Dealing With Change: Using the Conditional Change Model for Clinical Research

Mikel Aickin, PhD

Introduction

Virtually all clinical medicine is about change. The criteria for deciding whether a therapy has been successful nearly always include consideration of the degree to which the patient's initial condition has improved or to which a deteriorating condition has been stabilized. Both criteria depend on change. In the first case it is a rise in some measurement of benefit or drop in some measurement of burden, whereas in the second it is that a downward change has been prevented.

In clinical research, therefore, one of the most frequently used approaches is to compare changes in a treated group with corresponding changes in a control group. Perhaps the most notable pedagogic failing of statistics courses and textbooks is that they do not present the appropriate way to analyze data coming from this design, which explains why published analyses are so often suboptimal, if not actually incorrect. The purposes of this article are to explain what should be the default method of analyzing change data and to indicate how to compute and display the results graphically.

Regression to the Mean

One of the earliest observations, by Sir Frances Galton (1822-1922), was the tendency of change scores to be negatively related to baseline values. In fact, the regression procedure got its name from this phenomenon, which Galton called "regression to the mean." The notion that regression to the mean was a real biologic phenomenon supported the early 20th-century eugenics movement, especially in Great Britain. The great statistician RA Fisher (1890-1962) was an ardent participant in that movement. Although it is possible for there to be true regression to the mean, in most cases the phenomenon is artifactual. It arises from the fact that if, over time, a biologic quantity must remain in a certain range, to ensure survival of the organism, then it is automatic that high values at one time will tend to be followed by smaller values a little later on, and conversely low values at one time will tend to be followed by larger values a little later on. Thus, regression to the mean is an inevitable consequence of a time sequence of measurements needing to stay in some viable range. The fact that it is not a biologic effect but only a statistical one does not diminish its influence when one is looking at change.

Table 1. Glossary of terms and symbols

Term	Meaning
y_0	Baseline measurement
y_1	Endpoint measurement
x	Indicator variable (treatment)
$x = 0$	Control group
$x = 1$	Treated group
β	Study parameter
β_0	Mean change (control)
$\beta_0 + \beta_1$	Mean change (treated)
β_1	Difference (treatment effect)
e	Residual (patient-specific)
null	No treatment effect

The Conditional Change Model

The simple change design can be described as follows (Table 1): Each patient yields a measurement, y_0, at the start of the study, the so-called *baseline* measurement. At the end of the study, each produces another measurement on the same scale, y_1, the *endpoint* measurement. There is a further treatment variable x (such as a drug) that takes the value 0 for each patient in the control group and the value 1 for each patient in the treated group. It is called an *indicator* because it points to the treated patients. The purpose of the study is to compare pre-post changes $y_1 - y_0$ between the treated ($x = 1$) and control ($x = 0$) groups.

The most common advice in statistics texts is that this comparison should be made by applying a two-

Mikel Aickin, PhD, is a Research Professor in the Department of Family and Community Medicine at the University of Arizona in Tucson, AZ. E-mail: maickin@earthlink.net.

sample *t*-test to the change scores. The *null hypothesis* is that the mean changes in the two groups are equal. Although it is not particularly well-known, one can carry out the two-sample *t*-test with a linear regression program. Figure 1 shows the mathematical model for a single patient.

$y_1 - y_0$	=	$\beta_0 + \beta_1 x$	+	e
Change		**Parameters**		**Residual**
(endpoint – baseline)		(control and treated groups)		(patient-specific)

Figure 1. Mathematical model for a single patient.

I do not intend to go into the computation or theory behind this model—only to use it as a convenient language for thinking about the analysis. The y and x variables I have already defined. The β's are "parameters," which just means that they are imagined to be constant throughout any given study. The whole point of the model is to provide a way of interpreting and estimating these parameters. The *e* term represents a patient-specific variable, which accounts for the fact that change (the left side of the equation) is more than just a simple linear function of the treatment indicator. The left side of a regression equation is always thought of as the outcome, the result of some process, whereas the right side provides a mathematical explanation for the result. In textbooks the result is always written as just y, but here we want to think of change as the outcome, so the result is a difference: endpoint (y_1) minus baseline (y_0).

