IBM).

3. Results

3.1. Characteristics of investigated publications

The flow chart of the selection of the research subjects is summarized in Figure 1. An outline of the investigated publications is shown in Table 1 (total number was 1112). Most of the studies were large-scale research that exceeded 100 events. Publication whose first author is an expert in medical statistics is estimated to be 33.5% of the total, and in the remaining 67.7%, the proportion of publications in which an expert was included in co-authors was estimated to be 37.8%.

3.2. Descriptive statistics of the outcomes

Descriptive statistics of the outcomes are summarized in Table 2. The primary endpoint of our research, "Using only variables that were significant in univariate analysis" was estimated to occur in 6.4% (95%CI: 4.8-8.5%) of the overall publications. There was a big difference depending on whether an expert was the first author or not. It was observed in only 1.1% of the publications with the involvement of an expert as the first author, 12.2% if experts were not involved, and 3.5% if an expert was included as co-author. When an expert was included as the first author or co-author, it was 2.1%.

"Using too many explanatory variables for few events" was observed in 17.4% of the total, 19.0% if the first author is an expert, 22.1% if experts were not involved, and

11.5% if an expert was included as co-author. Since these are only for research with few events, the estimation accuracy was low. When an expert was included as the first author or co-author, it was 13.6%.

Regarding the preferred outcome, “Creating multiple models for the same outcome / factor relation,” like the primary outcome, the result greatly differed depending on whether the first author was an expert or not. If the first author is an expert, the preferred outcome was achieved 30.7% of the time. Otherwise, only 7.3% is achieved if the co-authorship did not contain experts, and 19.0% if an expert was included. In the case in which an expert was included as the first author or co-author, it was 26.2%. This outcome does not overlap with the algorithm "using only variables that are significant in univariate analysis" in which only one model was created basically. As can be seen from the above results, it was considered that when the authors included an expert, preferable analysis was carried out more frequently.

3.3. Subgroup analysis

Subsequently, the association between the number of events and the impact factor in each publication and the outcomes were assessed. As shown in Table 3, unfavorable results are observed in publications with fewer events and in journals with lower impact factors, independent from involvement of experts. In particular, where the number of cases was 50 or less and the study did not include experts, inappropriate multivariate analysis was observed with a high proportion of 20.2%. At the same time, construction of multiple models was implemented at a low proportion of 2.1%. When the impact factor is under 2 in studies in which experts were not involved, similar results have been observed (30.6% for the former, and 4.0% for the latter).

3.4. Further analysis for the association between involvement of experts in medical statistics and the quality of multivariate analysis

We assessed the association between the involvement of experts and the outcomes by adjusting for the two factors stratified above (Table 4). As a result, the odds ratio of the involvement of experts for "using only variables that are significant in univariate analysis" was 0.28 (95%CI: 0.15-0.53) which can be interpreted to be a large risk reduction.

If an expert was involved as the first author in the publication, the paper is expected to be an epidemiological study, and there should be an influence due to the difference in research characteristics on the result. If the first author is not an expert, the research could be a non-epidemiological research such as clinical research, and we focused on

how much improvement could be seen by involving an expert in these studies. As a result, even when an expert was involved only as a co-author, the risk decreased with an odds ratio of 0.42 (95%CI: 0.19-0.97). Likewise, for "Creating multiple models for the same outcome / factor relation," the result was favorable when an expert was included (OR 3.51. 95% CI: 1.88-6.58 for as any type of author, OR 2.36 for only as co-author, 95% CI: 1.03 - 5.38).

3.5. Nation-level investigation

Finally, we examined how much medical statistics experts are involved as co-authors when the first author is not an expert and its association with "using only variables that are significant in univariate analysis" for each country (of the first author).

First of all, 45% of all papers are reports from the United States, accounting for an overwhelming majority compared to other countries (Table 5). As shown in Figure 2, the correlation coefficients (weighting the number of publications) of “Proportion of publications with medical statistics experts as co-author within publications in which the first author is not an expert” with “proportion of publications with multivariate analysis using only variables that were significant in univariate analysis without manual selection of variables” showed an inverse correlation with R = -0.652. In this analysis, countries with more than 10 publications in which the first author is not an expert were used. North America and Northern Europe show relatively high expert involvement proportion, whereas East Asia has a low level of 20% or less except for Taiwan. For other European countries, there is variability in the result. The involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis. The details are summarized in Table 5.

4. Discussion

In this study, we focused on the algorithm called "use only variables that were significant in univariate analysis" as the inappropriate outcome which is often implemented mechanically without considering the influence of confounding and the relationship between variables. The result of 6.4% for this outcome was less than our expectation. However, considering that those who consult with us are "clinicians who conduct small-scale observational research (in Japan)," which was detected as a risk factor in this research, the research results are consistent with the expectation.

The reason why they adopt these methods seems to be based on the following idea.

- Regarding statistical significance as sacred: it has become a problem in recent years,

a statement concerning abuse of P value from American Statistical Association (ASA) was announced [7].

- Placing emphasis on being statistically "independent": some researchers think that it is totally meaningless unless the factor of interest is associated with their outcome independently of any existing variable.
- Thinking that not using significant variables in univariate analysis is considered arbitrary, and using non-significant variables in univariate analysis is also considered arbitrary.

Here, suppose adjuvant chemotherapy for a hypothetical cancer is performed frequently for cases with lymph node metastasis with strong association with recurrence. Although this adjuvant chemotherapy has the effect of preventing recurrence, univariate analysis shows weaker association than actual due to confounding by lymph node metastasis. However, with appropriate adjustment for lymph node metastasis, a significant inverse association was observed between the adjuvant chemotherapy with recurrence (example shown in Supplementary Table 3). If you apply an algorithm of using only variables that were significant in univariate analysis, the actual effect of adjuvant chemotherapy would be overlooked. Also, to investigate how confounding occurs in detail, it is necessary to create multiple models, and stratified analysis are very useful (Supplementary Table 4).

Variable selection is a critical problem in clinical studies with small sample size where it is unclear which factors should be adjusted. In such situations, variable selection dependent on P value in univariate analysis might be performed. Even though the number of covariates that can be entered at the same time is limited due to few events, a multifaceted approach such as creating multiple models should be helpful for causal interpretation. This is what we studied as a desirable outcome in this paper. For example, adjustments are made in multiple steps, such as crude (no adjustment) for model 1, age + sex for model 2, age + sex + another important factor A for model 3, and age + sex + another important factor B for model 4. Although this method should be recommended for studies with few events, there was a trend to omit this step in publications with fewer events (Table 3). Statistical multiplicity could be a problem with multiple models, however, we consider that it is not necessarily a severe problem because results from this approach are not independent and are highly correlated. Such sensitivity analysis with various statistical approaches is publicly recommended in clinical trials and analysis with missing data [8, 9].

Considering that multiple models are not created despite a small number of events and inappropriate analysis is often observed in a paper with a low impact factor, the reason why only significant variables are used is not caused only by the number of events, but by problems of the research system (including the absence of experts). In addition, the level of requirement from journals and the quality of peer review may be responsible.

Since medical and social influence from research is very large, and fair research performance is required, participation of biostatisticians is essential in clinical trials. However, ideally, experts should always participate in research even in observational studies because of the difficulty of appropriate adjustment for confounding including multivariate analysis. Even observational research can seriously affect clinical practice guidelines.

Based on the results of this study, the benefit of participation of medical statistics experts is obvious. Our results suggested that the proportion of experts' involvement is low in publications from East Asia, and there are relatively few publications in which the first author is an expert (Table 5). This would mean a shortage of such experts in these countries. The surveillance in 2011 by McKinsey Global Institute demonstrated that there are only small number of graduates with statistical training (including biostatistics) in Japan and China (2.66 and 1.31 graduates per 100 people in 2008, while 8.11, 13.58 and 12.47 for the United States, the United Kingdom, and France, respectively) [10]. The shortage of biostatisticians has been considered a problem in Japan, but infrastructure for training and developing biostatisticians has been developed rapidly in recent years [11].

However, it takes a long time to develop enough well-trained experts. In situations with a lack of medical statistics experts, it should be advisable to establish a system to disclose the data used for publication to enable the data to be analyzed (including multivariate analysis) by external experts as part of the peer review process. Here, "external" includes foreign experts or experts who are not acquainted personally with the research team. For new drug applications, researchers are obliged to submit the dataset of clinical trial standardized by the CDISC standard to regulatory authorities (Food and Drug Administration: FDA, Pharmaceuticals and Medical Devices Agency: PMDA, etc.) for further validation and additional analysis. Such standardization should be a model in constructing the system as described above.

Since clinicians performing clinical research are not necessarily full-time researchers and are usually very busy, they are the population that needs more support for medical statistics. In particular, those who are not involved in a huge research project (like a large epidemiological study) have difficulty accessing medical statistics experts. It is desirable to establish a support system for them within the peer review step regardless of the impact factor of the journal.

4.1. Limitations

1) Large-scale research was dominant in the study papers; the number of small-scale research in which there are possibly many problems was limited. Although it may have been sampled according to the number of events, it is difficult to extract that information by search words.
2) Since the definition of outcome is complicated, there are many possibilities of misclassification. Therefore, the reliability may be higher in the examination of the relative difference rather than absolute values.
3) The number of factors related to the quality of multivariate analysis are far more than those examined in this study.
4) Even papers classified under the undesirable outcome this time may not always be inappropriate as multivariate analysis. For example, when the purpose of multivariate analysis is to construct a predictive model, there is no problem if a model with high predictive power is finally created. Our outcomes should be considered as potential inappropriate/desirable use of multivariate analysis.

4.2. Conclusion

In publications about observational research in which the number of events is 50 or less without the involvement of medical statistics experts, more than 20% of publications may have problems in multivariate analysis. The involvement of experts was associated with desirable implementation of multivariate analysis independently of the number of events and the impact factor. The benefit of participation of medical statistics experts in the study is obvious. Since even observational research can be a source of important evidence in medical science, experts should be involved for proper confounding adjustment and interpretation of statistical models. We hope that this research will make medical researchers more conscious of the appropriate use of multivariate analysis.

Funding source

This study was supported by Grants-in-Aid for Scientific research (C), JSPS KAKENHI grant Number JP 26460764 (Fiscal-year 2014-16, Masanori Nojima).

Competing interest

There are no competing interests.

Author's contributions

MN: Conception and design of the study, writing the manuscript, analysis and interpretation of data. MT: Acquisition and interpretation of data and data, critically revision of the manuscript. FN: Supervising the overall research, and critically revision of the manuscript.

References

1. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG: Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1-12
2. Sun GW, Shook TL, Kay GL: Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907-16
3. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87
4. Rothman KJ: Epidemiology: An Introduction / Kenneth J. Rothman. - 2nd ed. Oxford 2012
5. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-9
6. Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503-10
7. Wasserstein RL, Lazar NA: The ASA's Statement on p-Values: Context, Process, and Purpose Am Stat 2016;70:129-33
8. International Conference on Harmonisation E9 Expert Working Group: ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. Stat Med 1999;18:1905–42
9. O'Neill RT, Temple R: The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol

Ther 2012;91:550-4

10. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH: Big data: The next frontier for innovation, competition, and productivity McKinsey Global Institute 2011
11. Carpenter J: How biostatistics is of rapidly growing importance in Japan Statistics Views 2013

Figure legends

Figure 1. Summary of the selection of publications investigated in this study.

Figure 2. A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are correlated inversely.

Table 1. Characteristics of publications investigated in this study.

			Number of publications (N = 1112)	%
The number of events	<21		47	4.2%
	21-50		122	11.0%
	51-100		96	8.6%
	100<		847	76.2%
Impact factor	Under 2		127	11.4%
	2-4<		160	14.4%
	4-6<		397	35.7%
	Over 6		428	38.5%
Medical statistics experts are included as	First author	Co-author		
	No	No	418	37.6%
	No	Yes	321	28.9%
	Yes	Either	373	33.5%

Table 2. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by whether medical statistics experts were included as author or not.

Events				95%CI	
			Proportion	Lower	Upper
1. Using only significant variables in univariate analysis					
			6.4%	4.8%	8.5%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	12.2%	8.7%	16.8%
	No	Yes	3.5%	2.0%	6.1%
	Yes	Either	1.1%	0.3%	3.5%
	1st author or co-author		2.1%	1.3%	3.6%
2. Using too many covariates for few events					
			17.4%	10.2%	28.0%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	22.1%	13.5%	33.9%
	No	Yes	11.5%	3.3%	33.1%
	Yes	Either	19.0%	3.8%	58.5%
	First author or co-author		13.6%	5.1%	31.5%
3. Constructing multiple multivariate models to assess the same outcome-factor association					
			14.4%	11.1%	18.3%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	7.3%	4.6%	11.4%
	No	Yes	19.0%	11.5%	29.7%
	Yes	Either	30.7%	23.0%	39.7%
	First author or co-author		26.2%	20.5%	32.9%

Table 3. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by the number of events, impact factor, and whether medical statistics experts were included as author or not.

Subgroup		Using only significant variables in univariate analysis			Constructing multiple multivariate models to assess the same outcome-factor association		
		Proportion	95%CI		Proportion	95%CI	
			Lower	Upper		Lower	Upper
Medical statistics experts included as first author or co-author	The number of events*						
No	<51	20.2%	12.5%	31.1%	2.1%	0.7%	5.9%
	51-100	9.4%	3.2%	24.7%	3.2%	1.1%	8.6%
	100<	8.6%	5.1%	14.2%	10.7%	6.3%	17.7%
Yes	<51	7.7%	2.9%	18.9%	12.6%	5.0%	28.2%
	51-100	4.0%	1.2%	13.0%	30.1%	16.5%	48.6%
	100<	1.6%	0.8%	3.2%	27.0%	20.6%	34.6%
Medical statistics experts included as first author or co-author	Impact factor						
No	Under 2	30.6%	17.1%	48.4%	4.0%	1.1%	13.7%
	2-4<	6.5%	2.4%	16.3%	3.4%	0.8%	13.1%
	4-6<	10.8%	5.8%	19.2%	11.7%	6.1%	21.5%
	Over 6	12.9%	7.5%	21.1%	9.0%	4.2%	18.4%
Yes	Under 2	6.0%	1.9%	17.2%	16.2%	5.4%	39.6%
	2-4<	3.1%	1.1%	8.6%	22.8%	10.5%	42.6%
	4-6<	0.2%	0.0%	1.1%	23.7%	16.1%	33.5%
	Over 6	3.5%	1.7%	6.9%	35.5%	25.9%	46.4%

*The category of "<21" has been integrated with the category "21 - 50" because of insufficient numbers

Table 4. Multivariate analysis for the assessment of the association between the absence of medical statistics experts and the use of inappropriate/desirable algorithms in multivariate analysis.

	Using only significant variables in univariate analysis			Constructing multiple multivariate models to assess the same outcome-factor association		
		95%CI			95%CI	
Factor	Odds ratio	Lower	Upper	Odds ratio	Lower	Upper
Medical statistics experts included as first author or co-author (vs. no experts)	0.28	0.15	0.53	3.51	1.88	6.58
Medical statistics experts included as first author or co-author (vs. no experts) when 1st author is clinicians or others	0.42	0.19	0.97	2.36	1.03	5.38

All models were adjusted for impact factor and the number of events.

Table 5. Summary of each country and proportion of publications in which medical statistics experts were included as co-author within the publications in which the first author is not an expert in these fields.

			Estimates		
			Publications in which the first author is NOT a medical statistics expert (%)	Medical experts are included as co-author within publications in which the first author is not an expert.	
Country	Total number of publications	Occupancy (%)		Proportion* (%)	95%CI*
USA	501	45.1	67.9	47.4	(40-54.9)
UK	63	5.7	48.2	22.0	(9.6-42.7)
China	51	4.6	84.5	6.7	(2.5-17.1)
Canada	48	4.3	67.4	50.7	(31.5-69.6)
Netherlands	46	4.1	73.1	37.4	(18.3-61.5)
Japan	45	4.0	81.2	15.3	(6.8-30.9)
South Korea	39	3.5	79.5	14.3	(4.9-35.1)
Sweden	38	3.4	40.0	45.3	(22.7-70)
Taiwan	29	2.6	91.3	38.8	(19.1-62.9)
Germany	27	2.4	80.1	41.7	(21.9-64.6)
Denmark	26	2.3	55.4	48.9	(23.9-74.5)
Italy	25	2.2	71.4	13.6	(4.1-36.3)
Australia	25	2.2	42.5	50.6	(16.4-84.3)
France	21	1.9	57.5	77.7	(46.5-93.3)
Spain	19	1.7	62.6	32.7	(11.8-63.8)
Brazil	13	1.2	51.1	4.6	(0.6-29.3)
Norway	11	1.0	48.4	44.8	(9.7-86)
Finland	8	0.7	85.8		
Switzerland	8	0.7	39.6		
Israel	7	0.6	60.9		
Singapore	6	0.5	92.8		
Belgium	6	0.5	64.8		
Turkey	5	0.4	100		
Austria	4	0.4	100		
South Africa	4	0.4	57.4		
Kenya	4	0.4	11.5		
Poland	3	0.3	100		
India	3	0.3	76.3		
Thailand	3	0.3	31.3		
Iran	3	0.3	34.2		
Greece	2	0.2	82.9		
Ireland	2	0.2	32.4		
Others	17	3.4	47.4		
Overall	1112	100	67.3	39.0	(32.2-45.4)

*Calculated only for countries with publications more than 10.

