Transparency considerations for describing statistical analyses in research

Sang Kyu Kwak¹, Jonghae Kim²

Departments of ¹Medical Statistics, ²Anesthesiology and Pain Medicine, School of Medicine, Daegu Catholic University, Daegu, Korea

Researchers who use the results of statistical analyses to draw conclusions about collected data must write a statistical analysis section in their manuscript. Describing statistical analyses in precise detail is as important as presenting the dosages of drugs and methodology of interventions. It is also essential for scientific accuracy and transparency in scientific research. We evaluated the quality of the statistical analysis sections of clinical research articles published in the Korean Journal of Anesthesiology between February 2020 and February 2021. Using a Likert scale where 1, 2, and 3 represented “not described at all,” “partially described,” and “fully described,” respectively, the following 6 items were assessed: 1) stating of the statistical analysis methods used, 2) rationale for and detailed description of the statistical analysis methods used, 3) parameters derived from the statistical analyses, 4) type and version of the statistical software package used, 5) significance level, and 6) sidedness of the test (one-sided vs. two-sided). The first 3 items evaluate issues directly related to the statistical analysis methods used and last 3 are indirectly related items. In all the included articles, the statistical analysis methods used were stated (score of 3). However, only 4 articles (12.9%) fully described the sidedness of the test (score of 3). Authors tend not to describe the sidedness of statistical analysis tests in the methodology section of clinical research articles. It is essential that the sidedness be described in research studies.

Keywords: Data analysis; Probability; Research; Software; Statistical data interpretation; Statistics.
methods, 4) type and version of the statistical software package used, 5) significance level, and 6) sidedness of the statistical test (one-sided vs. two-sided).

In this study, the statistical analysis sections of 31 clinical research papers published in the *Korean Journal of Anesthesiology* (KJA) between February 2020 and February 2021 were evaluated using the above-mentioned 6 items. Through this process, practical ways of increasing the transparency of research to build scientific evidence are presented.

**Materials and Methods**

In this study, the statistical analysis sections of clinical research papers published in the KJA between February 2020 and February 2021 were evaluated. The sample size calculation was not assessed. Editorials, review articles, statistical round articles, case reports, letters to the editor, and corrigendum, all of which do not include statistical analysis sections, were excluded from the analysis. Additionally, the statistical analysis sections of experimental studies were not evaluated.

**Evaluation items directly related to statistical analyses**

**Stating the statistical analysis methods used**

Statistical analyses are used to calculate the significance probability (probability value) of the test statistic (estimated based on the probability distribution corresponding to the statistical analysis method used) to test the hypothesis of a study. There are various probability distributions of test statistics, including t distribution, F distribution, and chi-square distribution. For example, if the distribution of the test statistic is a t distribution, the statistical analysis method is called a t-test. However, various statistical analysis methods use the t distribution: a one-sample t-test, an independent two-sample t-test, and a paired t-test. Therefore, the name of the statistical analysis methods used should be specifically and accurately presented.

**Rationale for and detailed description of the statistical analysis methods used**

Although the general descriptions of statistical analysis methods are readily available in related books, finding the best statistical analysis methods consistent with the research design and hypothesis is another issue. Hence, the rationale for using the statistical analysis methods should be clearly described in the statistical analysis section. If a t-test is used to assess the probability of the t-statistic calculated from the mean difference between 2 independent groups, researchers can simply state that the t-test was used to evaluate the statistical significance of the mean difference of the quantitative data between the 2 groups. However, some complicated statistical analysis methods require several steps and require the consideration of various factors. For example, when performing multiple linear regression or multiple logistic regression analyses, the variable selection method should be specified. If a principal component analysis is performed, the factor rotation method should be clear. Additionally, to fully and accurately describe propensity score matching, the adjustment method, caliper value, and ratio of pair-matching must be stated. Therefore, the statistical analysis section should describe in detail the rationale and steps taken for the statistical analysis methods used.

**Parameters derived from statistical analysis methods**

To test the hypothesis of a research study, statistics are estimated by a statistical analysis method. Statistics are used as the results of data analyses and determine whether the hypothesis is correct. For example, if a t-test is used to test whether there is a significant difference in means between 2 independent groups, the means and standard deviations of each group and the probability value that determines the significance of the mean difference are presented. The mean difference between the 2 groups and pooled standard deviation should be presented. If a logistic regression analysis is performed, the following parameters can be presented: the odds ratios of each variable with their 95% CIs and the probability value that determines whether the odds ratios are statistically different from 1, that is, whether the odds ratios are significant or not. Therefore, the parameters that are derived from statistical analysis methods and used as the results of a study, need to be presented in detail.

