PM R 7 (2015) 649-653

www.pmrjournal.org

Statistically Speaking

Dealing With Longitudinal Data

Kristin L. Sainani, PhD

Introduction

In longitudinal studies, researchers measure the same participants at multiple time points. The data are tricky because (1) observations from the same individual are correlated and (2) cross-sectional differences (or "between-subjects effects") are intertwined with longitudinal ones ("within-subjects effects"). These nuances necessitate the use of careful graphics and specialized statistical tests. In practice, researchers often fail to graph the data, leading to errors. They also commonly apply suboptimal tests to the data, or they waste valuable data by ignoring or combining time points.

This article reviews methods for visualizing and analyzing longitudinal data when the outcome is continuous. I will illustrate the principles with a simple longitudinal dataset (n = 41) involving 3 groups of women runners who had their spine bone densities measured at baseline, 1 year later, and 2 years later; this example dataset is based on real data [1] but has been modified for teaching purposes.

First, Graph Your Data

I am wary of any article involving longitudinal data that doesn't include a plot of the outcome variable against time. Formal statistical tests for longitudinal data are easily misinterpreted, but almost everything one needs to know is immediately apparent on the plot. Plots for continuous outcomes can be divided into 2 main types: those that treat time as discrete (mean plots), and those that treat time as continuous.

Mean Plots

Researchers typically measure participants at discrete time points—for example, at baseline as well as 1, 3, and 12 months later—and thus one can calculate the mean value of the outcome variable for each group at each time point. Plotting these means against time reveals longitudinal trends. For example, Figure 1 shows the changes in spine bone mineral density for the 3 groups of women runners; all had menstrual irregularities at baseline. Two groups (black and red lines) had improvements in menstrual function and corresponding increases in bone density, whereas one group (blue line) had no improvement in menstrual function and little change in bone density. For simplicity, I will refer to the groups as red, black, and blue for the remainder of the article.

It is useful to plot both the absolute value of the outcome (Figure 1, left panel) and the percent change since baseline (Figure 1, right panel). Absolute value graphs show between-subjects effects, including baseline differences between the groups. Percent change graphs isolate the within-subjects effects so the groups can be compared independent of any baseline differences.

Mean plots are simple to create and understand. They have two potential drawbacks, however. First, when plotting absolute values, missing data must be imputed; otherwise, the means may appear to increase or decrease simply because people with low or high values drop out over time. Second, mean plots are idealized, because they assume that everyone is measured at exactly the same time intervals; in reality, the timing of follow-up measurements may be variable.

Continuous Time Plots

Plots that treat time as continuous reveal more subtleties in the data. For example, Figure 2 shows runners' percent changes in bone density plotted against their exact measurement times; smoothing lines have been fit to each group's data to highlight the trends. This plot reveals considerable variability in the timing of runners' annual visits (runners had to travel to clinical visits, and some were late in scheduling or attending). The graph also shows individual-level values; for example, the maximum increase in bone density for any runner was 8%.

