Change Scores vs. Analysis of Covariance

A common research situation is comparing two groups using an interval outcome measure (eg, blood pressure, cholesterol, quality of life, pain score, test score, etc). Often, there is a baseline measurement and the research question is: Is the change in one group different from the change in the other group? There are two main approaches to analyzing these data: using change scores or using analysis of covariance.

Y = outcome measure; Y_1: baseline measurement; Y_2: follow–up measurement
Change from time 1 to time 2:

- Y_2–Y_1 or treatment group: ΔT
- Y_2–Y_1 for control group: ΔC

A potential problem with ***change scores*** is illustrated below:

By chance, time 1 average scores for treatment group might be lower than for control group. If so, then expect—due to regression to the mean—that the average scores in the treatment group will rise (toward the mean) and the average scores in the control group will fall (toward the mean), even though there is truly no treatment effect.

$\Delta T = +$ and $\Delta C = -$

Study results:

Average (ΔT) – Average (ΔC) = (+) – (–) = ++

This is a 2–independent–sample t–test comparing the mean difference in one group with the mean difference in the other group. The null hypothesis is that the *difference of differences* is zero. As noted above, due to chance and regression to the mean, the t–test can be erroneous.

The alternative strategy is an ***analysis of covariance*** (ANCOVA).

ANOVA: Y = X

Y is a continuous outcome measure, and
X indicates two or more groups to be compared (categorical variable)

ANCOVA: $Y_2 = X + Y_1$

Y_2 is a continuous variable measured at follow–up, and

X indicates two or more groups to be compared, and

Y_1 is the baseline measurement of Y

ANCOVA: Assessing change over time—3 steps:

First, stratify on the baseline score (which is a potential confounder):

Three strata:

Lower than average

Average

Higher than average

Second, calculate change scores within each strata. Again, assume that regression to the mean will cause 1) lower than average scores at baseline to increase at follow–up, and 2) higher than average scores at baseline will decrease at follow–up. (We are also assuming there is no true treatment effect.)

Strata	ΔT	ΔC
Lower	+	+
Average	0	0
Higher	–	–

Third, calculate study results within each stratum and then average across strata.

Strata	ΔT	ΔC	**ΔT – ΔC**
Lower	+	+	**0**
Average	**0**	**0**	**0**
Higher	**–**	**–**	**0**

Average (ΔT – ΔC) across strata: 0

Conclusions:

1. Both analysis methods (t–test of change scores and ANCOVA) address the same research question.
2. Both methods adjust for differences in baseline scores.
3. ANCOVA does a better job of dealing with possible regression to the mean (and similar problems).
4. Neither method addresses the important issue of interaction: Is there a different treatment effect among subjects with low and high baseline scores?