**Evaluation items indirectly related to statistical analysis methods**

**Type and version of statistical software package used**

Various statistical software packages are available, such as IBM SPSS Statistics (www.ibm.com/products/spss-statistics), R (www.r-project.org), SAS (www.sas.com), Minitab (www.minitab.com), MedCalc (www.medcalc.org), NCSS (www.ncss.com/software), and Excel (www.microsoft.com/en-us/microsoft-365/excel). The versions of these software packages are updated as new functions are added or old functions are upgraded. Since they have different manufacturers, the algorithms or methods used for the statistical analyses may be different. That is, the common statistical analyses with common data performed by different statistical software packages can produce results that are different from each other. Therefore, information regarding the software and the version
used for statistical analyses should be clearly stated in the statistical analysis section.

**Significance level**

It is mandatory to include the significance level for statistical hypothesis testing since conclusions are directly drawn from it. The significance level is the maximum probability that a type I error (an error of rejecting the null hypothesis even though the null hypothesis is correct) is tolerable. In a statistical hypothesis test, the calculated probability value is compared with the significance level set by the researcher to determine whether the null hypothesis can be rejected. The significance level was set at 0.05 (5%) in all 31 articles analyzed in this study. If the significance level is set at a relatively high value, such as 20% or 30%, rather than 5%, the results of the study may be unreliable. Therefore, the significance level is generally set no higher than 5%. Occasionally, the significance level is set at 1%. This does not mean that setting a significance level of 10% is incorrect. The authors should only clearly state that the significance level was set at 10%. Therefore, the significance level should be explicitly presented in the statistical analysis section.

**Sidedness of the test (one-sided vs. two-sided)**

Statistical hypothesis testing builds null and alternative hypotheses and is used to determine whether to reject the null hypothesis through a series of processes. To indicate inequality when formulating an alternative hypothesis, the symbol, “≠” should be used if the test is two-sided and “ >” or “ <” should be used if the test is one-sided. For example, if a two-sided t-test is used to determine whether the mean difference between 2 independent groups (A and B) is significant, the null hypothesis (H₀) and the alternative hypothesis (Hₐ) are as follows:

(1) \( H₀: \mu_A = \mu_B \) vs. \( Hₐ: \mu_A \neq \mu_B \)

where \( \mu_A \) is the population mean of group A, and \( \mu_B \) is the population mean of group B.

If a one-sided t-test is used to test whether the mean difference between 2 independent groups is significantly greater or less than the other, \( H₀ \) and \( Hₐ \) are shown as follows:

(2) \( H₀: \mu_A = \mu_B \) vs. \( Hₐ: \mu_A > \mu_B \) or \( H₀: \mu_A = \mu_B \) vs. \( Hₐ: \mu_A < \mu_B \)

The range of the t-statistic to determine statistical significance depends on the sidedness of the test. Likewise, different probabili-

---

https://doi.org/10.4097/kja.21203
performed to determine the odds ratios of the risk factors for inadequate emergence. In particular, the variable selection method for binary logistic regression was stepwise selection 1) (3 points). According to the type of data or statistical analysis methods used, the statistics were presented as counts (percentages) for frequencies, mean with standard deviation for continuous variables, and odds ratio (95% CI) for the logistic regression analysis results (3 points). The statistical analysis software package used was SPSS software ver. 22.0 (IBM Corp., USA) (3 points), the significance level was set at 0.05 (3 points), and a two-sided test was used (3 points).

Another example of a statistical analysis section, taken from a study conducted by Kaur et al. [2], is as follows:

“The normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed continuous variables are presented as the mean ± SD, whereas ordinal variables (NRS score) are presented as the mean ± SD (median). Means were also used to describe the ordinal data along with median. A one-way analysis of variance was used to compare the means among the three independent groups. The Kruskal-Wallis test was used followed by multiple comparisons (Bonferroni test) to compare the distribution of the NRS pain scores among the three study groups. A paired sample t-test was used to test the change in means between the pre to post observations. Fisher's exact test was used to compare the proportions between the groups. A two-sided P value of < 0.05 was considered statistically significant. Statistical Package for Social Sciences, version 23 (SPSS v.23, IBM Corp., USA) was used for data analysis.”