1934-1482/$ - see front matter
http://dx.doi.org/10.1016/j.pmrj.2015.04.009

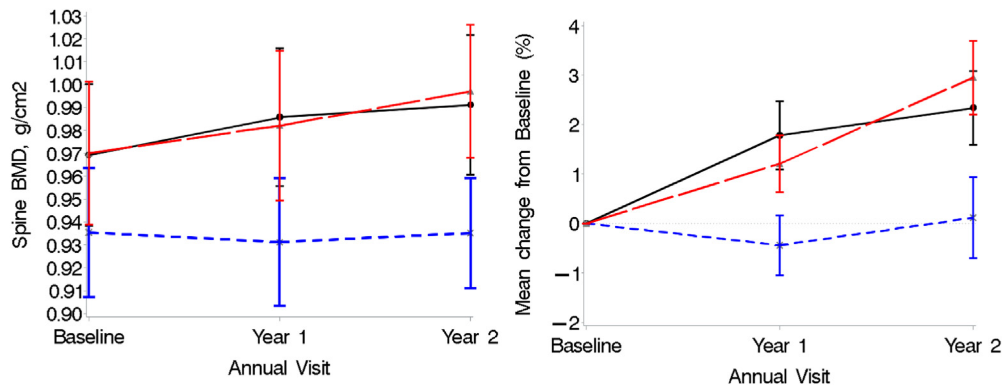

Figure 1. Plot of the mean spine bone mineral density (BMD) in g/cm^2 (left panel) and the mean percent change in bone density since baseline (right panel) versus follow-up time, by group (red, black, or blue). Error bars represent 1 standard error of the mean. The standard errors are greatly reduced in the percent change graph (right panel) because this plot removes between-person variability.

Approaches to Analysis

A wide range of methods are used to analyze longitudinal data. I will review 3 common approaches: data simplification, repeated-measures analysis of variance (ANOVA), and regression methods. Table 1 summarizes the results of applying each method to the example dataset. Note that the regression methods offer a considerable gain in statistical power (the *P* values for the effects are smaller).

In general, regression approaches are optimal; unfortunately, they lag behind in popularity. For example, a 2012 review of longitudinal studies in the anesthesiology literature found that 43% simplified the data, 36% used repeated-measures ANOVA, and only 21% used a regression approach [2].

Approach 1: Ignore or Collapse Data

Researchers are often more comfortable with statistical tests for cross-sectional data—such as 2-sample *t*-tests, ANOVA, and linear regression—than those for longitudinal data. Thus, they may choose to remove the repeated-measures aspect of the data by ignoring or combining time points.

For example, I compared the final bone densities of the 3 runner groups, adjusting for baseline bone density, using linear regression (Table 1). This approach may be appropriate for randomized trials in which one time point was prespecified as primary, but ignoring time points has obvious limitations; after all, why did the researchers bother to collect interim data if they only intended to throw it out?

An alternative approach is to combine all the repeated measurements into one summary measure, such as a slope, average, or area under the curve, and then to use this single measure in further analyses. For example, I calculated a linear regression slope for each woman in my example dataset; the slopes represent annual rates of change. Figure 3 shows the regression line for one woman; her slope (rate of change) was 0.009 g/cm^2 per year. The slopes were higher on average in the red and black groups than in the blue group, but these differences did not reach statistical significance (Table 1). Slope analysis has several merits; in fact, it has similarities to the regression methods (see Approach 3 later in this article). However, collapsing data generally results in a loss of information and statistical power.

Approach 2: Repeated-Measures ANOVA

Repeated-measures ANOVA is a specialized type of ANOVA that accounts for the correlation among repeated observations from the same individual. Despite its popularity, repeated-measures ANOVA has several drawbacks: it gives limited information, it can be tricky to interpret, and it involves restrictive assumptions.

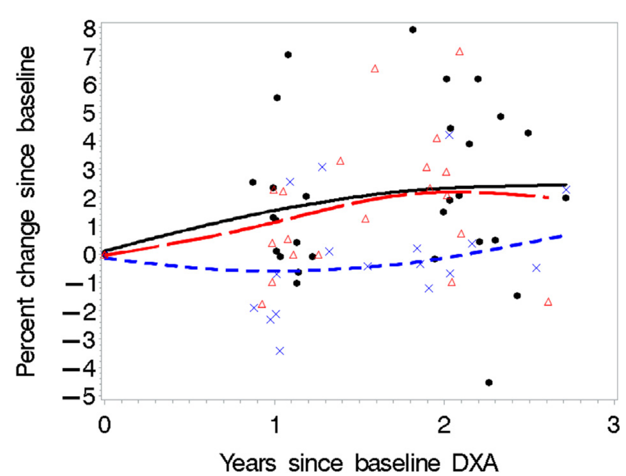

Figure 2. Plot of the percent change in spine bone mineral density versus time, where time is treated as a continuous variable (years since baseline measurement). A smoothing line has been fit to the data from each group to show the trends. DXA = dual-energy x-ray absorptiometry.