The model equation says that there is one mean change in the control group (β_0) and another in the treated group ($\beta_0 + \beta_1$). These can be deduced by the simple but universal technique of substituting the possible values of x on the right side of the equation. Thus, β_1 acquires its interpretation as the difference between the two mean changes, which is the whole point of the study.

In a remarkably useful but almost completely unknown book, Ian Plewis[1] argued strongly that in studies of change, one should include the baseline value in the analysis. The way to do this is to extend the above analytic model (Figure 2). The only difference between the two analyses is that y_0 appears on the right side of the equation in Figure 2. The model above thus says that changes (left side) are influenced by treatment (x) and baseline value (y_0).

$y_1 - y_0$	=	$\beta_0 + \beta_1 x$	+	$\beta_2 y_0$	+	e
Change		**Parameters**		**Baseline**		**Residual**
(endpoint – baseline)		(control and treated groups)		(measured value)		(patient-specific)

Figure 2. Analytic model extended to include baseline.

Analysis Steps

As a practical matter, the dataset for such a study looks like Table 2. The steps involved in the analysis are as follows:

• Compute the change as a new variable
• Subtract the mean baseline from the baseline values (centering)
• Regress the change on the treatment indicator and the centered baseline to produce Table 3.

Table 3 is a simplified version of what most statistical programs provide. The names of the right-side variables appear first, followed by effect estimates of their corresponding β's (under the "coefficient" column). The other two commonly displayed values are the standard deviation of the sampling distribution of the coefficient estimate (under the "SDE" column; most programs erroneously use SE, for "standard error," for what I call SDE, for "standard deviation of the estimate), and a p-value for testing the null hypothesis that the coefficient is in reality zero (so that the named variable would not appear in the true model equation).

Table 2. Example form of a dataset for studying change

Baseline	Endpoint	Treatment
140	138	0
141	123	0
157	140	1
138	136	1

Baseline is the value measured at the start of the study, and endpoint is measured at the end. The treatment variable is 0 for a control and 1 for a treated patient.

Table 3. Example regression computer output

Variable	Coefficient	SDE	p
Treatment	–4.122	2.178	0.058
Baseline	–.801	0.213	0.000
Intercept	1.531	1.260	0.224

Variables are the explanatory (right-side) variables in the regression equation. Coefficients are their estimated effects (β's), SDEs are their sampling standard deviations, and p are the p-values for the corresponding null hypothesis tests. The intercept term corresponds to β_0.

Baseline Differences

We can interpret the coefficient estimates by referring back to the model equation. To interpret β_0, plug the values $x = 0$ and $y_0 = 0$ into the right side. Of course $x = 0$ means "in the control group." Because we centered the baseline values, $y_0 = 0$ means "at the mean on baseline." Thus, β_0 stands for the mean change in a control patient who was exactly average at baseline. By substituting $x = 1$ and leaving $y_0 = 0$, we interpret $\beta_0 + \beta_1$ to be the mean change among treated patients who are exactly average at baseline. The β_2 parameter captures how much the mean changes differ, if we compare two patients in the same group but who differ by 1 unit at baseline. Usually β_2 is not of interest, so it is included in Table 3 only for completeness.

Single Regression Line

Because it will play a role in the following discussion, we can note that to interpret β_1 we could have substituted any fixed value for y_0. That is, β_1 represents the effect of treatment on mean change when we compare any two patients in different groups but who had exactly the same baseline value. It is in this sense that we say the analysis has been "adjusted for baseline." This is connected to the fact that if we were to graph the fitted regression equations in the two groups, like this

$$y_1 - y_0 = \begin{cases} \beta_0 + \beta_2 y_0 & (controls) \\ (\beta_0 + \beta_1) + \beta_2 y_0 & (treated) \end{cases}$$

then they would appear as parallel lines. The vertical distance between the two lines is always the same and is equal to the estimated β_1. Thus, β_1 captures the effect of treatment for any group of patients who have the same baseline values.