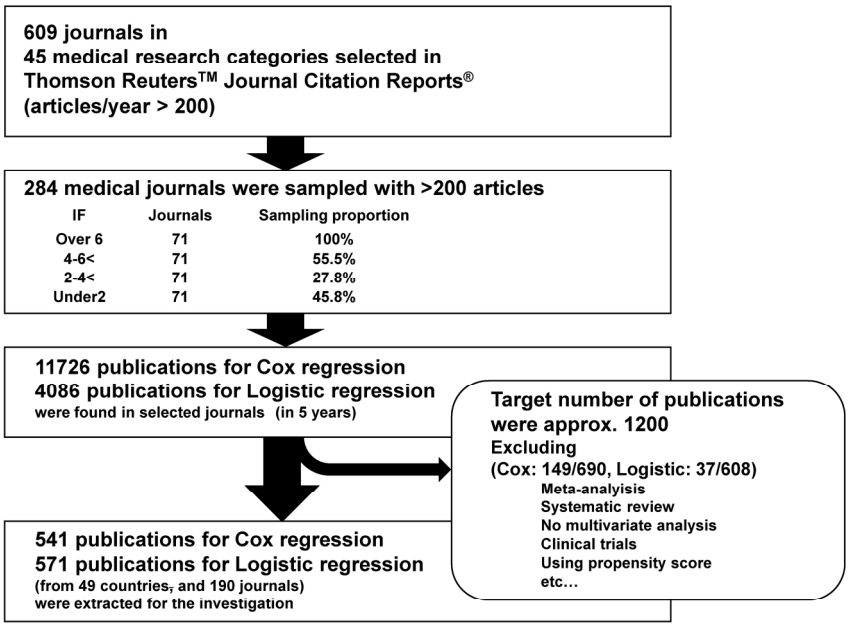

Figure 1.

Summary of the selection of publications investigated in this study.

190x142mm (300 x 300 DPI)

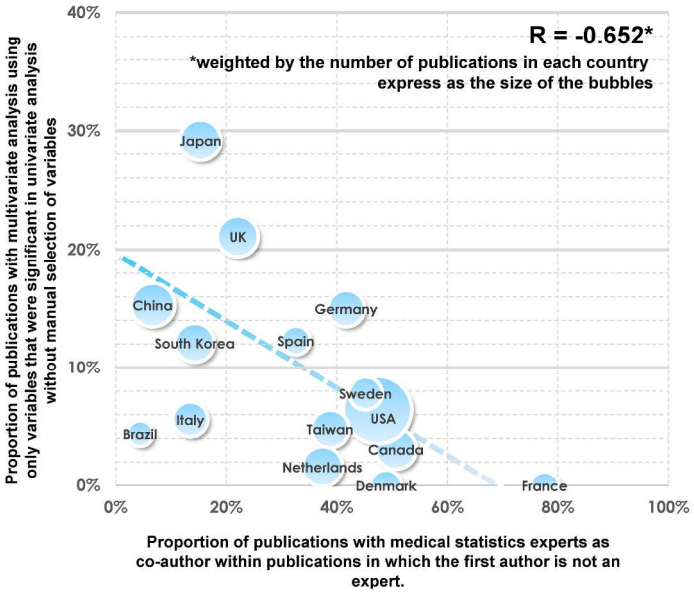

Figure 2.

A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are correlated inversely.

254x338mm (300 x 300 DPI)

Supplementary Table 1. Selected research filed in Thomson Reuter's Journal Citation Report (version 2014)

ALLERGY
ANESTHESIOLOGY
CARDIAC & CARDIOVASCULAR SYSTEMS
CLINICAL NEUROLOGY
CRITICAL CARE MEDICINE
DENTISTRY, ORAL SURGERY & MEDICINE
DERMATOLOGY
EMERGENCY MEDICINE
ENDOCRINOLOGY & METABOLISM
ENVIRONMENTAL SCIENCES
GASTROENTEROLOGY & HEPATOLOGY
GERIATRICS & GERONTOLOGY
HEALTH CARE SCIENCES & SERVICES
HEMATOLOGY
IMMUNOLOGY
INFECTIOUS DISEASES
INTEGRATIVE & COMPLEMENTARY MEDICINE
MEDICINE, GENERAL & INTERNAL
MEDICINE, RESEARCH & EXPERIMENTAL
NEUROSCIENCES
NURSING
NUTRITION & DIETETICS
OBSTETRICS & GYNECOLOGY
ONCOLOGY
OPHTHALMOLOGY
ORTHOPEDICS
OTORHINOLARYNGOLOGY
PATHOLOGY
PEDIATRICS
PERIPHERAL VASCULAR DISEASE
PHARMACOLOGY & PHARMACY
PSYCHIATRY
PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING
REHABILITATION
REPRODUCTIVE BIOLOGY
RESPIRATORY SYSTEM
RHEUMATOLOGY
SURGERY
TOXICOLOGY
TRANSPLANTATION
TROPICAL MEDICINE
UROLOGY & NEPHROLOGY
VIROLOGY
SUBSTANCE ABUSE

Supplementary Table 2. Journals selected for the investigation in this study.

2013 impact factor			
Over 6	4-<6	2-<4	Under 2
NEW ENGL J MED	ENVIRON MODELL SOFTW	TOXICON	TURK GOGUS KALP DAMA
LANCET	PEDIATRICS	J NEUROL SCI	RENAL FAILURE
JAMA-J AM MED ASSOC	PSYCHO-ONCOLOGY	AM J NEURORADIOL	ENVIRON MONIT ASSESS
J CLIN ONCOL	EXP NEUROL	PHYTOTHER RES	ZH NEVROL PSIKHIATR
BMJ-BRIT MED J	ALIMENT PHARM THER	INT J TUBERC LUNG D	ANIM REPROD SCI
NEURON	PLOS NEGLECT TROP D	J UROLOGY	NEUROL SCI
ENERG ENVIRON SCI	AM J OBSTET GYNECOL	AGR ECOSYST ENVIRON	J EMERG MED
J AM COLL CARDIOL	AM J PATHOL	EXP CELL RES	ENVIRON TOXICOL PHAR
NAT NEUROSCI	PAIN	DIABETES RES CLIN PR	BRAIN INJURY
CIRCULATION	INT J RADIAT ONCOL	OBES SURG	BMC PEDIATR
EUR HEART J	J AM MED INFORM ASSN	J VISION	AM J MED SCI
SCI TRANSL MED	THROMB HAEMOSTASIS	AM J INFECT CONTROL	WATER SCI TECHNOL
GASTROENTEROLOGY	J THROMB HAEMOST	ENVIRON TOXICOL CHEM	J STROKE CEREBROVASC
J EXP MED	ARTHRIT CARE RES	DRUG ALCOHOL DEPEN	CLINICS
J CLIN INVEST	EUR J CANCER	ECOL ECON	PROG UROL
AM J RESP CRIT CARE	AM J RESP CELL MOL	BMC NEUROL	ENVIRON SCI-PROC IMP
J ALLERGY CLIN IMMUN	PSYCHOL MED	VIRUS RES	J VIROL METHODS
HEPATOLOGY	BRIT J PHARMACOL	BIOL REPROD	BURNS
CIRC RES	AM J EPIDEMIOL	EUR J GASTROEN HEPAT	J NEUROSCI METH
J HEPATOL	RESUSCITATION	APPL CATAL A-GEN	J ORAL MAXIL SURG
NEUROSCI BIOBEHAV R	MOVEMENT DISORD	BREAST	PAK J MED SCI
BRAIN	BIOCHEM PHARMACOL	J NEURO-ONCOL	INT J ORAL MAX IMPL
BLOOD	NEUROBIOL AGING	SPINE J	ANN VASC SURG
BIOL PSYCHIAT	AM J KIDNEY DIS	EUR J PHARM SCI	KARDIOL POL
CLIN INFECT DIS	J TRANSL MED	TRANSPLANTATION	J CARDIOTHOR VASC AN
LEUKEMIA	GASTROINTEST ENDOSC	J PHARMACEUT BIOMED	CHINESE MED J-PEKING
CANCER RES	HAEMATOLOGICA	BMC PREGNANCY CHILDB	RHEUMATOL INT
ANN RHEUM DIS	RHEUMATOLOGY	AM J TROP MED HYG	B ENVIRON CONTAM TOX
DIABETES CARE	PROG NEURO-PSYCHOPH	J ENVIRON MANAGE	SUSTAINABILITY-BASEL
ONCOGENE	CLIN J AM SOC NEPHRO	TOXICOL IN VITRO	BONE JOINT J
KIDNEY INT	J AM COLL SURGEONS	MAGN RESON IMAGING	INT J CLIN EXP PATHO
DIABETES	J THORAC CARDIOV SUR	CORNEA	FOOT ANKLE INT
CEREB CORTEX	AM J SURG PATHOL	CHEMOSPHERE	EUR J OBSTET GYN R B
NEUROLOGY	REMOTE SENS ENVIRON	GEN COMP ENDOCR	ENVIRON MANAGE
GLOBAL CHANGE BIOL	J NUTR	CLIN ORAL IMPLAN RES	INT J GYNECOL CANCER
CLIN CANCER RES	OBESITY	BRIT J OPHTHALMOL	SURG TODAY
PLOS PATHOG	EUR RADIOL	TOXICOL APPL PHARM	ONCOL LETT
ARTHRITIS RHEUM-US	J AM ACAD DERMATOL	AM J CARDIOL	INTERNAL MED
NEUROPSYCHOPHARMACOL	INT J OBESITY	CLIN VACCINE IMMUNOL	J DRUGS DERMATOL
ANTIOXID REDOX SIGN	PHARM RES-DORDR	SLEEP MED	SKELETAL RADIOL
HYPERTENSION	J PHYSIOL-LONDON	CLIN EXP RHEUMATOL	PHARM BIOL
EMERG INFECT DIS	BIOL CONSERV	MOL VIS	PEDIATR EMERG CARE
BMC MED	ARTERIOSCL THROM VAS	J AM HEART ASSOC	PEDIATR CARDIOL
J CONTROL RELEASE	ENVIRON POLLUT	FOOD CHEM TOXICOL	EMERG MED J
ANN SURG	J NEUROCHEM	EUR J PHARMACOL	J CRANIOFAC SURG
STEM CELLS	ATHEROSCLEROSIS	ACTA TROP	AM J EMERG MED
CHEST	HUM REPROD	SPINE	ANTICANCER RES
EUR RESPIR J	AM HEART J	FRONT HUM NEUROSCI	ACTA NEUROCHIR
ENVIRON HEALTH PERSP	BREAST CANCER RES TR	MAGN RESON MED	PEDIATR RADIOL
HUM BRAIN MAPP	J CEREBR BLOOD F MET	NEUROSCIENCE	HEPATO-GASTROENTEROL
AM J CLIN NUTR	FERTIL STERIL	CURR MED CHEM	J CLIN NEUROSCI
DIABETOLOGIA	CAN J CARDIOL	J SEX MED	ACTA PAEDIATR
J NEUROSCI	RADIOTHER ONCOL	NUTRIENTS	INDIAN J SURG
J BONE MINER RES	J AM GERIATR SOC	NEPHROL DIAL TRANSPL	RESP PHYSIOL NEUROBI
ANN ONCOL	TOXICOL SCI	FRONT NEURAL CIRCUIT	DEUT MED WOCHENSCHR
AIDS	BONE	PRENATAL DIAG	J MATERN-FETAL NEO M
CLIN GASTROENTEROL H	LIVER INT	J GEN INTERN MED	INT J MED SCI
MOL THER	ENVIRON RES LETT	ARTHROSCOPY	INT J ENDOCRINOL
J INVEST DERMATOL	BRIT J ANAESTH	INT J ONCOL	OTOL NEUROTOL
J CLIN ENDOCR METAB	INFECT IMMUN	ENVIRON SCI POLLUT R	INT J PEDIATR OTORHI
RADIOLOGY	HEALTH AFFAIR	TRIALS	TERAPEVT ARKH
AM J TRANSPLANT	CANCER-AM CANCER SOC	INVEST OPHTH VIS SCI	ANZ J SURG
INT J CARDIOL	OSTEOPOROSIS INT	ARCH VIROL	J KOREAN MED SCI
OPHTHALMOLOGY	CANCER EPIDEM BIOMAR	AM J ROENTGENOL	OR SURG OR MED OR PA
ANESTHESIOLOGY	PSYCHOPHARMACOLOGY	UROL ONCOL-SEMIN ORI	J OBSTET GYNAECOL
CRIT CARE MED	ADDICTION	AM J PHYSIOL-GASTR L	IRAN J PUBLIC HEALTH
NEUROIMAGE	NEUROPHARMACOLOGY	QUAL LIFE RES	OTOLARYNG HEAD NECK
MOL CANCER THER	INT J CANCER	COLORECTAL DIS	J PAEDIATR CHILD H
CORTEX	J NUTR BIOCHEM	VIROL J	BMC COMPLEM ALTERN M
HEART	MOL CELL ENDOCRINOL	WASTE MANAGE	BRIT J ORAL MAX SURG
STROKE	MOL PHARMACOL	EUR J CLIN PHARMACOL	J ENVIRON SCI-CHINA

Supplementary Table 3. Example of multivariate analysis: logistic regression analysis for recurrence after surgery of hypothetical cancer with potential prognostic factors.

Univariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval	
			Lower	Upper
Adjuvant chemotherapy	0.101	0.45	0.17	1.17
Lymph node metastasis	<0.001	8.31	2.88	24.00
Biomarker positive	<0.001	17.11	5.38	54.39

Multivariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval		P value	Odds ratio	95% Confidence Interval	
			Lower	Upper			Lower	Upper
	Multivariate analysis 1				Multivariate analysis 2			
	Using only significant variables in univariate analysis				Using all potential prognostic factors			
Adjuvant chemotherapy	Not included				0.015	0.14	0.03	0.69
Lymph node metastasis	0.005	6.08	1.72	21.51	0.001	12.60	2.67	59.42
Biomarker positive	<0.001	13.77	3.99	47.48	<0.001	16.05	4.11	62.69
	Multivariate analysis 3				Multivariate analysis 4			
	Adjuvant chemotherapy + Lymph node metastasis				Adjuvant chemotherapy + Biomarker positive			
Adjuvant chemotherapy	0.013	0.18	0.05	0.70	0.093	0.35	0.10	1.19
Lymph node metastasis	<0.001	15.63	4.03	60.61	Not included			
Biomarker positive	Not included				<0.001	18.92	5.61	63.89

Inappropriate conclusion about adjuvant chemotherapy:
With multivariate analysis 1, adjuvant chemotherapy has no effect.

Desirable conclusion about adjuvant chemotherapy:
With multivariate analyses 2 to 4, adjuvant chemotherapy was inversely associated with recurrence after adjustment for lymph node metastasis.
Lymph node metastasis was a stronger confounder for the association between adjuvant chemotherapy and recurrence than the biomarker.

Supplementary Table 4. Cross-tabulation table for the association between adjuvant chemotherapy and recurrence stratified by lymph node metastasis for hypothetical cancer.

