Editorial

Appropriate design of research and statistical analyses: observational versus experimental studies

Hyun Kang

Department of Anesthesiology and Pain Medicine, Chung-Ang University College of Medicine, Seoul, Korea

With the recent increases in the number of published articles, following a trend of placing greater emphasis on the importance of research, there is an increasing need for the correct application of statistical methods and for the appropriate choice of valid experimental designs. Scientific research studies can be crudely divided into two general types: experimental studies and observational studies. Both study types have advantages and disadvantages, and the choice of study type depends on factors such as the purpose of the research and the nature of the phenomenon to be evaluated.

Experimental studies have higher internal validity; specifically, when the experiment is repeated under the same experimental conditions, the results will be the same. On the other hand, observational studies may have greater external validity; for example, the results of the study may be applicable to typical clinical practice. Because participants are assigned to control and treatment groups, and the conditions under which the study is conducted and data are collected are controlled by the researcher in experimental studies, factors that are of no interest can be eliminated or controlled. Thus, experimental studies can establish evidence of causation between variables, whereas observational studies can show only associations between variables [1].

Randomization is used for assigning participants to different groups in an experimental study. This process eliminates selection bias and confounding bias, and ensures that the groups are comparable despite the presence of factors other than the one being investigated. As Dr. Oh [2] has pointed out, the randomization process will, on average, evenly balance factors that were measured, were not measured, or could not be measured, and this justifies the statistical analysis [3]. However, randomization does not guarantee that there are no statistically significant differences in terms of baseline characteristics between groups. It only ensures that the differences between control and treatment groups in terms of baseline characteristics are due solely to chance. Accordingly, it must be remembered that even when randomization is executed correctly, baseline characteristics between control and treatment groups may still differ. For example, when simple randomization based on 20 baseline characteristics is used to assign participants to groups, the actual likelihood that at least one characteristic will, by chance alone, show a significant imbalance between the two groups is 64% at a two-sided value of $P < 0.05$ [4]. After a study has been conducted, clinically relevant imbalances should be dealt with by an adjusted analysis of the data. If imbalances considered to be important to the final results are expected, an analysis plan, including an adjusted analysis, can be included when the study is designed.

An observational study examines an existing association between variables based on observations of what is happening or has happened as a result of something else. Nothing is done to influence the results, and the participants are grouped based on their characteristics with respect to the variables and not by randomization. The researcher has no control over the study process or the allocation of participants in an observational study. This can result in bias masking of causality or in false suggestions of correlations.

Despite these limitations, observational studies are commonly used in situations in which experimental studies are inappropriate or impossible. Experimental studies are precluded when they 1)

Corresponding author: Hyun Kang, M.D., Ph.D., Department of Anesthesiology and Pain Medicine, Chung-Ang Universtiy College of Medicine, 224-1, Heuksuk-dong, Dongjak-gu, Seoul 156-756, Korea. Tel: 82-2-6299-2571, Fax: 82-2-6299-2585, E-mail: roman00@naver.com

are unethical; 2) involve rare diseases and patients; 3) include variables that are practically impossible to manipulate, such as inherent traits; or 4) are too costly and time-consuming to be conducted on a large scale. For example, an experimental study comparing the risk for developing lung cancer between smokers and non-smokers would raise ethical concerns, as making subjects smoke in order to assess the impact of smoking on lung cancer would deny participants of the right to make their own decision. Intubation difficulty scores are essentially an inherent trait and cannot be controlled; thus, the study by Seo et al. [5] is an example of the inability to practically manipulate a variable.

Although a poor source of data regarding causality of a treatment or intervention, observational studies can contribute important information, provided the data are analyzed and interpreted appropriately, with consideration of the biases and confounders [6]. To control for confounding arising from a lack of comparability between groups, methods such as matching, stratification, multivariate regression (multiple linear regression, multiple logistic regression), propensity scores, and instrumental variable analysis can be used. Using these methods, the level of one or more factors can be made a constant in order to evaluate the variation in outcome variables derived from a change in the risk factor of interest. These manipulations are referred to as 'statistical adjustments' or 'controlling' for confounding issues. However, these methods can only adjust for or control for known sources of bias under a specific set of assumptions.

