

Pretest-Posttest Comparison Group Designs: Analysis and Interpretation

JEFFREY A. GLINER, PH.D., GEORGE A. MORGAN, PH.D., AND ROBERT J. HARMON, M.D.

The pretest-posttest comparison group design is one of the most extensively used methods to evaluate clinical research, but it is often overanalyzed with more than one analysis when one is sufficient. From an example published in this *Journal*, we discuss parametric approaches that are often used to analyze this design and the strengths and limitations of each approach. We then comment on common nonparametric approaches. Last, we discuss methods to analyze this design when the treatment groups are not randomized (intact).

Design Description

The simplest case of the pretest-posttest comparison group design has one treatment group and one comparison group. Prior to the pretest, subjects are *randomly assigned* to groups or conditions. Random assignment is an important feature of the pretest-posttest comparison group design and separates it from nonequivalent (nonrandomized) group designs. Each group is measured prior to the intervention and after the intervention. Typically, one group receives a new treatment and the other group receives a treatment that has been used previously or a placebo. The purpose of this design is to allow the investigator to evaluate the new treatment relative to the previously used treatment. Figure 1, modified from Wood et al. (2001), published in this *Journal*, provides a schematic of the sequence of the pretest-posttest comparison group design.

The design is classified under the heading of mixed design because there are two independent variables. The between-groups independent variable is the treatment, and the within-

subjects or repeated-measures independent variable is change over time from pretest to posttest. Time is a within-subjects independent variable when two or more measures are recorded for each person. While the simplest description of the design has two levels of treatment and two levels of time, as seen in Figure 1, it is not uncommon to have three levels of treatment, such as two treatments and a control group. The number of levels of the between-groups independent variable makes a difference in the type of analysis selected, as does the scale of measurement of the dependent variable.

Analysis With Two Levels of Treatment and Two Levels of Time

Let's start with an example of the study by Wood et al. (2001). The objective of this study was to compare a group therapy

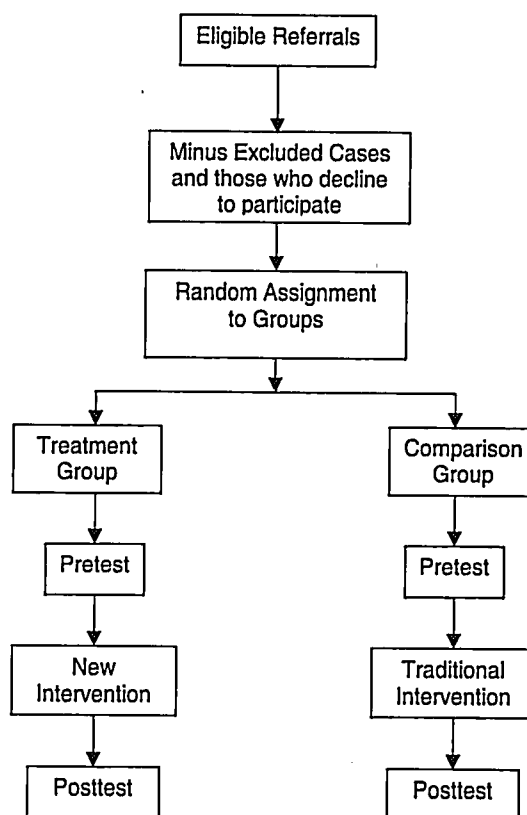


Fig. 1 Schematic of the sequence of the pretest-posttest comparison group design.

Accepted November 8, 2002.

Dr. Gliner is Professor of Occupational Therapy and Education, Colorado State University, Fort Collins. Dr. Morgan is Professor Emeritus of Education and Human Development, Colorado State University, Fort Collins, and Clinical Professor of Psychiatry, University of Colorado School of Medicine, Denver. Dr. Harmon is Professor of Psychiatry and Pediatrics and Director of The Irving Harris Program in Child Development and Infant Mental Health, University of Colorado School of Medicine, Denver.

The authors thank Helena Chmura Kraemer for a helpful critique and Nancy Plummer and Erica Snyder for manuscript preparation. Parts of the column are adapted, with permission from the publisher and the authors, from Gliner JA and Morgan GA (2000), *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis*. Mahwah, NJ: Erlbaum. Permission to reprint or adapt any part of this column must be obtained from Erlbaum.

Reprint requests to Dr. Harmon, CPH Room 2K04, UCHSC Box C268-52, 4200 East Ninth Avenue, Denver, CO 80262.