The statistical analysis methods used in this study were the Shapiro-Wilk test, one-way analysis of variance, Kruskal-Wallis test, paired sample t-test, and Fisher’s exact test (3 points). With regard to the rationale for each statistical analysis method used, the Shapiro-Wilk test was used to test the normality of the data, the one-way analysis of variance was used to compare the means among the 3 independent groups, the Kruskal Wallis test was used to compare the distribution of the NRS pain scores among the 3 study groups, the paired sample t-test was used to test the change in means between the pre- and post-observations, Fisher’s exact test was used to compare the proportions between the groups, and the Bonferroni test was used to adjust the probability values for multiple comparisons (3 points). According to the type of data, the statistics were presented as means ± standard deviations for normally distributed continuous variables and means ± standard deviations (median) for ordinal variables (3 points). The statistical analysis software package used was Statistical Package for Social Sciences version 23 (SPSS v.23, IBM Corp., USA) (3 points), the significance level was set at 0.05 (3 points), and two-sided tests were used (3 points).

The above methodology was applied to assess the statistical analysis sections of the remaining articles.

Results

From February 2020 to February 2021, a total of 111 papers were published in the KJA. Among them, 31 clinical research articles were included in the analysis after 10 editorials, 16 review articles, 5 statistical round articles, 15 case reports, 30 letters to editor, and 4 experimental studies were excluded.

The evaluation items, from highest to lowest rankings, are as follows: stating of the statistical analysis methods used, type and version of statistical software package, significance level, parameters derived from statistical analysis methods, rationale for and detailed description of the statistical analysis methods used, and sidedness of statistical tests (Table 1). All 31 articles stated the statistical analysis methods used appropriately (score of 3). Of the 31 articles, 29 (93.5%) presented the type and version of the statistical software package used (score of 3). The number (percentage) of the articles that received a score of 3 for the two items (parameters derived from statistical analysis methods and rationale for and details of statistical analysis methods) was 24 (77.4%) and 23 (74.2%), respectively. The studies by Lee et al. [3] and Hervás et al. [4] received scores of 2 for the “rationale for and details of the statistical analysis methods used” item because, although they stated that they used propensity score matching, no descriptions of the adjustment method, caliper value, or the ratio of pair-matching were found, which would significantly affect the results of the subsequent statistical analysis. Likewise, the papers by Lim et al. [5] and Tamboli et al. [6] reported performing statistical analyses without describing the rationale for and details of the statistical analysis methods used. Thus, they each received a score of 2 for this item. Lastly, the sidedness of statistical tests (one-sided vs. two-sided) was explicitly described in only 4 (12.9%) of the 31 articles, which was the lowest ranked evaluation item.

Discussion

Surprisingly, most of the articles did not specify the sidedness of the statistical tests used. The KJA author guidelines do not explicitly request the sidedness of statistical tests be included. However, it requires that probability values should be two-tailed except

---

1) It is unclear whether the authors performed backward elimination or stepwise selection to select the variables to be included in the regression model because they used the unclear term, “backward binary stepwise logistic regression.”
<table>
<thead>
<tr>
<th>No.</th>
<th>Last name of first author</th>
<th>Year</th>
<th>Volume</th>
<th>Issue</th>
<th>Evaluation items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gautam [7]</td>
<td>2021</td>
<td>74</td>
<td>1</td>
<td>★★★ ★★★ ★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>2</td>
<td>Lee [3]</td>
<td>2021</td>
<td>74</td>
<td>1</td>
<td>★★★ ★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>3</td>
<td>Hervás [4]</td>
<td>2021</td>
<td>74</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>4</td>
<td>Choi [8]</td>
<td>2021</td>
<td>74</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>5</td>
<td>Burton [9]</td>
<td>2021</td>
<td>74</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>6</td>
<td>Lee [10]</td>
<td>2020</td>
<td>73</td>
<td>6</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>8</td>
<td>Lee [12]</td>
<td>2020</td>
<td>73</td>
<td>6</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>10</td>
<td>Ryu [14]</td>
<td>2020</td>
<td>73</td>
<td>6</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>15</td>
<td>Aoyama [16]</td>
<td>2020</td>
<td>73</td>
<td>4</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>16</td>
<td>Chadha [17]</td>
<td>2020</td>
<td>73</td>
<td>4</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>17</td>
<td>Banik [18]</td>
<td>2020</td>
<td>73</td>
<td>4</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>18</td>
<td>Makarem [1]</td>
<td>2020</td>
<td>73</td>
<td>4</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>19</td>
<td>Udapi [19]</td>
<td>2020</td>
<td>73</td>
<td>3</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>20</td>
<td>Boscolo [20]</td>
<td>2020</td>
<td>73</td>
<td>3</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>21</td>
<td>Said [21]</td>
<td>2020</td>
<td>73</td>
<td>3</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>22</td>
<td>Azemati [22]</td>
<td>2020</td>
<td>73</td>
<td>3</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>23</td>
<td>Srivastava [23]</td>
<td>2020</td>
<td>73</td>
<td>2</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>26</td>
<td>Salama [26]</td>
<td>2020</td>
<td>73</td>
<td>2</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>27</td>
<td>Doo [27]</td>
<td>2020</td>
<td>73</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>28</td>
<td>Park [28]</td>
<td>2020</td>
<td>73</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>29</td>
<td>Ökmen [29]</td>
<td>2020</td>
<td>73</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>30</td>
<td>Choi [30]</td>
<td>2020</td>
<td>73</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
<tr>
<td>31</td>
<td>Danforth [31]</td>
<td>2020</td>
<td>73</td>
<td>1</td>
<td>★★★ ★★★ ★★★ ★★★★ ★★★★ ★★★</td>
</tr>
</tbody>
</table>

