Statistical review of 95 studies employing repeated-measures analysis of variance published in the Korean Journal of Anesthesiology

Sang-Il Park¹, Dong Kyu Lee², and Junyong In³

Department of Anesthesiology and Pain Medicine, ¹Chungnam National University College of Medicine, Daejeon, ²Korea University Guro Hospital, Seoul, ³Dongguk University Ilsan Medical Center, Goyang, Korea

Repeated-measures analysis of variance (RMANOVA) continues to be used widely for the analysis of repeated-measures data in anesthetic research. RMANOVA, like other parametric statistical tests, specifies several assumptions and requires specific description in publications. A recent article in the *Korean Journal of Anesthesiology (KJA)* provided valid information on RMANOVA [1].

We reviewed 493 studies published in *KJA* between January 2010 and December 2014. A total of 95 studies (19.3%) employed RMANOVA to analyze repeated-measures data. We reviewed these studies and summarized the following: type of variables, statistical assumptions (normally distributed population, homogeneity of variances, and sphericity assumption), interaction between two factors in the results, sample size calculation, post-hoc testing, and type of statistical package used.

In addition to a dependent variable (e.g. blood pressure), all the studies (n = 95) contained one within-subject factor (e.g. different measurement sites, such as left and right arm, or a time

series post-treatment, such as 0, 10, and 20 minutes in the same participant) as an independent variable. Some of these studies (n = 35) also contained one between-subject factor (e.g. different drugs or dosages administered to randomly assigned participants) as an additional independent variable. When a study involves one within-subject factor only, one-way RMANOVA is required, while “two-way RMANOVA” should be used for those that contain one within-subject factor and one between-subject factor. However, there is some confusion regarding the exact definition of “two-way RMANOVA,” although it is the most frequently used term in most *KJA* publications. A “true” two-way RMANOVA (or two-way ANOVA with repeated-measures in two factors) compares mean differences between data sets, split into “two within-subject factors.” Each data set is collected at one measurement point (one within-subject factor) under one treatment condition (another within-subject factor). For example, if every participant has their blood pressure measured three times (“the first within-subject factor”; for example, just before injection, just after injection, and 1 minute after injection) under propofol, thiopental sodium, and ketamine induction (“the second within-subject factor”), 2 weeks apart, nine data sets will be collected from a single group of participants. However, in anesthetic research, the participants are frequently randomized into two or more separate treatment conditions (e.g. each participant receives propofol, thiopental sodium, or ketamine induction [“between-subject factor”] and their blood pressure is measured three times [“within-subject factor”]); therefore, nine data sets will be collected from three groups of participants. Consequently, researchers should consider the differences between the “two-way RMANOVA” and the “two-way ANOVA with repeated-measures in one factor (also called mixed-model

Corresponding author: Junyong In, M.D.
Department of Anesthesiology and Pain Medicine, Dongguk University Ilsan Medical Center, 27, Dongguk-ro, Ilsandong-gu, Goyang 10326, Korea
Tel: 82-31-961-7875, Fax: 82-31-961-7864
E-mail: dragona1@dumc.or.kr
ORCID: <http://orcid.org/0000-0001-7403-4287>

Received: August 6, 2015.
Revised: 1st, November 2, 2015; 2nd, November 19, 2015.
Accepted: November 20, 2015.

Korean J Anesthesiol 2016 February 69(1): 97-99
<http://dx.doi.org/10.4097/kjae.2016.69.1.97>

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © the Korean Society of Anesthesiologists, 2016

Online access in <http://ekja.org>

ANOVA, mixed-design ANOVA, or mixed ANOVA),” the latter of which is the most frequently used.