Lymph node metastasis		No recurrence		recurrence		Total
		Number	%	Number	%	Number
Absent	Without adjuvant chemotherapy	22	73.3%	8	26.7%	30
	With adjuvant chemotherapy	22	91.7%	2	8.3%	24
	Total	44	81.5%	10	18.5%	54
Present	Without adjuvant chemotherapy	1	10.0%	9	90.0%	10
	With adjuvant chemotherapy	8	50.0%	8	50.0%	16
	Total	9	34.6%	17	65.4%	26
Overall	Without adjuvant chemotherapy	23	57.5%	17	42.5%	40
	With adjuvant chemotherapy	30	75.0%	10	25.0%	40
	Total	53	66.3%	27	33.8%	80

Chi-square test for 2x2 table without stratification (Overall): P = 0.098

Odds ratio: 0.45 95% Confidence Interval 0.17-1.17

Mantel-Haenszel test for stratified analysis: P = 0.013

Common odds ratio: 0.19 95% Confidence Interval 0.05-0.71

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract A cross-sectional study
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found See Abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported See Introduction section
Objectives	3	State specific objectives, including any prespecified hypotheses See Abstract and Introduction
Methods		
Study design	4	Present key elements of study design early in the paper See Materials and methods section (Selection of applicable journals and publications)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection See Materials and methods section (Selection of applicable journals and publications)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants See Materials and methods section (Selection of applicable journals and publications)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable See Materials and methods section (Surveillance and Outcomes)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group See Materials and methods section (Surveillance and Outcomes)
Bias	9	Describe any efforts to address potential sources of bias See Materials and methods section (Surveillance, Outcomes and Statistical analyses)
Study size	10	Explain how the study size was arrived at 1112 (see Results)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why See Materials and methods section (Surveillance, Outcomes and Statistical analyses)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding See Materials and methods section (Statistical analyses) and Results section
		(*b*) Describe any methods used to examine subgroups and interactions See Materials and methods section (Statistical analyses) and Results section
		(*c*) Explain how missing data were addressed See Materials and methods section (Surveillance, Outcomes and Statistical analyses)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy See Materials and methods section (Selection of applicable journals and

		publications)
		(*e*) Describe any sensitivity analyses See Materials and methods section (Statistical analyses) and Results section
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed See Results section (Characteristics of investigated publications and Descriptive statistics of the outcomes) and Figure 1
		(b) Give reasons for non-participation at each stage See Figure 1
		(c) Consider use of a flow diagram See Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders See Results section (Characteristics of investigated publications and Descriptive statistics of the outcomes)
		(b) Indicate number of participants with missing data for each variable of interest See Figure 1
Outcome data	15*	Report numbers of outcome events or summary measures See Results section (Characteristics of investigated publications and Descriptive statistics of the outcomes)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included See Tables
		(*b*) Report category boundaries when continuous variables were categorized See Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses See Tables
Discussion		
Key results	18	Summarise key results with reference to study objectives See the 1st paragraph in Discussion section
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias See Discussion section
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence See Discussion section
Generalisability	21	Discuss the generalisability (external validity) of the study results See Discussion section
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if

	applicable, for the original study on which the present article is based
	See Funding source section

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021129.R1
Article Type:	Research
Date Submitted by the Author:	03-Apr-2018
Complete List of Authors:	Nojima, Masanori; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Tokunaga, Mutsumi; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Nagamura, Fumitaka; the Institute of Medical Science, the University of Tokyo, Center for Translational Research
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Epidemiology
Keywords:	multivariate analysis, regression analysis, biostatistics, clinical research, observational research, medical statistics expert

A quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study

Masanori Nojima[1,2], Mutsumi Tokunaga[1,3], and Fumitaka Nagamura[1,2]

1. Center for Translational Research, the Institute of Medical Science, the University of Tokyo

2. The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

3. Department of Health and Social Behavior, the University of Tokyo, School of Public Health

Corresponding Author

Masanori Nojima

Center for Translational Research / The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

TEL: +81-6409-2340

Email: nojima@ims.u-tokyo.ac.jp

Abstract

Objective: To investigate under what circumstances inappropriate use of "multivariate analysis" is likely to occur and to identify the population that needs more support with medical statistics.

Study Design and Settings: The frequency of inappropriate regression model construction in multivariate analysis and related-factors were investigated in observational medical research publications.

Results: The inappropriate algorithm of using only variables that were significant in univariate analysis was estimated to occur at 6.4% (95%CI: 4.8-8.5%). This was observed in 1.1% of the publications with a medical statistics expert (hereinafter "expert") as the first author, 3.5% if an expert was included as co-author, and in 12.2% if experts were not involved. In the publications where the number of cases was 50 or less and the study did not include experts, inappropriate algorithm usage was observed with a high proportion of 20.2%. The odds ratio of the involvement of experts for this outcome was 0.28 (95%CI: 0.15-0.53). A further, nation-level, analysis showed that the involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis (R = -0.652).

Conclusion: Based on the results of this study, the benefit of participation of medical statistics experts is obvious. Experts should be involved for proper confounding adjustment and interpretation of statistical models.

Keywords

multivariate analysis; regression analysis; biostatistics; clinical research; observational research; medical statistics expert;

Strengths and limitations of this study

Strengths

- This is a unique research that quantitatively investigated the frequency and the factors leading to inappropriate use of algorithms for variable selection in multivariate analysis.
- We also evaluated the quantitative efficacy of the involvement of medical statistics experts, and the importance of experts' participation in medical research became clear.
- The association between absence of experts and inappropriate multivariate analysis was remarkable in the nation-level investigation.

Limitations

- There are many possibilities for outcome misclassification due to complicated definition, and the number of factors related to the quality of multivariate analysis are far more than those examined in this study.

1. Introduction

In the medical research field, "multivariate analysis" (some claim that it should be called "multivariable analysis"; the usage of this term is discussed later), typified by logistic regression or Cox regression, is widely used as a means of controlling confounding in observational research and creating a prognostic prediction model [1]. As statistical analysis software became widely used, multivariate analysis also became familiar to many medical researchers and clinicians. Although multivariate analysis is easily executed using software, understanding the statistical assumptions that constitute the premise of multivariate analysis and interpretation of the statistical model are very difficult for researchers who do not specialize in biostatistics. Moreover, common misconceptions have been formed among medical researchers who are not specialized in statistics, which can interfere with correct understanding and interpretation of the results.

An American medical journal, "Annals of Internal Medicine" (http://annals.org/aim/pages/AuthorInformationStatisticsOnly) describes its representative example as general statistical guidance on their website.

"Approaches that select factors for inclusion in a multivariable model only if the factors are 'statistically significant' in 'bivariate screening' are not optimal. A factor can be a confounder even if it is not statistically significant by itself because it changes the effect of the exposure of interest when it is included in the model, or because it is a confounder only when included with other covariates. … Better strategies than P value driven approaches for selecting variables are those that use external clinical judgment."

The problem with the algorithm in the first sentence of previous quotation has already been pointed out many times [1-3]. In Kenneth J. Rothman's "Epidemiology: An Introduction" [4], the author said, "The two primary ones (purposes) being to make predictions and to control for confounding." This algorithm ignores the true associated factor whose apparent association is weakened by confounding in univariate analysis, which is not reasonable for any purpose. However, although it is just personal

experience as statistical consultants, we receive many questions like, "Only variables that were significant in univariate analysis are included in multivariate analysis, right?"

Knowing in what situations such inappropriate analysis is being done should lead to improvement in the quality of statistical analysis in medical research. However, there are no reports that summarize how multivariate analysis is carried out, including whether medical statistical experts are involved or not.

Based on the above situation, we decided to investigate under what circumstances inappropriate use is likely to occur and to identify the population that needs more support. Since inappropriate use of multivariate analysis (particularly in variable selection for regression model construction) is found even in published papers, we investigated its frequency and related factors in publications. Considering the feasibility, time constraints, and difficulty in the survey, we examined the following items as outcomes: 1) using only variables that were significant in univariate analysis, 2) using too many explanatory variables for few events. Additionally, as a desirable multivariate analysis method, we also investigated whether several models were fitted for the same outcome and selected factors.

Many other things should be considered in multivariate analysis such as association of events with variables, premises on distribution of variables, and correlation between explanatory variables. Therefore, knowledge of both medical science and biostatistics is necessary to enable appropriate understanding of statistical models. We therefore assessed the association between medical statistics expert involvement (such as biostatistician and epidemiologist) and the outcomes. Based on this research, we found a high-risk population in the implementation of multivariate analysis and suggest improvement measures.

2. Materials and methods

2.1. Selection of applicable journals and publications

This study was conducted as a cross-sectional study. Here, target publications in this study are about medical research undertaking multivariate analysis. To target publications with various qualities and properties, a multistep sampling method was applied as described below. Briefly, we first selected scientific journals dealing with clinical medicine and epidemiology and then we sampled individual publications. Also,

for "multivariate analysis," we chose logistic regression and Cox regression which are frequently performed in medical research. Details are as follows:

1) Journals were selected from the journals listed in Thomson Reuter's Journal Citation Report. We first selected 45 medical research fields including 609 journals from the list in the website in 2014 ("JCR year" was 2013). Selected research fields were listed in Supplementary Table 1.
2) With simple sampling, many journals with a small number of citations could be selected. Therefore, sampling was stratified by the impact factor which is an indicator directly reflecting citation frequency. The journals were classified into the following four layers according to the impact factor: "<2 (less than 2)," "2-<4 (two to less than 4)," "4-<6 (four to less than 6),"and "6< (more than 6)."
3) Subsequently, we selected journals whose number of articles exceeds 200 / year to avoid journals with few articles and extracted all journals with impact factor of 6 or more (71 journals). The sampling rates of other strata were set to extract the same number (71 × 4 = 284 journals, listed in Supplementary Table 2). Sampling rates according to impact factor were: over 6: 100%, 4-6: < 55.5%, 2-4: < 27.8%, and under 2: 45.8%. Journals selected for the investigation in this study are listed with this information in Supplementary Table 2.
4) We searched for publications in which logistic regression / Cox regression was performed from selected journals in PubMed (within the past 5 years: 2011-2015). The search terms were "logistic + XXXX (journal name)" for logistic regression, and "hazard + XXXX (journal name)" for Cox regression, respectively. A publication database with 4086 (for logistic) and 11726 (for Cox) publications was constructed through the previously described process. Clinical trials were excluded when the word "random" or "trial" was included in the title or abstract. Meta-analysis was also excluded when the word "meta-analysis" was included in the title or abstract. All publications were from journals available through the University of Tokyo or open access articles.
5) To set the 95% confidence interval to the range of ± 3%, the target number of publications was 1200. To limit selection bias from choosing journals with many publications with multivariate analysis, the sampling rate was calculated by applying a power function with an exponent < 1 to the number of publications (for logistic regression: $0.34*N^{0.644}$, for Cox regression: $0.54/N^{0.644}$, N: the number of publications in each journal).
6) Ineligible publications that could not be excluded by the above steps were excluded afterwards, and 571 papers (for logistic) and 541 (for Cox) were selected as the

research subject. This number satisfies the target confidence interval set above.

2.2. Surveillance

The following information was collected from sampled publications by research assistants with knowledge of statistical analysis: affiliation of authors, country of the first author, method of variable selection for multivariate analysis (the primary outcome described below), number of the events (for multivariate analysis, categorized as: -20, 21-50, 51-100, and 101-), number of the covariates (categorized as: -2, 3-5, 6-10, 11-), etc. We decided whether authors or co-authors have expertise in biostatistics or epidemiology based on their affiliation. When the affiliation includes the following terms or related terms: epidemiology, public health, prevention, nutrition, social health, community health, occupational health, environmental health, population, global health, nutrition, biostatistics, statistics, mathematics, and clinical research, the author was considered a medical statistics expert (hereinafter, sometimes simply referred to as "expert") in this research. Affiliation and the outcomes were independently collected by different assistants to avoid affecting determination of their association. For outcome-specific (not research-specific) information such as the number of events and the number of covariates, basically the information on the primary endpoint was collected, and if not applicable, information on the multivariate analysis first appearing in the abstracts or results was collected.

Since studies with few events (the number of events was 100 or less at the preliminary review) often included inappropriate analyses, the first author confirmed careful collection of information for such studies. In addition, the outcome of "Fitting several models for the same outcome and selected factors" was surveyed by the first author. In this surveillance, for the studies where the number of events exceeds 100, because the number is extremely large, validation was carried out by 30% sampling.

2.3. Outcomes

All outcomes were defined as surrogates for the quality of multivariate analysis. The following were considered as inappropriate/desirable algorithms.

1. "Using only variables that were significant in univariate analysis" is the primary outcome for this study, which means that all variables screened with statistical significance in univariate analyses were automatically entered without manual selection of variables and without consideration for the relevance of variables. This includes cases when it is written as such in the method section or it is obvious that it was implemented as such from expression of the tables. It is excluded from the

event when variables were manually added or removed due to relevance to outcomes (such as a factor of interest or an established risk factor) or statistical consideration (such as multiple collinearity) after the screening in univariate analysis. However, it is not excluded when the stepwise method such as backward elimination method is only applied algorithmically for *post hoc* variable selection.

2. "Using too many explanatory variables for few events" is one of the secondary outcomes. This outcome was investigated only when the number of events for individual publication was equal to 50 or less and if the number of covariates was over 11 when the number of events was equal to 50 or less or the number of covariates was over 5 when the number of events was equal to 20 or less. The criterion was basically based on the study from Peduzzi et al. [5, 6], but because defining the exact number of events and covariates is sometimes very difficult, we relaxed that criterion; outcomes were taken only when the number of events is less than 50 and the number of covariates exceeds 20% of the number of events.
3. "Fitting several models for the same outcome and selected factors" was determined as a desirable outcome for multivariate analysis. It was defined as the event only if tables were included for multiple models (because of screening efficiency). A representative example of this outcome was a fixed outcome and factors of interest related to various adjustment of covariates such as "adjustment for age," "age + sex," "age + sex + other important factors," etc. Subgroup analysis and analysis on different outcomes are not included in this outcome.

Of course, there are many other points to be considered in multivariate analysis, such as multiple collinearity and use of intermediate variables, but these were not included at this time because it is difficult to gather information from publications from various research areas.

2.4. Statistical analyses

Statistical analyses for binomial outcomes were performed using weighted generalized estimating equation (distribution = binomial, link = logit) with robust variance. Weight was basically defined as the inverse of the following formula: sampling rate stratified by impact factor * sampling rate based on the number of each journal (investigated / published). The correlation coefficient weighted by the number of publications was calculated using a general linear model. All statistical analyses were performed using SPSS 23 (IBM).

2.5. Patient and Public involvement

Neither were involved.

3. Results

3.1. Characteristics of investigated publications

The flow chart of the selection of the research subjects is summarized in Figure 1. An outline of the investigated publications is shown in Table 1 (total number was 1112). Most of the studies were large-scale research that exceeded 100 events. Publication whose first author is an expert in medical statistics is estimated to be 33.5% of the total, and in the remaining 67.7%, the proportion of publications in which an expert was included in co-authors was estimated to be 37.8%.

3.2. Descriptive statistics of the outcomes

Descriptive statistics of the outcomes are summarized in Table 2. The primary outcome of our research, “Using only variables that were significant in univariate analysis” was estimated to occur in 6.4% (95%CI: 4.8-8.5%) of the overall publications. There was a big difference depending on whether an expert was the first author or not. It was observed in only 1.1% of the publications with the involvement of an expert as the first author, 12.2% if experts were not involved, and 3.5% if an expert was included as co-author. When an expert was included as the first author or co-author, it was 2.1%.

“Using too many explanatory variables for few events” was observed in 17.4% of the total, 19.0% if the first author is an expert, 22.1% if experts were not involved, and 11.5% if an expert was included as co-author. Since these are only for research with few events, the estimation accuracy was low. When an expert was included as the first author or co-author, it was 13.6%.

Regarding the preferred outcome, “Fitting several models for the same outcome and selected factors,” like the primary outcome, the result greatly differed depending on whether the first author was an expert or not. If the first author is an expert, the preferred outcome was achieved 30.7% of the time. Otherwise, only 7.3% is achieved if the co-authorship did not contain experts, and 19.0% if an expert was included. In the case in which an expert was included as the first author or co-author, it was 26.2%. This outcome does not overlap with the algorithm "using only variables that are significant in univariate analysis" in which only one model was created for model selection. As can be seen from the above results, when the authors included an expert, preferable analysis was carried out more frequently.

3.3. Subgroup analysis

Subsequently, the association between the number of events and the impact factor in each publication and the outcomes were assessed. As shown in Table 3, unfavorable results are observed in publications with fewer events and in journals with lower impact factors, independently from involvement of experts. In particular, where the number of cases was 50 or less and the study did not include experts, inappropriate multivariate analysis was observed with a high proportion of 20.2%. At the same time, “fitting several models” was implemented at a low proportion of 2.1%. When the impact factor is under 2 in studies in which experts were not involved, similar results have been observed (30.6% for the former, and 4.0% for the latter).

3.4. Further analysis for the association between involvement of experts in medical statistics and the quality of multivariate analysis

We assessed the association between the involvement of experts and the outcomes by adjusting for the two factors stratified above (Table 4). As a result, the odds ratio of the involvement of experts for "using only variables that are significant in univariate analysis" was 0.28 (95%CI: 0.15-0.53) which can be interpreted to be a large risk reduction.

If an expert was involved as the first author in the publication, the paper is expected to be an epidemiological study, and there should be an influence due to the difference in research characteristics on the result. If the first author is not an expert, the research could be a non-epidemiological research such as clinical research, and we focused on how much improvement could be seen by involving an expert in these studies. As a result, even when an expert was involved only as a co-author, the risk decreased with an odds ratio of 0.42 (95%CI: 0.19-0.97). Likewise, for "Fitting several models for the same outcome and selected factors," the result was favorable when an expert was included (OR 3.51. 95% CI: 1.88-6.58 for as any type of author, OR 2.36 for only as co-author, 95% CI: 1.03 - 5.38).

