Avoiding Common Data Analysis Pitfalls in Health Professions Education Research

Jimmie Leppink, PhD, postdoctoral fellow, School of Health Professions Education, Maastricht University; and Kulamakan M. Kulasegaram, PhD, assistant professor, Department of Family Medicine, University of Toronto

Two recently published AM Last Pages present several study design1 and quantitative analysis2 pitfalls in experimental research. In this AM Last Page, we present four common data analysis pitfalls in health professions education research from our perspective as researchers, reviewers, and editors. We use a fictitious study, with three skill test moments for two groups of residents, that shares elements with a wide variety of studies (Figures 1 and 2).

![Mean skill test score (0–50) for two groups of residents](image)

**Pitfall**

1. **Needless hypothesis testing**

   **Explanation and Advice**

   Hypothesis testing makes sense only when you wish to generalize the findings to a larger population. Hypothesis testing may be meaningless when samples are very small or very large, or when the entire population of interest is available.3

   **Advice:**

   - Presenting frequencies, means, standard deviations, ranges, other key descriptive numbers, and when possible a graphical representation (e.g., Figures 1 and 2) is a prerequisite, should occur before hypothesis testing, and is always meaningful to audiences.1
   - Be cautious in interpreting outcomes of statistical hypothesis tests, especially in small samples.3

2. **Reporting means but no standard deviations**

   When reporting means and mean differences to compare groups or conditions, standard deviations are sometimes omitted.

   **Advice:**

   - Standard deviations can be useful for evaluating means in terms of the size of the difference (effect size) and for conducting power analyses or required sample size calculations for subsequent studies that involve hypothesis testing.3
   - Reporting confidence intervals is a supplement to help readers understand the uncertainty around sample estimates.3

3. **Simple effects tests without testing for interaction**

   Studies that compare groups over time test group differences for each time point. This practice increases the likelihood of type I errors (i.e., concluding a difference exists when there is none) in some cases, and the likelihood of type II errors (i.e., concluding no difference when there is one) in other cases.

   **Advice:**

   - Unless you are dealing with time-point-specific comparisons with a study endpoint as the primary outcome, or a purely descriptive study, the appropriate analysis would be an omnibus approach, including group, time, and group-by-time interaction altogether in one analysis prior to post hoc testing at each time point.4
   - Start with an overall test on interaction and perform tests on group differences per time point only when there is a significant interaction (e.g., Figure 2); otherwise, focus on main effects of group and/or time (e.g., Figure 1)—whichever is of interest.4

4. **Testing on individual items that form one scale**

   Some studies report tests on group differences on individual items in a test or questionnaire which, based on appropriate psychometric analysis, can be considered indicators of the same skill or other underlying variable of interest (Figure 2), or of a limited number of constructs (Figure 4). In line with **Pitfall 3,** this increases the likelihood of type I and/or type II errors.

   **Advice:**

   - Perform appropriate psychometric analysis (e.g., factor analysis) to explore which items may be grouped together, and perform tests on group differences per summary for each group of items (Figures 1 and 2).4

**References:**


**Author contact:** Jimmie.Leppink@maastrichtuniversity.nl; Twitter: @JimmieLeppink