There are other factors to consider and describe in studies employing RMANOVA (Table 1). A paper by Wei et al. [2] is a good example of how to sufficiently describe RMANOVA and may assist other researchers in reporting this analysis in a manuscript. First, when researchers analyze repeated-measures data, they have to check whether the data can actually be analyzed using a RMANOVA. The dependent variable should be quantitative because RMANOVA is a parametric test. However, seven studies included qualitative variables as dependent variables, which were the level of thoracic block under spinal anesthesia, Numeric Pain Score, Ramsay Sedation Scale, modified Children's Hospital of Eastern Ontario Pain Scale, modified Yale Preoperative Anxiety Scale, and modified Observer's Assessment of Alertness/Sedation Scale. Second, for parametric testing, several assumptions have to be evaluated and satisfied. The normality of data is one of these and is commonly verified using either the Shapiro-Wilk test or the Kolmogorov-Smirnov test. In our review, only 8 studies provided a description of the normality testing procedure. If data fails the assumption of normality, researchers should transform their data to satisfy the assumption of normality or alternatively select a non-parametric test. Homogeneity of variances across a whole data set is another fundamental assumption of parametric tests and is frequently verified using the Levene's test or Bartlett's test. Violation of this assumption affects the F-statistic of the RMANOVA and reduces result reliability. All of the reviewed studies employing RMANOVA did not disclose whether homogeneity of variances was tested. In addition, variances of the differences between all combinations of data sets are required to be equal. This is referred to as the sphericity assumption [1]. If the sphericity assumption is violated, researchers need to use adjustments or consider performing alternative statistical methods for repeated-measures data such as multivariate analysis of variance or mixed-effects modeling [1]. However, with the exception of

two, all reviewed studies failed to state whether the sphericity assumption was satisfied. Third, a description of the interaction between the within-subject factor and the between-subject factor on the dependent variable is required in the results section because the interaction is the starting point for further statistical analysis [1]. Despite the importance of this interaction, none of the studies we reviewed demonstrated this adequately. Fourth, if a RMANOVA is planned to analyze primary outcomes, the sample size calculation should be based on RMANOVA. However, none the studies calculated and described this properly. We recommend Kang's article as a useful guide for calculating sample sizes for RMANOVA [3]. Fifth, the majority of studies presented statistical significance and post-hoc test results. There are a variety of post-hoc tests, each with individual characteristics. Therefore, researchers should be aware of the nature of each post-hoc test and describe them sufficiently in the literature. In the present review, only 59% of studies disclosed which post-hoc test had been employed. Finally, the statistical software package used also comprises important information for readers. Ten studies did not state which statistical package was used.

In conclusion, researchers not only should be aware of how to perform RMANOVA, but also should describe the statistical process and results correctly. They should refrain from simply listing a series of the statistical tests employed since this overcomplicates matters and does not carry any scientific basis. Researchers are also strongly recommended to consult a statistician for assistance with statistical analysis to ensure that the study process is adequate to test the statistical hypothesis and describe the results adequately.

Disclosure of Interests

S.P., D.K.L., and J.I. are members of the statistical round of the *KJA*. This letter was reviewed by the statistical round of the *KJA*.

Table 1. Summary of 95 Studies Employing Repeated-measures Analysis of Variance in the *Korean Journal of Anesthesiology* between 2010 and 2014

	Correctly described	Incorrectly described	Undescribed	Others
Normality test	8 (9%)	2 (2%)	83 (87%)	2 (2%)*
Homogeneity of variances test	0 (0%)	0 (0%)	95 (100%)	0 (0%)
Sphericity test	2 (2%)	0 (0%)	93 (98%)	0 (0%)
Interaction between two factors (n = 35) [†]	0 (0%)	0 (0%)	35 (100%)	0 (0%)
Sample size calculation (n = 64) [‡]	0 (0%)	31 (48%)	30 (47%)	3 (5%) [§]
Post-hoc test	56 (59%)	0 (0%)	36 (38%)	3 (3%)
Statistical package	85 (89%)	2 (2%) [¶]	8 (9%)	0 (0%)

*The name of the test is not described. [†]Studies using two-way analysis of variance with repeated-measures in one factor. [‡]Studies using repeated-measures analysis of variance for primary outcomes. [§]Studies using a pilot study or the Dixon's up and down method. ^{||}Results show no statistical significance. [¶]The statistical package version is not described.

ORCID

Sang-Il Park, <http://orcid.org/0000-0002-2026-6848>

Dong Kyu Lee, <http://orcid.org/0000-0002-4068-2363>

Junyong In, <http://orcid.org/0000-0001-7403-4287>

References

1. Lee Y. What repeated measures analysis of variances really tells us. *Korean J Anesthesiol* 2015; 68: 340-5.
2. Wei L, Deng X, Sui J, Wang L, Liu J. Dexmedetomidine improves intubating conditions without muscle relaxants in children after induction with propofol and remifentanyl. *Anesth Analg* 2015; 121: 785-90.
3. Kang H. Sample size determination for repeated measures design using G Power software. *Anesth Pain Med* 2015; 10: 6-15.