3.5. Nation-level investigation

Finally, we examined how much medical statistics experts are involved as co-authors when the first author is not an expert and its association with "using only variables that are significant in univariate analysis" for each country (of the first author).

First of all, 45% of all papers are reports from the United States, accounting for an overwhelming majority compared to other countries (Table 5). As shown in Figure 2, the correlation coefficients (weighting the number of publications) of “Proportion of

publications with medical statistics experts as co-author within publications in which the first author is not an expert" with "proportion of publications with multivariate analysis using only variables that were significant in univariate analysis without manual selection of variables" showed an inverse correlation with R = -0.652. In this analysis, countries with more than 10 publications in which the first author is not an expert were used. North America and Northern Europe show relatively high expert involvement proportion, whereas East Asia has a low level of 20% or less except for Taiwan. For other European countries, there is variability in the result. The involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis. The details are summarized in Table 5.

4. Discussion

In this study, we focused on the algorithm called "use only variables that were significant in univariate analysis" as the inappropriate outcome which is often implemented mechanically without considering the influence of confounding and the relationship between variables. The result of 6.4% for this outcome was less than our expectation. However, considering that those who consult with us are "clinicians who conduct small-scale observational research (in Japan)," which was detected as a risk factor in this research, the research results are consistent with the expectation.

The reason why they adopt these methods seems to be based on the following ideas.

- Regarding statistical significance as sacred: this has become a problem in recent years, a statement concerning abuse of P value from American Statistical Association (ASA) was issued [7].
- Placing emphasis on being statistically "independent": some researchers think that inclusion of a factor is totally meaningless unless the factor of interest is associated with their outcome independently of any included variables.
- Thinking that not using significant variables in univariate analysis is considered arbitrary, and using non-significant variables in univariate analysis is also considered arbitrary.

Here, suppose adjuvant chemotherapy for a hypothetical cancer is performed frequently for cases with lymph node metastasis with strong association with recurrence. Although this adjuvant chemotherapy has the effect of preventing recurrence, univariate analysis shows weaker association than actual due to confounding by lymph node metastasis. However, with appropriate adjustment for lymph node metastasis, a significant inverse

association was observed between the adjuvant chemotherapy with recurrence (example shown in Supplementary Table 3). If you apply an algorithm of using only variables that were significant in univariate analysis, the actual effect of adjuvant chemotherapy would be overlooked. Also, to investigate how confounding occurs in detail, it is necessary to create multiple models, and stratified analyses are very useful (Supplementary Table 4).

Variable selection for regression model construction is a critical problem in clinical studies with small sample sizes where it is unclear which factors should be adjusted. In such situations, variable selection dependent on P value in univariate analysis might be performed. Even though the number of covariates that can be entered at the same time is limited due to few events, a multifaceted approach such as fitting several models should be helpful for causal interpretation. This is what we studied as a desirable outcome in this paper. For example, adjustments are made in multiple steps, such as crude (no adjustment) for model 1, age + sex for model 2, age + sex + another important factor A for model 3, and age + sex + another important factor B for model 4. However, this step was tended to be omitted in publications with fewer events (Table 3). Statistical multiplicity could be a problem with multiple models; however, we consider that it is not necessarily a severe problem because results from this approach are not independent and are highly correlated. Such sensitivity analysis with various statistical approaches is publicly recommended in clinical trials and analysis with missing data [8, 9].

Considering that multiple models are not created despite a small number of events and inappropriate analysis is often observed in a paper with a low impact factor, the reason why only significant variables are used is not caused only by the number of events, but by problems of the research system (including the absence of experts). In addition, the level of requirement from journals and the quality of peer review may be responsible.

Since medical and social influence from research is very large, and fair research performance is required, participation of biostatisticians is essential in clinical trials. However, ideally, experts should always participate in research even in observational studies because of the difficulty of appropriate adjustment for confounding including multivariate analysis. Even observational research can seriously affect clinical practice guidelines.

Based on the results of this study, the benefit of participation of medical statistics experts is obvious. Our results suggested that the proportion of experts' involvement is low in publications from East Asia, and there are relatively few publications in which the first author is an expert (Table 5). This would mean a shortage of such experts in these countries. The surveillance in 2011 by McKinsey Global Institute demonstrated that there are only a small number of graduates with statistical training (including biostatistics) in Japan and China (2.66 and 1.31 graduates per 100 people in 2008, while 8.11, 13.58 and 12.47 for the United States, the United Kingdom, and France, respectively) [10]. The shortage of biostatisticians has been considered a problem in Japan, but infrastructure for training and developing biostatisticians has been developed rapidly in recent years [11].

However, it takes a long time to develop enough well-trained experts. In situations with a lack of medical statistics experts, it should be advisable to establish a system to disclose the data used for publication to enable the data to be analyzed (including multivariate analysis) by external experts as part of the peer review process. Here, "external" includes foreign experts or experts who are not acquainted personally with the research team. For new drug applications, researchers are obliged to submit the dataset of clinical trial standardized by the CDISC standard to regulatory authorities (Food and Drug Administration: FDA, Pharmaceuticals and Medical Devices Agency: PMDA, etc.) for further validation and additional analysis. Such standardization should be a model in constructing the system as described above.

Since clinicians performing clinical research are not necessarily full-time researchers and are usually very busy, they are the population that needs more support for medical statistics. In particular, those who are not involved in a huge research project (like a large epidemiological study) have difficulty accessing medical statistics experts. It is desirable to establish a support system for them within the peer review step regardless of the impact factor of the journal.

4.1. Limitations

1) Large-scale research was dominant in the study papers; the number of small-scale research in which there are possibly many problems was limited. Although it may have been sampled according to the number of events, it is difficult to extract that information by search words.
2) Since the definition of outcome is complicated, there are many possibilities of misclassification. Therefore, the reliability may be higher in the examination of the

relative difference rather than absolute values.

3) The number of factors related to the quality of multivariate analysis are far more than those examined in this study.

4) Even papers we classify under the undesirable outcome may not necessarily use an inappropriate form of multivariate analysis. For example, when the purpose of multivariate analysis is to construct a predictive model, there is no problem if a model with high predictive power is finally created. Our three outcomes should then be considered as “potentially inappropriate” / “desirable” use of multivariate analysis.

4.2. The controversy about the term “multivariate/univariate”

The term "multivariable/univariable analysis" instead of "multivariate/univariate analysis" is sometimes recommended for regression analyses because "variate" means random variable [12]. However, in most situations described as “multivariate analysis”, medical researchers’ intentions are clear: adjust for multiple covariates as explanatory variables in regression models. We therefore adopted "multivariate/univariate analysis" in this study as this usage is more common in today's medical literature [12]. See Supplementary Discussion for further details.

4.3. Conclusion

In publications about observational research in which the number of events is 50 or less without the involvement of medical statistics experts, more than 20% of publications may have problems in multivariate analysis. The involvement of experts was associated with desirable implementation of multivariate analysis independently of the number of events and the impact factor. The benefit of participation of medical statistics experts in the study is obvious. Since even observational research can be a source of important evidence in medical science, experts should be involved for proper confounding adjustment and interpretation of statistical models. We hope that this research will make medical researchers more cognizant of appropriate regression model construction in multivariate analysis.

Funding source

This study was supported by Grants-in-Aid for Scientific research (C), JSPS KAKENHI grant Number JP 26460764 (Fiscal-year 2014-16, Masanori Nojima).

Competing interests

There are no competing interests.

Author's contributions

MN: Conception and design of the study, writing the manuscript, analysis and interpretation of data. MT: Acquisition and interpretation of data and critical revision of the manuscript. FN: Supervising the overall research and critical revision of the manuscript.

Acknowledgements

I would like to thank a research assistant, Ms. Kasumi Okazaki, for collecting publications and detailed information. I would also like to thank a biostatistician, Dr. Tomohiro Shinozaki, for giving advanced statistical advice.

Data sharing statement

No additional data are available.

References

1. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG: Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1-12
2. Sun GW, Shook TL, Kay GL: Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907-16
3. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87
4. Rothman KJ: Epidemiology: An Introduction / Kenneth J. Rothman. - 2nd ed. Oxford 2012
5. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-9
6. Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503-10
7. Wasserstein RL, Lazar NA: The ASA's Statement on p-Values: Context, Process, and Purpose Am Stat 2016;70:129-33
8. International Conference on Harmonisation E9 Expert Working Group: ICH

Harmonised Tripartite Guideline. Statistical principles for clinical trials. Stat Med 1999;18:1905-42

9. O'Neill RT, Temple R: The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther 2012;91:550-4
10. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH: Big data: The next frontier for innovation, competition, and productivity McKinsey Global Institute 2011
11. Carpenter J: How biostatistics is of rapidly growing importance in Japan Statistics Views 2013
12. Hidalgo B, Goodman M. Multivariate or multivariable regression? Am J Public Health. 2013;103:39-40

Figure legends

Figure 1. Summary of the selection of publications investigated in this study.

Figure 2. A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are inversely correlated.

Table 1. Characteristics of publications investigated in this study.

			Number of publications (N = 1112)	%
The number of events	<21		47	4.2%
	21-50		122	11.0%
	51-100		96	8.6%
	100<		847	76.2%
Impact factor	Under 2		127	11.4%
	2-4<		160	14.4%
	4-6<		397	35.7%
	Over 6		428	38.5%
Medical statistics experts are included as	First author	Co-author		
	No	No	418	37.6%
	No	Yes	321	28.9%
	Yes	Either	373	33.5%

Table 2. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by whether medical statistics experts were included as author or not.

Outcomes			Proportion	95%CI Lower	95%CI Upper
1. Using only significant variables in univariate analysis					
			6.4%	4.8%	8.5%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	12.2%	8.7%	16.8%
	No	Yes	3.5%	2.0%	6.1%
	Yes	Either	1.1%	0.3%	3.5%
	1st author or co-author		2.1%	1.3%	3.6%
2. Using too many covariates for few events					
			17.4%	10.2%	28.0%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	22.1%	13.5%	33.9%
	No	Yes	11.5%	3.3%	33.1%
	Yes	Either	19.0%	3.8%	58.5%
	First author or co-author		13.6%	5.1%	31.5%
3. Fitting several models for the same outcome and selected factors					
			14.4%	11.1%	18.3%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	7.3%	4.6%	11.4%
	No	Yes	19.0%	11.5%	29.7%
	Yes	Either	30.7%	23.0%	39.7%
	First author or co-author		26.2%	20.5%	32.9%

Table 3. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by the number of events, impact factor, and whether medical statistics experts were included as author or not.

		Using only significant variables in univariate analysis			Fitting several models for the same outcome and selected factors		
			95%CI			95%CI	
Subgroup		Proportion	Lower	Upper	Proportion	Lower	Upper
Medical statistics experts included as first author or co-author	The number of events*						
No	<51	20.2%	12.5%	31.1%	2.1%	0.7%	5.9%
	51-100	9.4%	3.2%	24.7%	3.2%	1.1%	8.6%
	100<	8.6%	5.1%	14.2%	10.7%	6.3%	17.7%
Yes	<51	7.7%	2.9%	18.9%	12.6%	5.0%	28.2%
	51-100	4.0%	1.2%	13.0%	30.1%	16.5%	48.6%
	100<	1.6%	0.8%	3.2%	27.0%	20.6%	34.6%
Medical statistics experts included as first author or co-author	Impact factor						
No	Under 2	30.6%	17.1%	48.4%	4.0%	1.1%	13.7%
	2-4<	6.5%	2.4%	16.3%	3.4%	0.8%	13.1%
	4-6<	10.8%	5.8%	19.2%	11.7%	6.1%	21.5%
	Over 6	12.9%	7.5%	21.1%	9.0%	4.2%	18.4%
Yes	Under 2	6.0%	1.9%	17.2%	16.2%	5.4%	39.6%
	2-4<	3.1%	1.1%	8.6%	22.8%	10.5%	42.6%
	4-6<	0.2%	0.0%	1.1%	23.7%	16.1%	33.5%
	Over 6	3.5%	1.7%	6.9%	35.5%	25.9%	46.4%

*The category of "<21" has been integrated with the category "21 - 50" because of insufficient numbers

Table 4. The assessment of the association between the absence of medical statistics experts and the use of inappropriate/desirable algorithms in multivariate analysis with adjustment for potential confounders.

	Using only significant variables in univariate analysis			Fitting several models for the same outcome and selected factors		
		95%CI			95%CI	
Factor	Odds ratio	Lower	Upper	Odds ratio	Lower	Upper
Medical statistics experts included as first author or co-author (vs. no experts)	0.28	0.15	0.53	3.51	1.88	6.58
Medical statistics experts included as first author or co-author (vs. no experts) when 1st author is clinicians or others	0.42	0.19	0.97	2.36	1.03	5.38

All models were adjusted for impact factor and the number of events.

Table 5. Summary of each country and proportion of publications in which medical statistics experts were included as co-author within the publications in which the first author is not an expert in these fields.

Country	Total number of publications	Occupancy (%)	Estimates		
			Publications in which the first author is NOT a medical statistics expert (%)	Medical experts are included as co-author within publications in which the first author is not an expert.	
				Proportion* (%)	95%CI*
USA	501	45.1	67.9	47.4	(40-54.9)
UK	63	5.7	48.2	22.0	(9.6-42.7)
China	51	4.6	84.5	6.7	(2.5-17.1)
Canada	48	4.3	67.4	50.7	(31.5-69.6)
Netherlands	46	4.1	73.1	37.4	(18.3-61.5)
Japan	45	4.0	81.2	15.3	(6.8-30.9)
South Korea	39	3.5	79.5	14.3	(4.9-35.1)
Sweden	38	3.4	40.0	45.3	(22.7-70)
Taiwan	29	2.6	91.3	38.8	(19.1-62.9)
Germany	27	2.4	80.1	41.7	(21.9-64.6)
Denmark	26	2.3	55.4	48.9	(23.9-74.5)
Italy	25	2.2	71.4	13.6	(4.1-36.3)
Australia	25	2.2	42.5	50.6	(16.4-84.3)
France	21	1.9	57.5	77.7	(46.5-93.3)
Spain	19	1.7	62.6	32.7	(11.8-63.8)
Brazil	13	1.2	51.1	4.6	(0.6-29.3)
Norway	11	1.0	48.4	44.8	(9.7-86)
Finland	8	0.7	85.8		
Switzerland	8	0.7	39.6		
Israel	7	0.6	60.9		
Singapore	6	0.5	92.8		
Belgium	6	0.5	64.8		
Turkey	5	0.4	100		
Austria	4	0.4	100		
South Africa	4	0.4	57.4		
Kenya	4	0.4	11.5		
Poland	3	0.3	100		
India	3	0.3	76.3		
Thailand	3	0.3	31.3		
Iran	3	0.3	34.2		
Greece	2	0.2	82.9		
Ireland	2	0.2	32.4		
Others	17	3.4	47.4		
Overall	1112	100	67.3	39.0	(32.2-45.4)

*Calculated only for countries with publications more than 10.

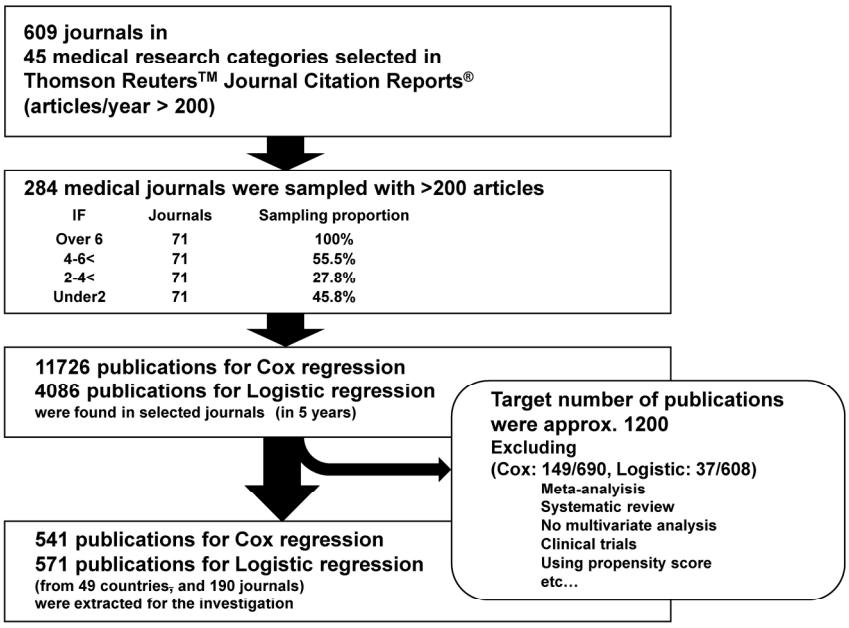

Figure 1.

Summary of the selection of publications investigated in this study.