Among the statistical adjustment methods, logistic regression analysis is popular and widely used. It is similar to linear regression, but is used for predicting the outcome of a categorical independent variable based on a calculated odds ratio (OR), which is a measure of the association between an exposure and an outcome. The OR reflects the odds of an outcome occurring given a particular exposure compared with the odds of the same outcome occurring in the absence of that exposure. In logistic regression analysis, the regression coefficient ($\beta 1$) of the equation is estimated, and the exponential function of the regression coefficient ($e^{\beta 1}$) is the OR associated with a one-unit increase in the exposure [7]. For categorical variables, ORs can be directly interpreted between groups. However, for continuous variables, ORs can be interpreted differently depending on the unit of the independent variable of interest.

As in other statistical analyses, there are two hypotheses of interest in logistic regression. The null hypothesis (H_0) is that all of the regression coefficients in the equation are zero. The alternate hypothesis (H_1) is that at least one of the regression coefficients in the equation is not zero, which would mean that the model derived from the logistic regression and currently being considered is accurate. In Seo's study, [5] the null hypothesis of the logistic regression analysis was that all of the regression coefficients in the equation predicting difficult

intubation have a value of zero. The alternate hypothesis was that at least one of the regression coefficients in the equation predicting difficult intubation differs significantly from zero, indicating that the model currently proposed to predict difficult intubation is accurate.

When interpreting the results of logistic regression, the absence of multicollinearity among independent variables should be evaluated. Multicollinearity means that two or more independent variables in a multiple regression or multiple logistic regression analysis are in fact highly correlated with each other. In the presence of multicollinearity, it is difficult to determine reliable estimates of individual coefficients, resulting in incorrect conclusions about the relationship between the dependent and independent variables. Thus, when performing multiple logistic regression using independent variables with similar characteristics, researchers should report whether multicollinerity was present, and if so, how it was treated in the statistical analysis. For example, in the study by Seo et al. [5], a discussion regarding multicollinearity and relevant statistical methods would alleviate possible doubts that the total airway score, upper lip bite test, head and neck movement, inter-incisor gap, body mass index, and Mallampati classification have similar characteristics in predicting a difficult airway. Additionally, reporting the overall model evaluation, goodness-of-fit statistics, and validation of predicted probabilities would also help to clarify and support the results.

In summary, experimental studies are considered to be more reliable than observational studies because the process can be controlled and randomization can eliminate bias and ensure comparable study groups in experimental studies. Furthermore, causality can be established in experimental studies. Nevertheless, when experimental studies are inappropriate or impossible, observational studies can provide important information, if the data are analyzed and interpreted using suitable statistical methods. The type of study to be performed should be determined based on the purpose of the study, the nature of the phenomenon, and the characteristics of the variables. The statistical methods should be appropriate for the design and hypothesis of the study, and should be applicable to the types and characteristics of the variables assessed in the study.

Improvements in research methodologies and increased understanding by readers of research articles will increase the debate regarding the correct application of statistics and the selection of appropriate study designs. This phenomenon may have positive ramifications by providing an opportunity to re-think research articles and by raising the quality of papers published in the *Korean Journal of Anesthesiology*. Finally, I thank Dr. Oh for the keen observations and encourage all readers to participate in this necessary debate.

References

1. Gray-Donald K, Kramer MS. Causality inference in observational vs. experimental studies. An empirical comparison. Am J Epidemiol 1988; 127: 885-92.

2. Oh SC. Possible statistical problems of the original paper titled "Predictors of difficult intubation": a bad start. Korean J Anesthesiol 2013; 65: 184-5.

3. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, et al. Randomized clinical trials. Perspectives on some recent ideas. N Engl J Med 1976; 295: 74-80.

4. Burgess DC, Gebski VJ, Keech AC. Baseline data in clinical trials. Med J Aust 2003; 179: 105-7.

5. Seo SH, Lee JG, Yu SB, Kim DS, Ryu SJ, Kim KH. Predictors of difficult intubation defined by the intubation difficulty scale (IDS): predictive value of 7 airway assessment factors. Korean J Anesthesiol 2012; 63: 491-7.

6. Morshed S, Tornetta P 3rd, Bhandari M. Analysis of observational studies: a guide to understanding statistical methods. J Bone Joint Surg Am 2009; 91 Suppl 3: 50-60.

7. Edwards AW. The measure of association in a 2×2 table. J Roy Statist Soc 1963; Series A 126: 109-14.