Number (percentage) of each score

| ★  | 0 (0.0)  | 3 (9.7) | 5 (16.1) | 2 (6.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
|★★ | 0 (0.0)  | 5 (16.1) | 2 (6.5) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) |
|★★★| 31 (100) | 23 (74.2) | 24 (77.4) | 29 (93.5) | 27 (87.1) | 4 (12.9) |

[Evaluation items]
1 Stating the statistical analysis methods used
2 Rationale for and detailed description of statistical analysis methods used
3 Parameters derived from statistical analysis methods
4 Type and version of statistical software package used
5 Significance level
6 Sidedness of statistical test (one-sided vs. two-sided)

[Symbols]
★: not described at all
★★: partially described
★★★: fully described

https://doi.org/10.4097/kja.21203
for study designs that require a one-tailed test. Since most of the authors of clinical research articles perform two-tailed tests, the readers are likely to think that the tests are two-tailed unless otherwise specified. However, it is not best practice to omit important information conventionally in scientific papers that must provide accurate information.

There are several important reasons why statistical analysis sections should include information on the 6 items mentioned in this article. First, scientific evidence should be available to support the appropriate use of statistical methods to test the research hypothesis. For example, it is inappropriate to perform parametric tests using data that do not fulfill the normality assumption. Instead, non-parametric tests should be performed for non-normally distributed data. Performing a t-test twice (A vs. B and A vs. C) or 3 times (A vs. B, A vs. C, and B vs. C) to compare the means of 3 independent groups (A, B, and C) is also inappropriate. One-way analysis of variance should be used to test the null hypothesis regarding the difference in means of 3 groups. If the null hypothesis is rejected, a post-hoc test is necessary. Second, the transparency of a study can be ensured by describing how the statistical analysis method was performed with the type and version of the statistical software package, significance level, and the sidedness of statistical tests used. For example, when performing logistic regression analyses, whether covariates are included (and if so, which covariates) and the significance level of the univariate analyses, which is used to determine which independent variables to include in the multivariable analysis, should be described. In propensity score matching, the results of the analysis vary depending on the propensity score correction method, caliper value, and the ratio of pair-matching, and thus these should be described in detail. The type and version of the statistical software package, significance level, and sidedness of the tests should also be presented. Third, describing the rationale or purpose for choosing the selected analysis methods provides evidence that the analyses were appropriate to reveal the causal relationship between the variables measured in the study, making it scientific. For example, when an independent two-sample t-test is used, a statement such as “After the normality assumption had been met by the Kolmogorov-Smirnov test, independent two-sample t-test was used to compare the pain scores between the 2 analgesics” could be used.

However, an important limitation of this study should be considered. We did not check whether the results of all the statistical analyses described in the statistical analysis section were presented clearly in the results section. It is therefore unclear whether the results of the statistical analyses that were not described in the statistical analysis section are presented in the results section. However, we believe that the peer review system of this journal has already resolved this issue.

Researchers need to recognize what should be included when describing statistical analyses in research proposals, results reports, theses, and articles. The inclusion of the above-mentioned items, which are directly and indirectly related to statistical analyses, in the methodology section improves the transparency and scientific nature of a study. In particular, for complicated analyses or the use of multiple statistical analyses, more effort to describe the essential points we have proposed is needed to clarify how the presented results were obtained. While it is difficult to generalize our results since we did not assess all articles that have been published, we emphasize that researchers should pay particular attention to the items that were rated poorly in this study (sidedness of statistical tests [one-sided vs. two-sided], rationale for and details of the statistical analysis methods used, and parameters derived from statistical analysis methods) and should describe them clearly in their statistical analysis sections.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Author Contributions

Sang Kyu Kwak (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Supervision; Validation; Writing – original draft; Writing – review & editing)
Jonghae Kim (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Supervision; Validation; Writing – original draft; Writing – review & editing)

ORCID

Sang Kyu Kwak, https://orcid.org/0000-0003-0398-5514
Jonghae Kim, https://orcid.org/0000-0003-1222-0054

References