190x142mm (300 x 300 DPI)

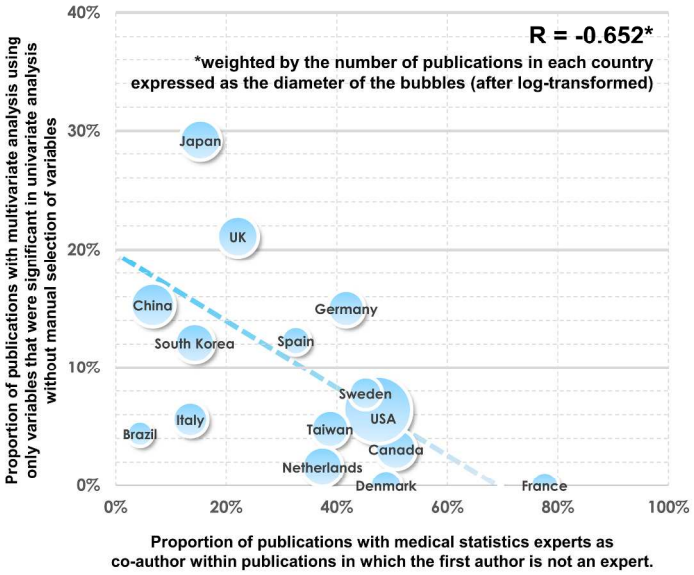

Figure 2.

A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are inversely correlated.

254x338mm (300 x 300 DPI)

Supplementary Table 1. Selected research filed in Thomson Reuter's Journal Citation Report (version 2014)

ALLERGY
ANESTHESIOLOGY
CARDIAC & CARDIOVASCULAR SYSTEMS
CLINICAL NEUROLOGY
CRITICAL CARE MEDICINE
DENTISTRY, ORAL SURGERY & MEDICINE
DERMATOLOGY
EMERGENCY MEDICINE
ENDOCRINOLOGY & METABOLISM
ENVIRONMENTAL SCIENCES
GASTROENTEROLOGY & HEPATOLOGY
GERIATRICS & GERONTOLOGY
HEALTH CARE SCIENCES & SERVICES
HEMATOLOGY
IMMUNOLOGY
INFECTIOUS DISEASES
INTEGRATIVE & COMPLEMENTARY MEDICINE
MEDICINE, GENERAL & INTERNAL
MEDICINE, RESEARCH & EXPERIMENTAL
NEUROSCIENCES
NURSING
NUTRITION & DIETETICS
OBSTETRICS & GYNECOLOGY
ONCOLOGY
OPHTHALMOLOGY
ORTHOPEDICS
OTORHINOLARYNGOLOGY
PATHOLOGY
PEDIATRICS
PERIPHERAL VASCULAR DISEASE
PHARMACOLOGY & PHARMACY
PSYCHIATRY
PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING
REHABILITATION
REPRODUCTIVE BIOLOGY
RESPIRATORY SYSTEM
RHEUMATOLOGY
SURGERY
TOXICOLOGY
TRANSPLANTATION
TROPICAL MEDICINE
UROLOGY & NEPHROLOGY
VIROLOGY
SUBSTANCE ABUSE

Supplementary Table 2. Journals selected for the investigation in this study.

2013 impact factor			
Over 6	4-<6	2-<4	Under 2
NEW ENGL J MED	ENVIRON MODELL SOFTW	TOXICON	TURK GOGUS KALP DAMA
LANCET	PEDIATRICS	J NEUROL SCI	RENAL FAILURE
JAMA-J AM MED ASSOC	PSYCHO-ONCOLOGY	AM J NEURORADIOL	ENVIRON MONIT ASSESS
J CLIN ONCOL	EXP NEUROL	PHYTOTHER RES	ZH NEVROL PSIKHIATR
BMJ-BRIT MED J	ALIMENT PHARM THER	INT J TUBERC LUNG D	ANIM REPROD SCI
NEURON	PLOS NEGLECT TROP D	J UROLOGY	NEUROL SCI
ENERG ENVIRON SCI	AM J OBSTET GYNECOL	AGR ECOSYST ENVIRON	J EMERG MED
J AM COLL CARDIOL	AM J PATHOL	EXP CELL RES	ENVIRON TOXICOL PHAR
NAT NEUROSCI	PAIN	DIABETES RES CLIN PR	BRAIN INJURY
CIRCULATION	INT J RADIAT ONCOL	OBES SURG	BMC PEDIATR
EUR HEART J	J AM MED INFORM ASSN	J VISION	AM J MED SCI
SCI TRANSL MED	THROMB HAEMOSTASIS	AM J INFECT CONTROL	WATER SCI TECHNOL
GASTROENTEROLOGY	J THROMB HAEMOST	ENVIRON TOXICOL CHEM	J STROKE CEREBROVASC
J EXP MED	ARTHRIT CARE RES	DRUG ALCOHOL DEPEN	CLINICS
J CLIN INVEST	EUR J CANCER	ECOL ECON	PROG UROL
AM J RESP CRIT CARE	AM J RESP CELL MOL	BMC NEUROL	ENVIRON SCI-PROC IMP
J ALLERGY CLIN IMMUN	PSYCHOL MED	VIRUS RES	J VIROL METHODS
HEPATOLOGY	BRIT J PHARMACOL	BIOL REPROD	BURNS
CIRC RES	AM J EPIDEMIOL	EUR J GASTROEN HEPAT	J NEUROSCI METH
J HEPATOL	RESUSCITATION	APPL CATAL A-GEN	J ORAL MAXIL SURG
NEUROSCI BIOBEHAV R	MOVEMENT DISORD	BREAST	PAK J MED SCI
BRAIN	BIOCHEM PHARMACOL	J NEURO-ONCOL	INT J ORAL MAX IMPL
BLOOD	NEUROBIOL AGING	SPINE J	ANN VASC SURG
BIOL PSYCHIAT	AM J KIDNEY DIS	EUR J PHARM SCI	KARDIOL POL
CLIN INFECT DIS	J TRANSL MED	TRANSPLANTATION	J CARDIOTHOR VASC AN
LEUKEMIA	GASTROINTEST ENDOSC	J PHARMACEUT BIOMED	CHINESE MED J-PEKING
CANCER RES	HAEMATOLOGICA	BMC PREGNANCY CHILDB	RHEUMATOL INT
ANN RHEUM DIS	RHEUMATOLOGY	AM J TROP MED HYG	B ENVIRON CONTAM TOX
DIABETES CARE	PROG NEURO-PSYCHOPH	J ENVIRON MANAGE	SUSTAINABILITY-BASEL
ONCOGENE	CLIN J AM SOC NEPHRO	TOXICOL IN VITRO	BONE JOINT J
KIDNEY INT	J AM COLL SURGEONS	MAGN RESON IMAGING	INT J CLIN EXP PATHO
DIABETES	J THORAC CARDIOV SUR	CORNEA	FOOT ANKLE INT
CEREB CORTEX	AM J SURG PATHOL	CHEMOSPHERE	EUR J OBSTET GYN R B
NEUROLOGY	REMOTE SENS ENVIRON	GEN COMP ENDOCR	ENVIRON MANAGE
GLOBAL CHANGE BIOL	J NUTR	CLIN ORAL IMPLAN RES	INT J GYNECOL CANCER
CLIN CANCER RES	OBESITY	BRIT J OPHTHALMOL	SURG TODAY
PLOS PATHOG	EUR RADIOL	TOXICOL APPL PHARM	ONCOL LETT
ARTHRITIS RHEUM-US	J AM ACAD DERMATOL	AM J CARDIOL	INTERNAL MED
NEUROPSYCHOPHARMACOL	INT J OBESITY	CLIN VACCINE IMMUNOL	J DRUGS DERMATOL
ANTIOXID REDOX SIGN	PHARM RES-DORDR	SLEEP MED	SKELETAL RADIOL
HYPERTENSION	J PHYSIOL-LONDON	CLIN EXP RHEUMATOL	PHARM BIOL
EMERG INFECT DIS	BIOL CONSERV	MOL VIS	PEDIATR EMERG CARE
BMC MED	ARTERIOSCL THROM VAS	J AM HEART ASSOC	PEDIATR CARDIOL
J CONTROL RELEASE	ENVIRON POLLUT	FOOD CHEM TOXICOL	EMERG MED J
ANN SURG	J NEUROCHEM	EUR J PHARMACOL	J CRANIOFAC SURG
STEM CELLS	ATHEROSCLEROSIS	ACTA TROP	AM J EMERG MED
CHEST	HUM REPROD	SPINE	ANTICANCER RES
EUR RESPIR J	AM HEART J	FRONT HUM NEUROSCI	ACTA NEUROCHIR
ENVIRON HEALTH PERSP	BREAST CANCER RES TR	MAGN RESON MED	PEDIATR RADIOL
HUM BRAIN MAPP	J CEREBR BLOOD F MET	NEUROSCIENCE	HEPATO-GASTROENTEROL
AM J CLIN NUTR	FERTIL STERIL	CURR MED CHEM	J CLIN NEUROSCI
DIABETOLOGIA	CAN J CARDIOL	J SEX MED	ACTA PAEDIATR
J NEUROSCI	RADIOTHER ONCOL	NUTRIENTS	INDIAN J SURG
J BONE MINER RES	J AM GERIATR SOC	NEPHROL DIAL TRANSPL	RESP PHYSIOL NEUROBI
ANN ONCOL	TOXICOL SCI	FRONT NEURAL CIRCUIT	DEUT MED WOCHENSCHR
AIDS	BONE	PRENATAL DIAG	J MATERN-FETAL NEO M
CLIN GASTROENTEROL H	LIVER INT	J GEN INTERN MED	INT J MED SCI
MOL THER	ENVIRON RES LETT	ARTHROSCOPY	INT J ENDOCRINOL
J INVEST DERMATOL	BRIT J ANAESTH	INT J ONCOL	OTOL NEUROTOL
J CLIN ENDOCR METAB	INFECT IMMUN	ENVIRON SCI POLLUT R	INT J PEDIATR OTORHI
RADIOLOGY	HEALTH AFFAIR	TRIALS	TERAPEVT ARKH
AM J TRANSPLANT	CANCER-AM CANCER SOC	INVEST OPHTH VIS SCI	ANZ J SURG
INT J CARDIOL	OSTEOPOROSIS INT	ARCH VIROL	J KOREAN MED SCI
OPHTHALMOLOGY	CANCER EPIDEM BIOMAR	AM J ROENTGENOL	OR SURG OR MED OR PA
ANESTHESIOLOGY	PSYCHOPHARMACOLOGY	UROL ONCOL-SEMIN ORI	J OBSTET GYNAECOL
CRIT CARE MED	ADDICTION	AM J PHYSIOL-GASTR L	IRAN J PUBLIC HEALTH
NEUROIMAGE	NEUROPHARMACOLOGY	QUAL LIFE RES	OTOLARYNG HEAD NECK
MOL CANCER THER	INT J CANCER	COLORECTAL DIS	J PAEDIATR CHILD H
CORTEX	J NUTR BIOCHEM	VIROL J	BMC COMPLEM ALTERN M
HEART	MOL CELL ENDOCRINOL	WASTE MANAGE	BRIT J ORAL MAX SURG
STROKE	MOL PHARMACOL	EUR J CLIN PHARMACOL	J ENVIRON SCI-CHINA

Supplementary Table 3. Example of multivariate analysis: logistic regression analysis for recurrence after surgery of hypothetical cancer with potential prognostic factors.

Univariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval	
			Lower	Upper
Adjuvant chemotherapy	0.101	0.45	0.17	1.17
Lymph node metastasis	<0.001	8.31	2.88	24.00
Biomarker positive	<0.001	17.11	5.38	54.39

Multivariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval		P value	Odds ratio	95% Confidence Interval	
			Lower	Upper			Lower	Upper
	Multivariate analysis 1				Multivariate analysis 2			
	Using only significant variables in univariate analysis				Using all potential prognostic factors			
Adjuvant chemotherapy	Not included				0.015	0.14	0.03	0.69
Lymph node metastasis	0.005	6.08	1.72	21.51	0.001	12.60	2.67	59.42
Biomarker positive	<0.001	13.77	3.99	47.48	<0.001	16.05	4.11	62.69
	Multivariate analysis 3				Multivariate analysis 4			
	Adjuvant chemotherapy + Lymph node metastasis				Adjuvant chemotherapy + Biomarker positive			
Adjuvant chemotherapy	0.013	0.18	0.05	0.70	0.093	0.35	0.10	1.19
Lymph node metastasis	<0.001	15.63	4.03	60.61	Not included			
Biomarker positive	Not included				<0.001	18.92	5.61	63.89

Inappropriate conclusion about adjuvant chemotherapy:
With multivariate analysis 1, adjuvant chemotherapy has no effect.

Desirable conclusion about adjuvant chemotherapy:
With multivariate analyses 2 to 4, adjuvant chemotherapy was inversely associated with recurrence after adjustment for lymph node metastasis.
Lymph node metastasis was a stronger confounder for the association between adjuvant chemotherapy and recurrence than the biomarker.

Supplementary Table 4. Cross-tabulation table for the association between adjuvant chemotherapy and recurrence stratified by lymph node metastasis for hypothetical cancer.

Lymph node metastasis		No recurrence		recurrence		Total
		Number	%	Number	%	Number
Absent	Without adjuvant chemotherapy	22	73.3%	8	26.7%	30
	With adjuvant chemotherapy	22	91.7%	2	8.3%	24
	Total	44	81.5%	10	18.5%	54
Present	Without adjuvant chemotherapy	1	10.0%	9	90.0%	10
	With adjuvant chemotherapy	8	50.0%	8	50.0%	16
	Total	9	34.6%	17	65.4%	26
Overall	Without adjuvant chemotherapy	23	57.5%	17	42.5%	40
	With adjuvant chemotherapy	30	75.0%	10	25.0%	40
	Total	53	66.3%	27	33.8%	80

Chi-square test for 2x2 table without stratification (Overall): P = 0.098
Odds ratio: 0.45 95% Confidence Interval 0.17-1.17

Mantel-Haenszel test for stratified analysis: P = 0.013
Common odds ratio: 0.19 95% Confidence Interval 0.05-0.71

Supplementary Discussion

The controversy about the term "multivariate/univariate"

The term "multivariable/univariable analysis" instead of "multivariate/univariate analysis" is sometimes recommended for regression analyses by several authors and guidelines because "variate" means random variable in statistics terminology [12]. If we literally follow the definition, "multivariate analysis" may only cover non-regression type analyses for multiple random variables (e.g., principal component analysis and factor analysis) or regression analyses with multiple outcome variables (e.g., multivariate analysis of variance). However, in most situations described as "multivariate analysis", medical researchers' intentions are clear: adjust for multiple covariates as explanatory variables in regression models. In fact, we usually model the conditional expectation E(Y|X) by regression analysis in observational studies where the joint distribution (X, Y) is not controlled by researchers. We thus believe that "multivariate adjustment" or "multivariate analysis" is not necessarily misuse of the terminology. We therefore adopted "multivariate/univariate analysis" in this study as this usage is more common in today's medical literature [12].

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract p.1: "a cross-sectional study"
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found p.2: See the abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported pp.3-4: See the 1st to 5th paragraphs in the introduction section
Objectives	3	State specific objectives, including any prespecified hypotheses p.1 and p.4: See the abstract and the 6th and last paragraphs in the introduction section
Methods		
Study design	4	Present key elements of study design early in the paper pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable pp.6-7: See the materials and methods section (2.2 Surveillance and 2.3. Outcomes)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group pp.6-7: See the materials and methods section (2.2 Surveillance and 2.3. Outcomes)
Bias	9	Describe any efforts to address potential sources of bias pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4. Statistical analyses)
Study size	10	Explain how the study size was arrived at p.8: See the results section (3.1. Characteristics of investigated publications)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4 Statistical analyses)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding pp.7-9: See Materials and methods section (2.4. Statistical analyses) and Results section (3.3. Subgroup analysis and 3.4. Further analysis for...)
		(*b*) Describe any methods used to examine subgroups and interactions See Materials and methods section (Statistical analyses) and Results section

		(*c*) Explain how missing data were addressed pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4 Statistical analyses)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
		(*e*) Describe any sensitivity analyses pp.7-10: See Materials and methods section (2.4. Statistical analyses) and Results section (3.3. Subgroup analysis, 3.4. Further analysis for... and 3.5. Nation-level investigation)
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed p.8 and p.21 (figure): See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes) and Figure 1
		(b) Give reasons for non-participation at each stage p. 21: See Figure 1
		(c) Consider use of a flow diagram p.21: See Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders p.8: See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes)
		(b) Indicate number of participants with missing data for each variable of interest p.21: See Figure 1
Outcome data	15*	Report numbers of outcome events or summary measures p.8: See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included pp.16-20: See Tables
		(*b*) Report category boundaries when continuous variables were categorized pp.16-20: See Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses pp.16-20: See Tables
Discussion		
Key results	18	Summarise key results with reference to study objectives p.10: See the 1st paragraph in the discussion section
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias p.12: See the discussion section (4.1. Limitations)

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p.13: See the discussion section (4.3. Conclusion)
Generalisability	21	Discuss the generalisability (external validity) of the study results pp.10-13: See the whole discussion section (but in particular, intensively described in the 6th and 7th paragraphs, 4.1. Limitations and 4.3. Conclusion)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p.13: See Funding source section

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021129.R2
Article Type:	Research
Date Submitted by the Author:	10-Apr-2018
Complete List of Authors:	Nojima, Masanori; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Tokunaga, Mutsumi; the Institute of Medical Science, the University of Tokyo, Center for Translational Research Nagamura, Fumitaka; the Institute of Medical Science, the University of Tokyo, Center for Translational Research
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Epidemiology
Keywords:	multivariate analysis, regression analysis, biostatistics, clinical research, observational research, medical statistics expert

A quantitative investigation of inappropriate regression model construction and the importance of medical statistics experts in observational medical research: a cross-sectional study

Masanori Nojima[1,2], Mutsumi Tokunaga[1,3], and Fumitaka Nagamura[1,2]

1. Center for Translational Research, the Institute of Medical Science, the University of Tokyo

2. The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

3. Department of Health and Social Behavior, the University of Tokyo, School of Public Health

Corresponding Author

Masanori Nojima

Center for Translational Research / The Division of Advanced Medicine Promotion, the Institute of Medical Science, the University of Tokyo

TEL: +81-6409-2340

Email: nojima@ims.u-tokyo.ac.jp

Abstract

Objective: To investigate under what circumstances inappropriate use of "multivariate analysis" is likely to occur and to identify the population that needs more support with medical statistics.