DOI: 0.1097/01.CHI.0000046809.95464.BE

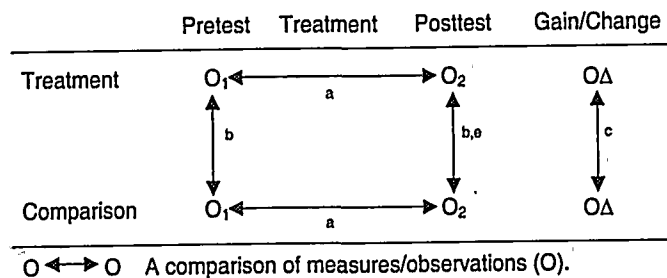
treatment with a routine care treatment "...in adolescents who had deliberately harmed themselves on at least two occasions within a year" (p. 1246). The Wood et al. (2001) study had two levels of the between-groups independent variable, group therapy and routine care. Also, there were two levels of the within-subjects independent variable, a pretest and a posttest 7 months later. Adolescents were randomly assigned to groups. The study fits the criteria for a pretest-posttest comparison group design. The authors used a number of dependent variables; however, for our example, we selected suicidal thinking measured by the Suicidal Ideation Questionnaire. This 30-item questionnaire has scores ranging from 0 to 180, with high scores indicating higher suicidal ideation. We demonstrate how the study could be analyzed using several different statistical procedures, some appropriate and some not appropriate (Fig. 2).

Analysis of Pretest to Posttest Scores Within Each Group. One approach is to compare *within* each group separately from pretest to posttest, and if the treatment condition is significant but the comparison condition is not, then the treatment is assumed successful. Unfortunately, it is not uncommon for both treatment and comparison groups to show gains (with the treatment being statistically significant and the comparison not), even though the difference in gain between the two groups is quite small. Although these results seem to indicate that the treatment worked, the improvement could be due to something else (e.g., maturation), and the results do not show that the new treatment worked better than the comparison treatment. This procedure should not be used to analyze the pretest-posttest comparison group design.

Analysis of Pretest Scores and Posttest Scores Separately. It is not uncommon to see this design analyzed by making one comparison between pretest scores of the two groups and then making a second comparison between posttest scores of the two groups. If there was no difference at the pretest comparison and a sig-

nificant difference at the posttest comparison, then a conclusion is reached that the treatment was successful. The problem with this approach is that the pretest analysis is not needed because participants were randomly assigned to groups. Therefore, a statistical test on the pretest scores of the two groups, such as an independent-samples *t* test, only tests the randomization between the two groups and would be expected to be statistically significant 5% of the time. In addition, adding a pretest analysis to the posttest analysis inflates the type I error above 5%. The posttest analysis between the two groups is all that is needed. A statistically significant difference between the two groups at the posttest in favor of the treatment group would lead to the conclusion that the treatment is successful relative to the control group with this measure. The major problem with this approach is that you cannot take advantage of pretest scores and thus the analysis is less powerful than other analyses.

Gain Score Approach. The gain score approach is the most straightforward approach for the analysis of this design. The gain score approach involves subtracting the pretest scores from the posttest scores within each group. This creates just one independent variable with only two groups or levels, the treatment group and the comparison group. The gain scores become the dependent variable. As we reported in an earlier article in this *Journal*, the proper analysis for this design is an independent-samples *t* test. This tests whether the means of the gain scores for the two groups are equal. However, one should be cautious when using the gain score approach because the reliability of gain scores is often suspect, especially if there is not evidence for strong reliability of the measurement instrument. The gain score approach used by Wood et al. (2001) is described in the statistical analysis section: "...changes from baseline were calculated for the outcomes, and *t* tests for independent samples were used to compare the two arms of the trial" (p. 1248). They found no statistically significant difference between the



- a. Analysis of pretest to posttest within each group. Not recommended.
- b. Analysis of pretest and posttest scores separately. Analyze only the posttest comparison.
- c. Gain score analysis. Usually a good approach.
- d. Mixed analysis of variance (not shown). The interaction *F* provides the same information as the gain score *F*.
- e. ANCOVA (with adjusted posttests) the most powerful approach if assumptions are met. We suggest that this be done using multiple regression.

Fig. 2 Comparison of five approaches to analyzing the pretest-posttest comparison group design. ANCOVA = analysis of covariance.

two groups for suicidal thinking, and they reported their results in the form of confidence intervals for the group difference in mean gain scores (Table 1).

