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 Fumitaka Nagamura1,2

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% to 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 coauthor 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 OR of the involvement of experts for this outcome was 0.28 (95% CI 0.15 to 0.53). A further, nation-level, analysis showed that the involvement of experts and the implementation of unfavourable multivariate analysis are associated at the nation-level analysis (R² = −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.

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 specialise in biostatistics. Moreover, common misconceptions have been formed among medical researchers who are not specialised 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
research undertaking multivariate analysis. To target Here, target publications in this study are about medical
This study was conducted as a cross-sectional study.
selection of applicable journals and publications
MAterIAls And MethOds
population in the implementation of multivariate analysis
outcomes. Based on this research, we found a high-risk
ment (such as biostatistician and epidemiologist) and the
understanding of statistical models. We therefore assessed
science and biostatistics is necessary to enable appropriate
on distribution of variables and correlation between explan-
iAnalysis is included in multivariate analysis, right?’
Knowing in what situations such inappropriate analysis
is being done should lead to improvement in the quality
statistical analysis in medical research. However, there
are no reports that summarise how multivariate analysis is
being carried out, including whether medical statistical experts
are involved or not.
Based on the above situation, we decided to investiga-
ate under what circumstances inappropriate use is likely
to occur and to identify the population that needs more
support. Since inappropriate use of multivariate analysis (par-
cularly in variable selection for regression model
construction) is found even in published papers, we inves-
tigated 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 anal-
ysis, 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 explana-
tory 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 involve-
ment (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.

MATERIALS AND METHODS
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 multi-
step 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 that 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 jour-
nals from the list in the website in 2014 (‘JCR year’
was 2013). Selected research fields are listed in online
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: ‘Under 2’, ‘2 to <4
(two to less than 4)’, ‘4 to <6 (four to less than 6)’ and
‘6 or over’. 3. Subsequently, we selected journals whose number of
articles exceeds 200/year to avoid journals with few ar-
ticles 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 online supplementary table 2). Sam-
pling rates according to impact factor were: 6 or over:
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 online supplementary
table 2.
4. We searched for publications in which logistic re-
gression/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, re-
spectively. A publication database with 4086 (for logis-
tic) and 11726 (for Cox) publications was constructed
through the previously described process. Clinical tri-
als 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 in-
cluded in the title or abstract. All publications were
from journals available through the University of To-
kyo or open access articles.
5. To set the 95% CI 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 cal-
culated by applying a power function with an expo-
nent <1 to the number of publications (for logistic reg-
ression: 0.54×N0.664, for Cox regression: 0.54/N0.664,
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
Nojima M, et al. BMJ Open 2018;8:e021129. doi:10.1136/bmjopen-2017-021129
papers (for logistic) and 541 (for Cox) were selected as the research subject. This number satisfies the target CI set above.

**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, categorised as: <21, 21–50, 51–100 and >100), number of the covariates (categorised as: <3, 3–5, 6–10, and >10), etc.

We decided whether authors or coauthors 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 end point 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 determined as a desirable outcome in 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.

**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. If the number of covariates was over 10 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, it was defined as the event. The criterion was basically based on the study from Peduzzi et al., 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 <51 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.

**STATISTICAL ANALYSES**

Statistical analyses for binomial outcomes were performed using weighted generalised 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 V.23 (IBM).

**PATIENT AND PUBLIC INVOLVEMENT**

Neither were involved.

**RESULTS**

**Characteristics of investigated publications**

The flow chart of the selection of the research subjects is summarised 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 coauthors was estimated to be 37.8%.

**DESCRIPTIVE STATISTICS OF THE OUTCOMES**

Descriptive statistics of the outcomes are summarised 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% to 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 coauthor. When an expert was included as the first author or coauthor, 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 3.5% if an expert was included as coauthor. Since these are only for research with few events, the estimation accuracy was low. When an expert was included as the first author or coauthor, 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 coauthorship 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 coauthor, 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.

**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, unfavourable 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).

**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 OR of the involvement of experts for ‘using only variables that are significant in univariate analysis’ was 0.28 (95% CI 0.15 to 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 coauthor, the risk decreased with an OR of 0.42 (95% CI 0.19 to 0.97). Likewise, for ‘Fitting several models for the same outcome and selected factors’, the result was favourable when an expert was included (OR 3.51; 95% CI 1.88 to 6.58 for any type of author, OR 2.36 for only as coauthor, 95% CI 1.03 to 5.38).

**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 (%) | 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 Coauthor                                                   |            |        |           |           |
| No No                                                                    | 12.2       | 8.7    | 16.8      |           |
| No Yes                                                                   | 3.5        | 2.0    | 6.1       |           |
| Yes Either                                                               | 1.1        | 0.3    | 3.5       |           |
| First author or coauthor                                                | 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 Coauthor                                                   |            |        |           |           |
| 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 coauthor                                                | 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 Coauthor                                                   |            |        |           |           |
| 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 coauthor                                                | 26.2       | 20.5   | 32.9      |           |
Finally, we examined how much medical statistics experts are involved as coauthors 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 USA, accounting for an overwhelming majority compared with 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 coauthor 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 >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 involvement.

| Subgroup                                      | The number of events* | Using only significant variables in univariate analysis | Fitting several models for the same outcome and selected factors |
|-----------------------------------------------|-----------------------|-------------------------------------------------------|---------------------------------------------------------------|
| Medical statistics experts included as first author or coauthor | 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 coauthor | Impact factor | Using only significant variables in univariate analysis | Fitting several models for the same outcome and selected factors |
|-----------------------------------------------------------------|--------------|-------------------------------------------------------|---------------------------------------------------------------|
| 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 | 6 or over: 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 | 6 or over: 3.5, 1.7, 6.9 | 35.5, 25.9, 46.4 |