Study Design and Settings: The frequency of inappropriate regression model construction in multivariate analysis and related-factors were investigated in observational medical research publications.

Results: The inappropriate algorithm of using only variables that were significant in univariate analysis was estimated to occur at 6.4% (95%CI: 4.8-8.5%). This was observed in 1.1% of the publications with a medical statistics expert (hereinafter "expert") as the first author, 3.5% if an expert was included as co-author, and in 12.2% if experts were not involved. In the publications where the number of cases was 50 or less and the study did not include experts, inappropriate algorithm usage was observed with a high proportion of 20.2%. The odds ratio of the involvement of experts for this outcome was 0.28 (95%CI: 0.15-0.53). A further, nation-level, analysis showed that the involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis (R = -0.652).

Conclusion: Based on the results of this study, the benefit of participation of medical statistics experts is obvious. Experts should be involved for proper confounding adjustment and interpretation of statistical models.

Keywords

multivariate analysis; regression analysis; biostatistics; clinical research; observational research; medical statistics expert;

Strengths and limitations of this study

Strengths

- This is a unique research quantitatively investigating the frequency and the factors leading to inappropriate use of algorithms for variable selection in multivariate analysis.
- We also evaluated the quantitative efficacy of the involvement of medical statistics experts, and the importance of experts' participation in medical research became clear.
- The association between absence of experts and inappropriate multivariate analysis was remarkable in the nation-level investigation.

Limitations

- There are many possibilities for outcome misclassification due to complicated definitions, and the number of factors related to the quality of multivariate analysis are far more than those examined in this study.

1. Introduction

In the medical research field, "multivariate analysis" (some claim that it should be called "multivariable analysis"; the usage of this term is discussed later), typified by logistic regression or Cox regression, is widely used as a means of controlling confounding in observational research and creating a prognostic prediction model [1]. As statistical analysis software became widely used, multivariate analysis also became familiar to many medical researchers and clinicians. Although multivariate analysis is easily executed using software, understanding the statistical assumptions that constitute the premise of multivariate analysis and interpretation of the statistical model are very difficult for researchers who do not specialize in biostatistics. Moreover, common misconceptions have been formed among medical researchers who are not specialized in statistics, which can interfere with correct understanding and interpretation of the results.

An American medical journal, "Annals of Internal Medicine" (http://annals.org/aim/pages/AuthorInformationStatisticsOnly) describes its representative example as general statistical guidance on their website.

"Approaches that select factors for inclusion in a multivariable model only if the factors are 'statistically significant' in 'bivariate screening' are not optimal. A factor can be a confounder even if it is not statistically significant by itself because it changes the effect of the exposure of interest when it is included in the model, or because it is a confounder only when included with other covariates. … Better strategies than P value driven approaches for selecting variables are those that use external clinical judgment."

The problem with the algorithm in the first sentence of the previous quotation has already been pointed out many times [1-3]. In Kenneth J. Rothman's "Epidemiology: An Introduction" [4], the author said, "The two primary ones (purposes) being to make predictions and to control for confounding." This algorithm ignores the true associated factor whose apparent association is weakened by confounding in univariate analysis, which is not reasonable for any purpose. However, although it is just personal

experience as statistical consultants, we receive many questions like, "Only variables that were significant in univariate analysis are included in multivariate analysis, right?"

Knowing in what situations such inappropriate analysis is being done should lead to improvement in the quality of statistical analysis in medical research. However, there are no reports that summarize how multivariate analysis is carried out, including whether medical statistical experts are involved or not.

Based on the above situation, we decided to investigate under what circumstances inappropriate use is likely to occur and to identify the population that needs more support. Since inappropriate use of multivariate analysis (particularly in variable selection for regression model construction) is found even in published papers, we investigated its frequency and related factors in publications. Considering the feasibility, time constraints, and difficulty in the survey, we examined the following items as outcomes: 1) using only variables that were significant in univariate analysis, 2) using too many explanatory variables for few events. Additionally, as a desirable multivariate analysis method, we also investigated whether several models were fitted for the same outcome and sets of selected factors.

Many other things should be considered in multivariate analysis such as association of events with variables, premises on distribution of variables, and correlation between explanatory variables. Therefore, knowledge of both medical science and biostatistics is necessary to enable appropriate understanding of statistical models. We therefore assessed the association between medical statistics expert involvement (such as biostatistician and epidemiologist) and the outcomes. Based on this research, we found a high-risk population in the implementation of multivariate analysis and suggest improvement measures.

2. Materials and methods

2.1. Selection of applicable journals and publications

This study was conducted as a cross-sectional study. Here, target publications in this study are about medical research undertaking multivariate analysis. To target publications with various qualities and properties, a multistep sampling method was applied as described below. Briefly, we first selected scientific journals dealing with clinical medicine and epidemiology and then we sampled individual publications. Also,

for "multivariate analysis," we chose logistic regression and Cox regression which are frequently performed in medical research. Details are as follows:

1) Journals were selected from the journals listed in Thomson Reuter's Journal Citation Report. We first selected 45 medical research fields including 609 journals from the list in the website in 2014 ("JCR year" was 2013). Selected research fields were listed in Supplementary Table 1.
2) With simple sampling, many journals with a small number of citations could be selected. Therefore, sampling was stratified by the impact factor which is an indicator directly reflecting citation frequency. The journals were classified into the following four layers according to the impact factor: "<2 (less than 2)," "2-<4 (two to less than 4)," "4-<6 (four to less than 6),"and "6< (more than 6)."
3) Subsequently, we selected journals whose number of articles exceeds 200 / year to avoid journals with few articles and extracted all journals with impact factor of 6 or more (71 journals). The sampling rates of other strata were set to extract the same number (71 × 4 = 284 journals, listed in Supplementary Table 2). Sampling rates according to impact factor were: over 6: 100%, 4-6: < 55.5%, 2-4: < 27.8%, and under 2: 45.8%. Journals selected for the investigation in this study are listed with this information in Supplementary Table 2.
4) We searched for publications in which logistic regression / Cox regression was performed from selected journals in PubMed (within the past 5 years: 2011-2015). The search terms were "logistic + XXXX (journal name)" for logistic regression, and "hazard + XXXX (journal name)" for Cox regression, respectively. A publication database with 4086 (for logistic) and 11726 (for Cox) publications was constructed through the previously described process. Clinical trials were excluded when the word "random" or "trial" was included in the title or abstract. Meta-analysis was also excluded when the word "meta-analysis" was included in the title or abstract. All publications were from journals available through the University of Tokyo or open access articles.
5) To set the 95% confidence interval to the range of ± 3%, the target number of publications was 1200. To limit selection bias from choosing journals with many publications with multivariate analysis, the sampling rate was calculated by applying a power function with an exponent < 1 to the number of publications (for logistic regression: $0.34*N^{0.644}$, for Cox regression: $0.54/N^{0.644}$, N: the number of publications in each journal).
6) Ineligible publications that could not be excluded by the above steps were excluded afterwards, and 571 papers (for logistic) and 541 (for Cox) were selected as the

research subject. This number satisfies the target confidence interval set above.

2.2. Surveillance

The following information was collected from sampled publications by research assistants with knowledge of statistical analysis: affiliation of authors, country of the first author, method of variable selection for multivariate analysis (the primary outcome described below), number of the events (for multivariate analysis, categorized as: -20, 21-50, 51-100, and 101-), number of the covariates (categorized as: -2, 3-5, 6-10, 11-), etc. We decided whether authors or co-authors have expertise in biostatistics or epidemiology based on their affiliation. When the affiliation includes the following terms or related terms: epidemiology, public health, prevention, nutrition, social health, community health, occupational health, environmental health, population, global health, nutrition, biostatistics, statistics, mathematics, and clinical research, the author was considered a medical statistics expert (hereinafter, sometimes simply referred to as "expert") in this research. Affiliation and the outcomes were independently collected by different assistants to avoid affecting determination of their association. For outcome-specific (not research-specific) information such as the number of events and the number of covariates, basically the information on the primary endpoint was collected, and if not applicable, information on the multivariate analysis first appearing in the abstracts or results was collected.

Since studies with few events (the number of events was 100 or less at the preliminary review) often included inappropriate analyses, the first author confirmed careful collection of information for such studies. In addition, the outcome of "Fitting several models for the same outcome and selected factors" was surveyed by the first author. In this surveillance, for the studies where the number of events exceeds 100, because the number is extremely large, validation was carried out by 30% sampling.

2.3. Outcomes

All outcomes were defined as surrogates for the quality of multivariate analysis. The following were considered as inappropriate/desirable algorithms.

1. "Using only variables that were significant in univariate analysis" is the primary outcome for this study, which means that all variables screened with statistical significance in univariate analyses were automatically entered without manual selection of variables and without consideration for the relevance of variables. This includes cases when it is written as such in the method section or it is obvious that it was implemented as such from expression of the tables. It is excluded from the

event when variables were manually added or removed due to relevance to outcomes (such as a factor of interest or an established risk factor) or statistical consideration (such as multiple collinearity) after the screening in univariate analysis. However, it is not excluded when the stepwise method such as backward elimination method is only applied algorithmically for *post hoc* variable selection.

2. "Using too many explanatory variables for few events" is one of the secondary outcomes. This outcome was investigated only when the number of events for individual publication was equal to 50 or less and if the number of covariates was over 11 when the number of events was equal to 50 or less or the number of covariates was over 5 when the number of events was equal to 20 or less. The criterion was basically based on the study from Peduzzi et al. [5, 6], but because defining the exact number of events and covariates is sometimes very difficult, we relaxed that criterion; outcomes were taken only when the number of events is less than 50 and the number of covariates exceeds 20% of the number of events.
3. "Fitting several models for the same outcome and selected factors" was determined as a desirable outcome for multivariate analysis. It was defined as the event only if tables were included for multiple models (because of screening efficiency). A representative example of this outcome was a fixed outcome and factors of interest related to various adjustment of covariates such as "adjustment for age," "age + sex," "age + sex + other important factors," etc. Subgroup analysis and analysis on different outcomes are not included in this outcome.

Of course, there are many other points to be considered in multivariate analysis, such as multiple collinearity and use of intermediate variables, but these were not included at this time because it is difficult to gather information from publications from various research areas.

2.4. Statistical analyses

Statistical analyses for binomial outcomes were performed using weighted generalized estimating equations (distribution = binomial, link = logit) with robust variance. Weight was basically defined as the inverse of the following formula: sampling rate stratified by impact factor * sampling rate based on the number of each journal (investigated / published). The correlation coefficient weighted by the number of publications was calculated using a general linear model. All statistical analyses were performed using SPSS 23 (IBM).

2.5. Patient and Public involvement

Neither were involved.

3. Results

3.1. Characteristics of investigated publications

The flow chart of the selection of the research subjects is summarized in Figure 1. An outline of the investigated publications is shown in Table 1 (total number was 1112). Most of the studies were large-scale research that exceeded 100 events. Publication whose first author is an expert in medical statistics is estimated to be 33.5% of the total, and in the remaining 67.7%, the proportion of publications in which an expert was included in co-authors was estimated to be 37.8%.

3.2. Descriptive statistics of the outcomes

Descriptive statistics of the outcomes are summarized in Table 2. The primary outcome of our research, “Using only variables that were significant in univariate analysis” was estimated to occur in 6.4% (95%CI: 4.8-8.5%) of the overall publications. There was a big difference depending on whether an expert was the first author or not. It was observed in only 1.1% of the publications with the involvement of an expert as the first author, 12.2% if experts were not involved, and 3.5% if an expert was included as co-author. When an expert was included as the first author or co-author, it was 2.1%.

“Using too many explanatory variables for few events” was observed in 17.4% of the total, 19.0% if the first author is an expert, 22.1% if experts were not involved, and 11.5% if an expert was included as co-author. Since these are only for research with few events, the estimation accuracy was low. When an expert was included as the first author or co-author, it was 13.6%.

Regarding the preferred outcome, “Fitting several models for the same outcome and selected factors,” like the primary outcome, the result greatly differed depending on whether the first author was an expert or not. If the first author is an expert, the preferred outcome was achieved 30.7% of the time. Otherwise, only 7.3% is achieved if the co-authorship did not contain experts, and 19.0% if an expert was included. In the case in which an expert was included as the first author or co-author, it was 26.2%. This outcome does not overlap with the algorithm "using only variables that are significant in univariate analysis" in which only one model was created for model selection. As can be seen from the above results, when the authors included an expert, preferable analysis was carried out more frequently.

3.3. Subgroup analysis

Subsequently, the association between the number of events and the impact factor in each publication and the outcomes were assessed. As shown in Table 3, unfavorable results are observed in publications with fewer events and in journals with lower impact factors, independently from involvement of experts. In particular, where the number of cases was 50 or less and the study did not include experts, inappropriate multivariate analysis was observed with a high proportion of 20.2%. At the same time, “fitting several models” was implemented at a low proportion of 2.1%. When the impact factor is under 2 in studies in which experts were not involved, similar results have been observed (30.6% for the former, and 4.0% for the latter).

3.4. Further analysis for the association between involvement of experts in medical statistics and the quality of multivariate analysis

We assessed the association between the involvement of experts and the outcomes by adjusting for the two factors stratified above (Table 4). As a result, the odds ratio of the involvement of experts for "using only variables that are significant in univariate analysis" was 0.28 (95%CI: 0.15-0.53) which can be interpreted to be a large risk reduction.

If an expert was involved as the first author in the publication, the paper is expected to be an epidemiological study, and there should be an influence due to the difference in research characteristics on the result. If the first author is not an expert, the research could be a non-epidemiological research such as clinical research, and we focused on how much improvement could be seen by involving an expert in these studies. As a result, even when an expert was involved only as a co-author, the risk decreased with an odds ratio of 0.42 (95%CI: 0.19-0.97). Likewise, for "Fitting several models for the same outcome and selected factors," the result was favorable when an expert was included (OR 3.51. 95% CI: 1.88-6.58 for as any type of author, OR 2.36 for only as co-author, 95% CI: 1.03 - 5.38).

3.5. Nation-level investigation

Finally, we examined how much medical statistics experts are involved as co-authors when the first author is not an expert and its association with "using only variables that are significant in univariate analysis" for each country (of the first author).

First of all, 45% of all papers are reports from the United States, accounting for an overwhelming majority compared to other countries (Table 5). As shown in Figure 2, the correlation coefficients (weighting the number of publications) of “Proportion of

publications with medical statistics experts as co-author within publications in which the first author is not an expert" with "proportion of publications with multivariate analysis using only variables that were significant in univariate analysis without manual selection of variables" showed an inverse correlation with R = -0.652. In this analysis, countries with more than 10 publications in which the first author is not an expert were used. North America and Northern Europe show relatively high expert involvement proportion, whereas East Asia has a low level of 20% or less except for Taiwan. For other European countries, there is variability in the result. The involvement of experts and the implementation of unfavorable multivariate analysis are associated at the nation-level analysis. The details are summarized in Table 5.

4. Discussion

In this study, we focused on the algorithm called "use only variables that were significant in univariate analysis" as the inappropriate outcome which is often implemented mechanically without considering the influence of confounding and the relationship between variables. The result of 6.4% for this outcome was less than our expectation. However, considering that those who consult with us are "clinicians who conduct small-scale observational research (in Japan)," which was detected as a risk factor in this research, the research results are consistent with the expectation.