Mixed Analysis of Variance Approach. This is a less common approach to the analysis of the pretest-posttest comparison group design. Because the design is a mixed design, this analysis appears to be the proper analysis. Since there are two independent variables in this design, the analysis yields three different *F* ratios, one for between groups, one for change over time, and an interaction between treatment and time. The only *F* of interest for this design is the treatment by time interaction. It has been demonstrated that the interaction *F* provides identical information to the gain score *t* (or *F* if more than two groups), which, as demonstrated above, is a simpler approach. Therefore, we do not recommend this analysis of the pretest-posttest comparison design.

Analysis of Covariance. This approach, favored by many researchers, is a statistical method used to reduce error variance. When used in the analysis of the pretest-posttest comparison group design, the analysis of covariance (ANCOVA) changes the design from a mixed design to a single-factor design. The ANCOVA makes use of differences in the pretest scores among conditions to reduce error variance by adjusting posttest scores. Once these adjustments have been made to the posttest scores, the analysis is applied only to the posttest scores. Use of ANCOVA in the pretest-posttest comparison group design allows the researcher to use the pretest as the *covariate* and to adjust posttest scores (*variate*) based on a significant linear relationship between the pretest scores (*covariate*) and posttest scores (*variate*). It should be noted that gain scores, instead of posttest scores, could be adjusted using ANCOVA. The rationale behind this approach is that there are usually pretest differences between the treatment and control groups prior to the intervention. Examination of the pretest scores from the Wood et al. (2001) study (Table 1) demonstrates that the pretest scores are higher (or worse) for the group therapy condition by about 5 points prior to the intervention. Thus the ANCOVA approach would adjust that group's posttest scores downward based on the linear regression between pretest and posttest scores.

While the ANCOVA approach is common with the pretest-posttest comparison group design, two assumptions *must* be satisfied. The first is that the relationship between the pretest scores and the posttest scores must be linear. The second assumption is that the regression slopes for each pretest-posttest relationship must be homogeneous (or regression lines must be parallel). This latter assumption is often *not* satisfied in the analysis of the pretest-posttest comparison group design, leading to two problems. First, research is often reported using ANCOVA without satisfying this assumption, making the conclusions invalid. Second, the researcher discovering the violation must reanalyze the data using one of the other approaches

TABLE 1
Change From Baseline to 7 Months on the
Suicidal Ideation Questionnaire

	Baseline			Change From Baseline to 7 Months		
	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD
Group therapy	32	89.1	44.4	28	47.3	50.5
Routine care	28	83.9	51.1	27	39.7	46.7
Mean difference		5.2			7.5	
95% CI of difference		(-19.5 to 29.9)			(-18.8 to 33.9)	

Data from Woods et al., 2001. CI = confidence interval.

mentioned above. A better solution to the problem is to use the ANCOVA approach through multiple linear regression.

Multiple Linear Regression Approach. Since ANCOVA is a special case of multiple regression, it can be performed using multiple regression. This statistical approach, while less common than other approaches used with this design, is a powerful approach (Kraemer, personal communication, October 13, 2002). We will discuss multiple regression in the next article for this *Journal*; however, it is important to present the topic here, at least in a brief description, because of its relevance for this design. For the approach presented here, the posttest scores become the dependent or criterion variable. The independent or predictor variables are the pretest scores (the *covariate*), the groups (treatment and comparison), and the interaction between the pretest scores and the groups. The multiple regression analysis yields tests of significance for each of the predictor variables. This allows the researcher to test both assumptions of the ANCOVA using multiple linear regression. However, if the ANCOVA assumptions are not met, the analysis still appropriately assesses the impact of the treatment on the posttest scores.

Recommendations. Kraemer (personal communication, October 13, 2002) recommended, in order of power from least to most, posttest-only analysis, gain score analysis, and ANCOVA performed by multiple regression. The ANCOVA, if assumptions are satisfied, is considered to be more powerful than the gain score approach because the variability due to error is reduced (Stevens, 1999).

Analysis With More Than Two Levels of the Treatment Variable

The example by Wood et al. (2001) had just two conditions or groups with a pretest and a posttest. However, if there were three or more groups, the three methods recommended above would still apply. For the posttest-only analysis, a single-factor ANOVA (instead of a *t* test) would be applied to the posttest scores of the three groups. Assuming a significant *F* is found, it would be followed by appropriate post hoc tests to identify specific differences. Likewise, for the gain score approach with

three conditions, a single-factor ANOVA would be used instead of the independent-samples *t* test, after gain scores were obtained. Again, since there are more than two conditions, a post hoc test must follow the single-factor ANOVA, if there was a statistically significant overall *F*. For the ANCOVA approach, performed through multiple regression, the variable coding is a little more complex for three or more groups, but the analysis is appropriate.