*Nation-level investigation*

Finally, we examined how much medical statistics experts are involved as coauthors 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 USA, accounting for an overwhelming majority compared with 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 coauthor 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 >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 involvement.

| 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 | Fitting several models for the same outcome and selected factors |
| Group                                           | Proportion (%) | 95% CI Lower (%) Upper (%) | Proportion (%) | 95% CI Lower (%) Upper (%) |
| No                                              | 20.2, 12.5, 31.1 | 2.1, 0.7, 5.9 | 9.4, 3.2, 24.7 | 3.2, 1.1, 8.6 | 8.6, 5.1, 14.2 | 10.7, 6.3, 17.7 |
| Yes                                             | 7.7, 2.9, 18.9 | 12.6, 5.0, 28.2 | 4.0, 1.2, 13.0 | 30.1, 16.5, 48.6 | 1.6, 0.8, 3.2 | 27.0, 20.6, 34.6 |

| Medical statistics experts included as first author or coauthor | Impact factor | Using only significant variables in univariate analysis | Fitting several models for the same outcome and selected factors |
|-----------------------------------------------------------------|--------------|-------------------------------------------------------|---------------------------------------------------------------|
| 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 | 6 or over: 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 | 6 or over: 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

| Factor | Using only significant variables in univariate analysis | Fitting several models for the same outcome and selected factors |
|--------|-------------------------------------------------------|---------------------------------------------------------------|
|        | OR         | 95% CI Lower (%) Upper (%) | OR         | 95% CI Lower (%) Upper (%) |
| Medical statistics experts included as first author or coauthor (vs no experts) | 0.28, 0.15, 0.53 | 3.51, 1.88, 6.58 | 0.42, 0.19, 0.97 | 2.36, 1.03, 5.38 |

All models were adjusted for impact factor and the number of events.
20% or less except for Taiwan. For other European countries, there is variability in the result. The involvement of experts and the implementation of unfavourable multivariate analysis are associated at the nation-level analysis. The details are summarised in table 5.

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

| Country     | Total number of publications | Occupancy (%) | Publications in which the first author is NOT a medical statistics expert (%) | Medical experts are included as coauthor within publications in which the first author is not an expert. |
|-------------|------------------------------|---------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| USA         | 501                          | 45.1          | 67.9                                                                            | 47.4 (%) 40 to 54.9                                                                               |
| UK          | 63                           | 5.7           | 48.2                                                                            | 22.0 (%) 9.6 to 42.7                                                                               |
| China       | 51                           | 4.6           | 84.5                                                                            | 6.7 (%) 2.5 to 17.1                                                                               |
| Canada      | 48                           | 4.3           | 67.4                                                                            | 50.7 (%) 31.5 to 69.6                                                                               |
| The Netherlands | 46                       | 4.1           | 73.1                                                                            | 37.4 (%) 18.3 to 61.5                                                                               |
| Japan       | 45                           | 4.0           | 81.2                                                                            | 15.3 (%) 6.8 to 30.9                                                                               |
| South Korea | 39                           | 3.5           | 79.5                                                                            | 14.3 (%) 4.9 to 35.1                                                                               |
| Sweden      | 38                           | 3.4           | 40.0                                                                            | 45.3 (%) 22.7 to 70                                                                                |
| Taiwan      | 29                           | 2.6           | 91.3                                                                            | 38.8 (%) 19.1 to 62.9                                                                               |
| Germany     | 27                           | 2.4           | 80.1                                                                            | 41.7 (%) 21.9 to 64.6                                                                               |
| Denmark     | 26                           | 2.3           | 55.4                                                                            | 48.9 (%) 23.9 to 74.5                                                                               |
| Italy       | 25                           | 2.2           | 71.4                                                                            | 13.6 (%) 4.1 to 36.3                                                                               |
| Australia   | 25                           | 2.2           | 42.5                                                                            | 50.6 (%) 16.4 to 84.3                                                                               |
| France      | 21                           | 1.9           | 57.5                                                                            | 77.7 (%) 46.5 to 93.3                                                                               |
| Spain       | 19                           | 1.7           | 62.6                                                                            | 32.7 (%) 11.8 to 63.8                                                                               |
| Brazil      | 13                           | 1.2           | 51.1                                                                            | 4.6 (%) 0.6 to 29.3                                                                                |
| Norway      | 11                           | 1.0           | 48.4                                                                            | 44.8 (%) 9.7 to 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 to 45.4                                                                               |

*Calculated only for countries with publications >10.

**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
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 coauthors. Inappropriate use of multivariate analysis and presence of experts are inversely correlated.

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 was issued7 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 online 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 (see online 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 USA, the UK 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 analysed (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 standardised by the CDISC standard to regulatory authorities (Food and Drug Administration, Pharmaceuticals and Medical Devices Agency, etc) for further validation and additional analysis. Such standardisation 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.

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.

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 online supplementary discussion for further details.

CONCLUSION
In publications about observational research in which the number of events is 50 or less without the involvement of medical statistics experts, >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 aware of the importance of medical statistics experts.
more cognizant of appropriate regression model construction in multivariate analysis.

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

Contributors 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.

Funding This study was supported by grants-in-aid for scientific research (C), JSPS KAKENHI grant number JP 26460764 (fiscal-year 2014-2016, Masanori Nojima).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES
1. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:e1001221–12.
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, 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. 2nd edn. Oxford, 2012.
5. Peduzzi P, Concato J, Kemper E, et al. 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, et al. 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. 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, et al. 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.