The reason why they adopt these methods seems to be based on the following ideas.

- Regarding statistical significance as sacred: this has become a problem in recent years, a statement concerning abuse of P values from the American Statistical Association (ASA) was issued [7] in 2016.
- Placing emphasis on being statistically "independent": some researchers think that inclusion of a factor is totally meaningless unless the factor of interest is associated with their outcome independently of any included variables.
- Thinking that not using significant variables in univariate analysis is considered arbitrary, and using non-significant variables in univariate analysis is also considered arbitrary.

Here, suppose adjuvant chemotherapy for a hypothetical cancer is performed frequently for cases with lymph node metastasis with strong association with recurrence. Although this adjuvant chemotherapy has the effect of preventing recurrence, univariate analysis shows weaker association than actual due to confounding by lymph node metastasis. However, with appropriate adjustment for lymph node metastasis, a significant inverse

association was observed between the adjuvant chemotherapy with recurrence (example shown in Supplementary Table 3). If you apply an algorithm of using only variables that were significant in univariate analysis, the actual effect of adjuvant chemotherapy would be overlooked. Also, to investigate how confounding occurs in detail, it is necessary to create multiple models, and stratified analyses are very useful (Supplementary Table 4).

Variable selection for regression model construction is a critical problem in clinical studies with small sample sizes where it is unclear which factors should be adjusted. In such situations, variable selection dependent on P value in univariate analysis might be performed. Even though the number of covariates that can be entered at the same time is limited due to few events, a multifaceted approach such as fitting several models should be helpful for causal interpretation. This is what we studied as a desirable outcome in this paper. For example, adjustments are made in multiple steps, such as crude (no adjustment) for model 1, age + sex for model 2, age + sex + another important factor A for model 3, and age + sex + another important factor B for model 4. However, this step tended to be omitted in publications with fewer events (Table 3). Statistical multiplicity could be a problem with multiple models; however, we consider that it is not necessarily a severe problem because results from this approach are not independent and are highly correlated. Such sensitivity analysis with various statistical approaches is publicly recommended in clinical trials and analysis with missing data [8, 9].

Considering that multiple models are not created despite a small number of events and inappropriate analysis is often observed in a paper with a low impact factor, the reason why only significant variables are used is not caused only by the number of events, but by problems of the research system (including the absence of experts). In addition, the level of requirement from journals and the quality of peer review may be responsible.

Since medical and social influence from research is very large, and fair research performance is required, participation of biostatisticians is essential in clinical trials. However, ideally, experts should always participate in research even in observational studies because of the difficulty of appropriate adjustment for confounding including multivariate analysis. Even observational research can seriously affect clinical practice guidelines.

Based on the results of this study, the benefit of participation of medical statistics

experts is obvious. Our results suggested that the proportion of experts' involvement is low in publications from East Asia, and there are relatively few publications in which the first author is an expert (Table 5). This would mean a shortage of such experts in these countries. The surveillance in 2011 by McKinsey Global Institute demonstrated that there are only a small number of graduates with statistical training (including biostatistics) in Japan and China (2.66 and 1.31 graduates per 100 people in 2008, while 8.11, 13.58 and 12.47 for the United States, the United Kingdom, and France, respectively) [10]. The shortage of biostatisticians has been considered a problem in Japan, but infrastructure for training and developing biostatisticians has been developed rapidly in recent years [11].

However, it takes a long time to develop enough well-trained experts. In situations with a lack of medical statistics experts, it should be advisable to establish a system to disclose the data used for publication to enable the data to be analyzed (including multivariate analysis) by external experts as part of the peer review process. Here, "external" includes foreign experts or experts who are not acquainted personally with the research team. For new drug applications, researchers are obliged to submit the dataset of clinical trial standardized by the CDISC standard to regulatory authorities (Food and Drug Administration: FDA, Pharmaceuticals and Medical Devices Agency: PMDA, etc.) for further validation and additional analysis. Such standardization should be a model in constructing the system as described above.

Since clinicians performing clinical research are not necessarily full-time researchers and are usually very busy, they are the population that needs more support for medical statistics. In particular, those who are not involved in a huge research project (like a large epidemiological study) have difficulty accessing medical statistics experts. It is desirable to establish a support system for them within the peer review step regardless of the impact factor of the journal.

4.1. Limitations

1) Large-scale research was dominant in the study papers; the number of small-scale research in which there are possibly many problems was limited. Although it may have been sampled according to the number of events, it is difficult to extract that information by search words.
2) Since the definition of outcome is complicated, there are many possibilities of misclassification. Therefore, the reliability may be higher in the examination of the relative difference rather than absolute values.

3) The number of factors related to the quality of multivariate analysis are far more than those examined in this study.
4) Even papers we classify under the undesirable outcome may not necessarily use an inappropriate form of multivariate analysis. For example, when the purpose of multivariate analysis is to construct a predictive model, there is no problem if a model with high predictive power is finally created. Our three outcomes should then be considered as “potentially inappropriate” / “desirable” use of multivariate analysis.

4.2. The controversy about the term “multivariate/univariate”

The term "multivariable/univariable analysis" instead of "multivariate/univariate analysis" is sometimes recommended for regression analyses because "variate" means random variable [12]. However, in most situations described as “multivariate analysis”, medical researchers’ intentions are clear: adjust for multiple covariates as explanatory variables in regression models. We therefore adopted "multivariate/univariate analysis" in this study as this usage is more common in today's medical literature [12]. See the Supplementary Discussion for further details.

4.3. Conclusion

In publications about observational research in which the number of events is 50 or less without the involvement of medical statistics experts, more than 20% of publications may have problems in multivariate analysis. The involvement of experts was associated with desirable implementation of multivariate analysis independently of the number of events and the impact factor. The benefit of participation of medical statistics experts in the study is obvious. Since even observational research can be a source of important evidence in medical science, experts should be involved for proper confounding adjustment and interpretation of statistical models. We hope that this research will make medical researchers more cognizant of appropriate regression model construction in multivariate analysis.

Funding source

This study was supported by Grants-in-Aid for Scientific research (C), JSPS KAKENHI grant Number JP 26460764 (Fiscal-year 2014-16, Masanori Nojima).

Competing interests

There are no competing interests.

Author's contributions

MN: Conception and design of the study, writing the manuscript, analysis and interpretation of data. MT: Acquisition and interpretation of data and critical revision of the manuscript. FN: Supervising the overall research and critical revision of the manuscript.

Acknowledgements

We would like to thank a research assistant, Ms. Kasumi Okazaki, for collecting publications and detailed information. We would also like to thank a biostatistician, Dr. Tomohiro Shinozaki, for giving advanced statistical advice.

Data sharing statement

No additional data are available.

References

1. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG: Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1-12
2. Sun GW, Shook TL, Kay GL: Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol 1996;49:907-16
3. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87
4. Rothman KJ: Epidemiology: An Introduction / Kenneth J. Rothman. - 2nd ed. Oxford 2012
5. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR: A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-9
6. Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503-10
7. Wasserstein RL, Lazar NA: The ASA's Statement on p-Values: Context, Process, and Purpose Am Stat 2016;70:129-33
8. International Conference on Harmonisation E9 Expert Working Group: ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. Stat Med

1999;18:1905-42

9. O'Neill RT, Temple R: The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther 2012;91:550-4
10. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH: Big data: The next frontier for innovation, competition, and productivity McKinsey Global Institute 2011
11. Carpenter J: How biostatistics is of rapidly growing importance in Japan Statistics Views 2013
12. Hidalgo B, Goodman M. Multivariate or multivariable regression? Am J Public Health. 2013;103:39-40

Figure legends

Figure 1. Summary of the selection of publications investigated in this study.

Figure 2. A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are inversely correlated.

Table 1. Characteristics of publications investigated in this study.

			Number of publications (N = 1112)	%
The number of events	<21		47	4.2%
	21-50		122	11.0%
	51-100		96	8.6%
	100<		847	76.2%
Impact factor	Under 2		127	11.4%
	2-4<		160	14.4%
	4-6<		397	35.7%
	Over 6		428	38.5%
Medical statistics experts are included as	First author	Co-author		
	No	No	418	37.6%
	No	Yes	321	28.9%
	Yes	Either	373	33.5%

Table 2. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by whether medical statistics experts were included as author or not.

Outcomes			Proportion	95%CI Lower	95%CI Upper
1. Using only significant variables in univariate analysis					
			6.4%	4.8%	8.5%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	12.2%	8.7%	16.8%
	No	Yes	3.5%	2.0%	6.1%
	Yes	Either	1.1%	0.3%	3.5%
	1st author or co-author		2.1%	1.3%	3.6%
2. Using too many covariates for few events					
			17.4%	10.2%	28.0%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	22.1%	13.5%	33.9%
	No	Yes	11.5%	3.3%	33.1%
	Yes	Either	19.0%	3.8%	58.5%
	First author or co-author		13.6%	5.1%	31.5%
3. Fitting several models for the same outcome and selected factors					
			14.4%	11.1%	18.3%
Subgroup analysis	Medical statistics experts are included as				
	First author	Co-author			
	No	No	7.3%	4.6%	11.4%
	No	Yes	19.0%	11.5%	29.7%
	Yes	Either	30.7%	23.0%	39.7%
	First author or co-author		26.2%	20.5%	32.9%

Table 3. Estimated proportions of publications using inappropriate/desirable algorithms in multivariate analysis stratified by the number of events, impact factor, and whether medical statistics experts were included as author or not.

		Using only significant variables in univariate analysis			Fitting several models for the same outcome and selected factors		
			95%CI			95%CI	
Subgroup		Proportion	Lower	Upper	Proportion	Lower	Upper
Medical statistics experts included as first author or co-author	The number of events*						
No	<51	20.2%	12.5%	31.1%	2.1%	0.7%	5.9%
	51-100	9.4%	3.2%	24.7%	3.2%	1.1%	8.6%
	100<	8.6%	5.1%	14.2%	10.7%	6.3%	17.7%
Yes	<51	7.7%	2.9%	18.9%	12.6%	5.0%	28.2%
	51-100	4.0%	1.2%	13.0%	30.1%	16.5%	48.6%
	100<	1.6%	0.8%	3.2%	27.0%	20.6%	34.6%
Medical statistics experts included as first author or co-author	Impact factor						
No	Under 2	30.6%	17.1%	48.4%	4.0%	1.1%	13.7%
	2-4<	6.5%	2.4%	16.3%	3.4%	0.8%	13.1%
	4-6<	10.8%	5.8%	19.2%	11.7%	6.1%	21.5%
	Over 6	12.9%	7.5%	21.1%	9.0%	4.2%	18.4%
Yes	Under 2	6.0%	1.9%	17.2%	16.2%	5.4%	39.6%
	2-4<	3.1%	1.1%	8.6%	22.8%	10.5%	42.6%
	4-6<	0.2%	0.0%	1.1%	23.7%	16.1%	33.5%
	Over 6	3.5%	1.7%	6.9%	35.5%	25.9%	46.4%

*The category of "<21" has been integrated with the category "21 - 50" because of insufficient numbers

Table 4. The assessment of the association between the absence of medical statistics experts and the use of inappropriate/desirable algorithms in multivariate analysis with adjustment for potential confounders.

	Using only significant variables in univariate analysis			Fitting several models for the same outcome and selected factors		
		95%CI			95%CI	
Factor	Odds ratio	Lower	Upper	Odds ratio	Lower	Upper
Medical statistics experts included as first author or co-author (vs. no experts)	0.28	0.15	0.53	3.51	1.88	6.58
Medical statistics experts included as first author or co-author (vs. no experts) when 1st author is clinicians or others	0.42	0.19	0.97	2.36	1.03	5.38

All models were adjusted for impact factor and the number of events.

Table 5. Summary of each country and proportion of publications in which medical statistics experts were included as co-author within the publications in which the first author is not an expert in these fields.

Country	Total number of publications	Occupancy (%)	Estimates		
			Publications in which the first author is NOT a medical statistics expert (%)	Medical experts are included as co-author within publications in which the first author is not an expert.	
				Proportion* (%)	95%CI*
USA	501	45.1	67.9	47.4	(40-54.9)
UK	63	5.7	48.2	22.0	(9.6-42.7)
China	51	4.6	84.5	6.7	(2.5-17.1)
Canada	48	4.3	67.4	50.7	(31.5-69.6)
Netherlands	46	4.1	73.1	37.4	(18.3-61.5)
Japan	45	4.0	81.2	15.3	(6.8-30.9)
South Korea	39	3.5	79.5	14.3	(4.9-35.1)
Sweden	38	3.4	40.0	45.3	(22.7-70)
Taiwan	29	2.6	91.3	38.8	(19.1-62.9)
Germany	27	2.4	80.1	41.7	(21.9-64.6)
Denmark	26	2.3	55.4	48.9	(23.9-74.5)
Italy	25	2.2	71.4	13.6	(4.1-36.3)
Australia	25	2.2	42.5	50.6	(16.4-84.3)
France	21	1.9	57.5	77.7	(46.5-93.3)
Spain	19	1.7	62.6	32.7	(11.8-63.8)
Brazil	13	1.2	51.1	4.6	(0.6-29.3)
Norway	11	1.0	48.4	44.8	(9.7-86)
Finland	8	0.7	85.8		
Switzerland	8	0.7	39.6		
Israel	7	0.6	60.9		
Singapore	6	0.5	92.8		
Belgium	6	0.5	64.8		
Turkey	5	0.4	100		
Austria	4	0.4	100		
South Africa	4	0.4	57.4		
Kenya	4	0.4	11.5		
Poland	3	0.3	100		
India	3	0.3	76.3		
Thailand	3	0.3	31.3		
Iran	3	0.3	34.2		
Greece	2	0.2	82.9		
Ireland	2	0.2	32.4		
Others	17	3.4	47.4		
Overall	1112	100	67.3	39.0	(32.2-45.4)

*Calculated only for countries with publications more than 10.

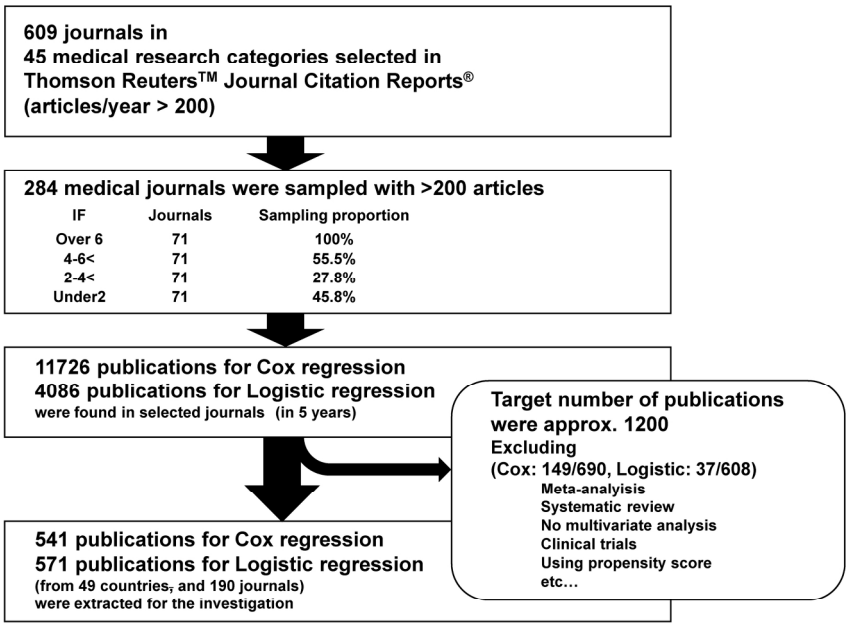

Figure 1.

Summary of the selection of publications investigated in this study.

190x142mm (300 x 300 DPI)

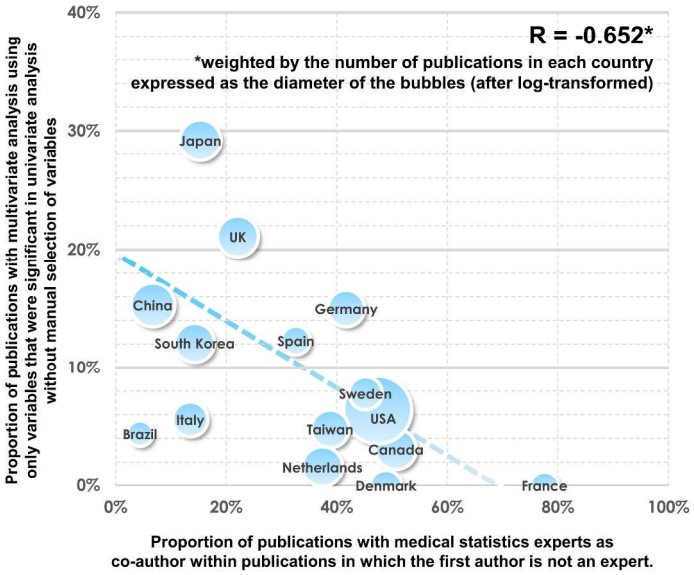

Figure 2.