Analysis With Nonparametric Measures

When the data to be analyzed in the pretest-posttest comparison group design are ordinal (and not normally distributed) or nominal/dichotomous, nonparametric analyses should be undertaken. With ordinal data, a gain score approach could be used. Then, a Mann-Whitney *U* would be applied if there are just two conditions, or a Kruskal-Wallis test would be used for more than two conditions. ANCOVA cannot be used in this situation.

Often when clinical importance is being considered, posttest data are dichotomized based on a clinically relevant cut point and then a statistical analysis is performed. It is recommended that if *continuous* data are to be dichotomized for clinical relevance, then effect size indices such as number needed to treat or absolute risk ratio be reported *without* significance testing.

Nonequivalent (Intact) Group Designs With a Pretest and Posttest

An essential feature of the pretest-posttest comparison group design is random assignment of participants to groups. When this feature cannot be accomplished (e.g., using different hos-

pitals or classrooms as intact groups), the design is referred to as a nonequivalent group design with a pretest and posttest. With this design, the ANCOVA *should not* be applied because the population means on the covariate cannot be assumed to be equal (Huck, 2000). Thus the posttest-adjusted means could be biased. Stevens (1999) pointed out, "The fact is that inferring cause-effect from intact groups is treacherous, regardless of the type of statistical analysis. Therefore, the task is to do the best we can and exercise considerable caution..." (p. 324).

Conclusion

In conclusion, many different statistical approaches are used to analyze the pretest-posttest comparison group design. The ANCOVA using multiple linear regression appears to be the best statistical approach because, if the assumptions are satisfied, it is the most powerful analysis. If the ANCOVA assumptions are not satisfied, the analysis can still be used and the researcher does not have to reanalyze the data. We urge caution of any interpretation resulting from analysis of this design when participants have not been randomly assigned to groups (intact groups). We also advise against the statistical analysis of data that have been dichotomized artificially, but suggest descriptive indices such as effect sizes be reported.

REFERENCES

- Huck SW (2000), *Reading Statistics and Research*, 3rd ed. New York: Longman
 Stevens JP (1999), *Intermediate Statistics: A Modern Approach*, 2nd ed. Mahwah, NJ: Erlbaum
 Wood A, Trainor G, Rothwell JU, Moore A, Harrington R (2001), Randomized trial of group therapy for repeated deliberate self-harm in adolescents. *J Am Acad Child Adolesc Psychiatry* 40:1246-1253

The next article in this series appears in the June 2003 issue:

Use and Interpretation of Multiple Regression

Nancy L. Leech, Jeffrey A. Gliner, George A. Morgan, Robert J. Harmon

Women's Health After Pregnancy and Child Outcomes at Age 3 Years: A Prospective Cohort Study. Robert S. Kahn, MD, MPH, Barry Zuckerman, MD, Howard Bauchner, MD, Charles J. Homer, MD, Paul H. Wise, MD

Objectives: This study examined the persistence and comorbidity of women's physical and mental health conditions after pregnancy and the association of these conditions with child outcomes. *Methods:* A national cohort of women who recently gave birth were surveyed in 1988 and again in 1991. We examined longitudinal data on maternal poor physical health, depressive symptoms, and smoking, and maternal report of child outcomes (at age -3 years). *Results:* Women's poor physical health and smoking had strong, graded associations with children's physical health and behavior problems, whereas women's depressive symptoms were associated with children's delayed language and behavior problems. *Conclusions:* Substantial persistence and comorbidity of women's health conditions exist after pregnancy with adverse effects on early child outcomes. Child health professionals should support services and policies that promote women's health outside the context of pregnancy. *Am J Public Health* 2002;92:1312-1318. Copyright 2002 by the American Public Health Association.

A vertical bar on the left side of the page, consisting of a series of colored squares (red, orange, yellow, green, blue) stacked vertically, with a small red diamond at the top.

COPYRIGHT INFORMATION

TITLE: Pretest-Posttest Comparison Group Designs: Analysis and Interpretation

SOURCE: J Am Acad Child Adolesc Psychiatry 42 no4 Ap 2003

WN: 0309102114017

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited.

Copyright 1982-2003 The H.W. Wilson Company. All rights reserved.