A scatter plot for the correlation between the proportion of publications using an inappropriate algorithm in multivariate analysis and the proportion of publications in which medical statistics experts were included as co-authors. Inappropriate use of multivariate analysis and presence of experts are inversely correlated.

254x338mm (300 x 300 DPI)

Supplementary Table 1. Selected research filed in Thomson Reuter's Journal Citation Report (version 2014)

ALLERGY
ANESTHESIOLOGY
CARDIAC & CARDIOVASCULAR SYSTEMS
CLINICAL NEUROLOGY
CRITICAL CARE MEDICINE
DENTISTRY, ORAL SURGERY & MEDICINE
DERMATOLOGY
EMERGENCY MEDICINE
ENDOCRINOLOGY & METABOLISM
ENVIRONMENTAL SCIENCES
GASTROENTEROLOGY & HEPATOLOGY
GERIATRICS & GERONTOLOGY
HEALTH CARE SCIENCES & SERVICES
HEMATOLOGY
IMMUNOLOGY
INFECTIOUS DISEASES
INTEGRATIVE & COMPLEMENTARY MEDICINE
MEDICINE, GENERAL & INTERNAL
MEDICINE, RESEARCH & EXPERIMENTAL
NEUROSCIENCES
NURSING
NUTRITION & DIETETICS
OBSTETRICS & GYNECOLOGY
ONCOLOGY
OPHTHALMOLOGY
ORTHOPEDICS
OTORHINOLARYNGOLOGY
PATHOLOGY
PEDIATRICS
PERIPHERAL VASCULAR DISEASE
PHARMACOLOGY & PHARMACY
PSYCHIATRY
PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING
REHABILITATION
REPRODUCTIVE BIOLOGY
RESPIRATORY SYSTEM
RHEUMATOLOGY
SURGERY
TOXICOLOGY
TRANSPLANTATION
TROPICAL MEDICINE
UROLOGY & NEPHROLOGY
VIROLOGY
SUBSTANCE ABUSE

Supplementary Table 2. Journals selected for the investigation in this study.

2013 impact factor			
Over 6	4-<6	2-<4	Under 2
NEW ENGL J MED	ENVIRON MODELL SOFTW	TOXICON	TURK GOGUS KALP DAMA
LANCET	PEDIATRICS	J NEUROL SCI	RENAL FAILURE
JAMA-J AM MED ASSOC	PSYCHO-ONCOLOGY	AM J NEURORADIOL	ENVIRON MONIT ASSESS
J CLIN ONCOL	EXP NEUROL	PHYTOTHER RES	ZH NEVROL PSIKHIATR
BMJ-BRIT MED J	ALIMENT PHARM THER	INT J TUBERC LUNG D	ANIM REPROD SCI
NEURON	PLOS NEGLECT TROP D	J UROLOGY	NEUROL SCI
ENERG ENVIRON SCI	AM J OBSTET GYNECOL	AGR ECOSYST ENVIRON	J EMERG MED
J AM COLL CARDIOL	AM J PATHOL	EXP CELL RES	ENVIRON TOXICOL PHAR
NAT NEUROSCI	PAIN	DIABETES RES CLIN PR	BRAIN INJURY
CIRCULATION	INT J RADIAT ONCOL	OBES SURG	BMC PEDIATR
EUR HEART J	J AM MED INFORM ASSN	J VISION	AM J MED SCI
SCI TRANSL MED	THROMB HAEMOSTASIS	AM J INFECT CONTROL	WATER SCI TECHNOL
GASTROENTEROLOGY	J THROMB HAEMOST	ENVIRON TOXICOL CHEM	J STROKE CEREBROVASC
J EXP MED	ARTHRIT CARE RES	DRUG ALCOHOL DEPEN	CLINICS
J CLIN INVEST	EUR J CANCER	ECOL ECON	PROG UROL
AM J RESP CRIT CARE	AM J RESP CELL MOL	BMC NEUROL	ENVIRON SCI-PROC IMP
J ALLERGY CLIN IMMUN	PSYCHOL MED	VIRUS RES	J VIROL METHODS
HEPATOLOGY	BRIT J PHARMACOL	BIOL REPROD	BURNS
CIRC RES	AM J EPIDEMIOL	EUR J GASTROEN HEPAT	J NEUROSCI METH
J HEPATOL	RESUSCITATION	APPL CATAL A-GEN	J ORAL MAXIL SURG
NEUROSCI BIOBEHAV R	MOVEMENT DISORD	BREAST	PAK J MED SCI
BRAIN	BIOCHEM PHARMACOL	J NEURO-ONCOL	INT J ORAL MAX IMPL
BLOOD	NEUROBIOL AGING	SPINE J	ANN VASC SURG
BIOL PSYCHIAT	AM J KIDNEY DIS	EUR J PHARM SCI	KARDIOL POL
CLIN INFECT DIS	J TRANSL MED	TRANSPLANTATION	J CARDIOTHOR VASC AN
LEUKEMIA	GASTROINTEST ENDOSC	J PHARMACEUT BIOMED	CHINESE MED J-PEKING
CANCER RES	HAEMATOLOGICA	BMC PREGNANCY CHILDB	RHEUMATOL INT
ANN RHEUM DIS	RHEUMATOLOGY	AM J TROP MED HYG	B ENVIRON CONTAM TOX
DIABETES CARE	PROG NEURO-PSYCHOPH	J ENVIRON MANAGE	SUSTAINABILITY-BASEL
ONCOGENE	CLIN J AM SOC NEPHRO	TOXICOL IN VITRO	BONE JOINT J
KIDNEY INT	J AM COLL SURGEONS	MAGN RESON IMAGING	INT J CLIN EXP PATHO
DIABETES	J THORAC CARDIOV SUR	CORNEA	FOOT ANKLE INT
CEREB CORTEX	AM J SURG PATHOL	CHEMOSPHERE	EUR J OBSTET GYN R B
NEUROLOGY	REMOTE SENS ENVIRON	GEN COMP ENDOCR	ENVIRON MANAGE
GLOBAL CHANGE BIOL	J NUTR	CLIN ORAL IMPLAN RES	INT J GYNECOL CANCER
CLIN CANCER RES	OBESITY	BRIT J OPHTHALMOL	SURG TODAY
PLOS PATHOG	EUR RADIOL	TOXICOL APPL PHARM	ONCOL LETT
ARTHRITIS RHEUM-US	J AM ACAD DERMATOL	AM J CARDIOL	INTERNAL MED
NEUROPSYCHOPHARMACOL	INT J OBESITY	CLIN VACCINE IMMUNOL	J DRUGS DERMATOL
ANTIOXID REDOX SIGN	PHARM RES-DORDR	SLEEP MED	SKELETAL RADIOL
HYPERTENSION	J PHYSIOL-LONDON	CLIN EXP RHEUMATOL	PHARM BIOL
EMERG INFECT DIS	BIOL CONSERV	MOL VIS	PEDIATR EMERG CARE
BMC MED	ARTERIOSCL THROM VAS	J AM HEART ASSOC	PEDIATR CARDIOL
J CONTROL RELEASE	ENVIRON POLLUT	FOOD CHEM TOXICOL	EMERG MED J
ANN SURG	J NEUROCHEM	EUR J PHARMACOL	J CRANIOFAC SURG
STEM CELLS	ATHEROSCLEROSIS	ACTA TROP	AM J EMERG MED
CHEST	HUM REPROD	SPINE	ANTICANCER RES
EUR RESPIR J	AM HEART J	FRONT HUM NEUROSCI	ACTA NEUROCHIR
ENVIRON HEALTH PERSP	BREAST CANCER RES TR	MAGN RESON MED	PEDIATR RADIOL
HUM BRAIN MAPP	J CEREBR BLOOD F MET	NEUROSCIENCE	HEPATO-GASTROENTEROL
AM J CLIN NUTR	FERTIL STERIL	CURR MED CHEM	J CLIN NEUROSCI
DIABETOLOGIA	CAN J CARDIOL	J SEX MED	ACTA PAEDIATR
J NEUROSCI	RADIOTHER ONCOL	NUTRIENTS	INDIAN J SURG
J BONE MINER RES	J AM GERIATR SOC	NEPHROL DIAL TRANSPL	RESP PHYSIOL NEUROBI
ANN ONCOL	TOXICOL SCI	FRONT NEURAL CIRCUIT	DEUT MED WOCHENSCHR
AIDS	BONE	PRENATAL DIAG	J MATERN-FETAL NEO M
CLIN GASTROENTEROL H	LIVER INT	J GEN INTERN MED	INT J MED SCI
MOL THER	ENVIRON RES LETT	ARTHROSCOPY	INT J ENDOCRINOL
J INVEST DERMATOL	BRIT J ANAESTH	INT J ONCOL	OTOL NEUROTOL
J CLIN ENDOCR METAB	INFECT IMMUN	ENVIRON SCI POLLUT R	INT J PEDIATR OTORHI
RADIOLOGY	HEALTH AFFAIR	TRIALS	TERAPEVT ARKH
AM J TRANSPLANT	CANCER-AM CANCER SOC	INVEST OPHTH VIS SCI	ANZ J SURG
INT J CARDIOL	OSTEOPOROSIS INT	ARCH VIROL	J KOREAN MED SCI
OPHTHALMOLOGY	CANCER EPIDEM BIOMAR	AM J ROENTGENOL	OR SURG OR MED OR PA
ANESTHESIOLOGY	PSYCHOPHARMACOLOGY	UROL ONCOL-SEMIN ORI	J OBSTET GYNAECOL
CRIT CARE MED	ADDICTION	AM J PHYSIOL-GASTR L	IRAN J PUBLIC HEALTH
NEUROIMAGE	NEUROPHARMACOLOGY	QUAL LIFE RES	OTOLARYNG HEAD NECK
MOL CANCER THER	INT J CANCER	COLORECTAL DIS	J PAEDIATR CHILD H
CORTEX	J NUTR BIOCHEM	VIROL J	BMC COMPLEM ALTERN M
HEART	MOL CELL ENDOCRINOL	WASTE MANAGE	BRIT J ORAL MAX SURG
STROKE	MOL PHARMACOL	EUR J CLIN PHARMACOL	J ENVIRON SCI-CHINA

Supplementary Table 3. Example of multivariate analysis: logistic regression analysis for recurrence after surgery of hypothetical cancer with potential prognostic factors.

Univariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval	
			Lower	Upper
Adjuvant chemotherapy	0.101	0.45	0.17	1.17
Lymph node metastasis	<0.001	8.31	2.88	24.00
Biomarker positive	<0.001	17.11	5.38	54.39

Multivariate Analysis

Potential prognostic factors	P value	Odds ratio	95% Confidence Interval		P value	Odds ratio	95% Confidence Interval	
			Lower	Upper			Lower	Upper
	Multivariate analysis 1				Multivariate analysis 2			
	Using only significant variables in univariate analysis				Using all potential prognostic factors			
Adjuvant chemotherapy	Not included				0.015	0.14	0.03	0.69
Lymph node metastasis	0.005	6.08	1.72	21.51	0.001	12.60	2.67	59.42
Biomarker positive	<0.001	13.77	3.99	47.48	<0.001	16.05	4.11	62.69
	Multivariate analysis 3				Multivariate analysis 4			
	Adjuvant chemotherapy + Lymph node metastasis				Adjuvant chemotherapy + Biomarker positive			
Adjuvant chemotherapy	0.013	0.18	0.05	0.70	0.093	0.35	0.10	1.19
Lymph node metastasis	<0.001	15.63	4.03	60.61	Not included			
Biomarker positive	Not included				<0.001	18.92	5.61	63.89

Inappropriate conclusion about adjuvant chemotherapy:
With multivariate analysis 1, adjuvant chemotherapy has no effect.

Desirable conclusion about adjuvant chemotherapy:
With multivariate analyses 2 to 4, adjuvant chemotherapy was inversely associated with recurrence after adjustment for lymph node metastasis.
Lymph node metastasis was a stronger confounder for the association between adjuvant chemotherapy and recurrence than the biomarker.

Supplementary Table 4. Cross-tabulation table for the association between adjuvant chemotherapy and recurrence stratified by lymph node metastasis for hypothetical cancer.

Lymph node metastasis		No recurrence		recurrence		Total
		Number	%	Number	%	Number
Absent	Without adjuvant chemotherapy	22	73.3%	8	26.7%	30
	With adjuvant chemotherapy	22	91.7%	2	8.3%	24
	Total	44	81.5%	10	18.5%	54
Present	Without adjuvant chemotherapy	1	10.0%	9	90.0%	10
	With adjuvant chemotherapy	8	50.0%	8	50.0%	16
	Total	9	34.6%	17	65.4%	26
Overall	Without adjuvant chemotherapy	23	57.5%	17	42.5%	40
	With adjuvant chemotherapy	30	75.0%	10	25.0%	40
	Total	53	66.3%	27	33.8%	80

Chi-square test for 2x2 table without stratification (Overall): P = 0.098
Odds ratio: 0.45 95% Confidence Interval 0.17-1.17

Mantel-Haenszel test for stratified analysis: P = 0.013
Common odds ratio: 0.19 95% Confidence Interval 0.05-0.71

Supplementary Discussion

The controversy about the term "multivariate/univariate"

The term "multivariable/univariable analysis" instead of "multivariate/univariate analysis" is sometimes recommended for regression analyses by several authors and guidelines because "variate" means random variable in statistics terminology [12]. If we literally follow the definition, "multivariate analysis" may only cover non-regression type analyses for multiple random variables (e.g., principal component analysis and factor analysis) or regression analyses with multiple outcome variables (e.g., multivariate analysis of variance). However, in most situations described as "multivariate analysis", medical researchers' intentions are clear: adjust for multiple covariates as explanatory variables in regression models. In fact, we usually model the conditional expectation E(Y|X) by regression analysis in observational studies where the joint distribution (X, Y) is not controlled by researchers. We thus believe that "multivariate adjustment" or "multivariate analysis" is not necessarily misuse of the terminology. We therefore adopted "multivariate/univariate analysis" in this study as this usage is more common in today's medical literature [12].

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract p.1: "a cross-sectional study"
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found p.2: See the abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported pp.3-4: See the 1st to 5th paragraphs in the introduction section
Objectives	3	State specific objectives, including any prespecified hypotheses p.1 and p.4: See the abstract and the 6th and last paragraphs in the introduction section
Methods		
Study design	4	Present key elements of study design early in the paper pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable pp.6-7: See the materials and methods section (2.2 Surveillance and 2.3. Outcomes)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group pp.6-7: See the materials and methods section (2.2 Surveillance and 2.3. Outcomes)
Bias	9	Describe any efforts to address potential sources of bias pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4. Statistical analyses)
Study size	10	Explain how the study size was arrived at p.8: See the results section (3.1. Characteristics of investigated publications)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4 Statistical analyses)
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding pp.7-9: See Materials and methods section (2.4. Statistical analyses) and Results section (3.3. Subgroup analysis and 3.4. Further analysis for...)
		(*b*) Describe any methods used to examine subgroups and interactions See Materials and methods section (Statistical analyses) and Results section

		(*c*) Explain how missing data were addressed pp.6-7: See the materials and methods section (2.2. Surveillance, 2.3. Outcomes and 2.4 Statistical analyses)
		(*d*) If applicable, describe analytical methods taking account of sampling strategy pp.4-6: See the materials and methods section (2.1. Selection of applicable journals and publications)
		(*e*) Describe any sensitivity analyses pp.7-10: See Materials and methods section (2.4. Statistical analyses) and Results section (3.3. Subgroup analysis, 3.4. Further analysis for... and 3.5. Nation-level investigation)
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed p.8 and p.21 (figure): See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes) and Figure 1
		(b) Give reasons for non-participation at each stage p. 21: See Figure 1
		(c) Consider use of a flow diagram p.21: See Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders p.8: See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes)
		(b) Indicate number of participants with missing data for each variable of interest p.21: See Figure 1
Outcome data	15*	Report numbers of outcome events or summary measures p.8: See Results section (3.1. Characteristics of investigated publications and 3.2. Descriptive statistics of the outcomes)
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included pp.16-20: See Tables
		(*b*) Report category boundaries when continuous variables were categorized pp.16-20: See Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses pp.16-20: See Tables
Discussion		
Key results	18	Summarise key results with reference to study objectives p.10: See the 1st paragraph in the discussion section
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias p.12: See the discussion section (4.1. Limitations)

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p.13: See the discussion section (4.3. Conclusion)
Generalisability	21	Discuss the generalisability (external validity) of the study results pp.10-13: See the whole discussion section (but in particular, intensively described in the 6th and 7th paragraphs, 4.1. Limitations and 4.3. Conclusion)
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p.13: See Funding source section

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.