Three Advantages: Smaller Error, Similar Groups, Less Artifact

There are three primary reasons for preferring the conditional change model over the *t*-test (Table 4). The first is purely statistical: that the SDE of the effect of interest (β_1) is nearly always smaller under the conditional change model. This means that the estimate of treatment effect is more precise, and it has implications for the chance of detecting a real effect using null hypothesis testing.

The second reason is that if there is an imbalance between the control and treatment groups with respect to baseline values, this undermines the whole logic of

> **If the relationship between baseline and change is different in the two groups, then does it not follow that this might be a consequence of treatment?**

the study, in that the comparison of treatment versus control is not made across two "similar" groups. Statisticians frequently claim that randomization removes this concern, but this is an argument based on large-sample theory, which does not apply to small studies and may apply inadequately to most studies. Thus the conditional change model is seen as an attempt to lessen, if not remove, baseline differences.

The third reason is related to both of the first two. Although it is possible for there to be true regression to the mean, in most cases the phenomenon is artifactual. Thus, the final argument for the conditional change model is that it tends to reduce the artifactual effect of regression to the mean.

The Unavoidable Warnings

Outliers

Although the conditional change model should probably become the default for analyzing change, it is not without its difficulties. First, badly outlying values on the baseline measurement can cause serious damage to the estimate of the treatment effect because regression is sensitive to outliers. This is not, incidentally, an argument in favor of the *t*-test, because it is also unduly affected by outliers. Thus, it is always wise to view the baseline distributions graphically and perhaps to take some evasive action. One method, formerly widely used and now nearly abandoned, is Winsorization. One picks the largest and smallest reasonable values, and then those above the largest value are rounded down to it and those below the smallest value are rounded up to it.

Differing Regression Lines

The second potential problem is that if one separately fitted regression lines (of change, on baseline) within the two treatment groups, one might get quite different lines. In this case, the rationale for fitting a single line to both groups, which is part of the conditional change model, might seem unjustified. There is actually a deeper issue here than just model fitting. If the relationship between baseline and change is different in the two groups, then does it not follow that this might be a consequence of treatment? For example, suppose that in the control group there is the usual negative relationship between change and baseline that is predicted by regression to the mean, but there is no relationship in the treated group. Part of the effect of treatment may then have been to detach changes from baseline values.

Regardless of the source, differing regression lines

pose a real conceptual problem, which can be seen in terms of the model:

$$y_1 - y_0 = \beta_0 + \beta_1 x + (\beta_2 + \beta_3 x)y_0 + e$$

Here the effect of baseline explicitly depends on which group the patient is in. This is accomplished in the analysis by adding an interaction term, xy_0, to the list of explanatory variables. Now if we consider two patients, one treated and one control with identical values of y_0, the difference between their mean changes is $\beta_1 + \beta_3 y_0$. (This is again deduced by substituting trial values on the right side of the equation.) This says that the treatment effect depends on the baseline value, so that there is no single well-defined treatment effect. The statistical way around this is to center the baseline, as I have recommended, so that the nominal treatment effect on the computer output (β_1) is the effect of treatment at the mean of the baseline. This is a reasonable convention, but it does not solve the conceptual problem. Although one might be annoyed

at the conditional change model for raising such an issue, if there are differential treatment effects then it would seem important to report them, which would never happen with the *t*-test approach.

Three Measurement Points

The third difficulty only arises if one extends the conditional change model beyond the comparison of baseline and endpoint values. For example, imagine a design in which the outcome variable is measured three times on each patient: baseline, midstudy, and endstudy. To measure the early effect, the change from baseline to midstudy might be subjected to conditional change analysis. One might then go on to investigate late changes by applying the same model to midstudy and endstudy values, with midstudy values now playing the role of baseline. The reason this can go awry can be seen in terms of regression to the mean. If the treatment produces an early beneficial effect, then midstudy values will be higher in the treated patients than in the control patients. The midstudy values will not be com-

Adjusting Data: Graphical Display of Conditional Change Results

After having presented a statistical analysis, researchers may want to show the data graphically, to give a richer impression of the results. The danger here is that the conditional change anaysis adjusts for baseline but the graphed data are unadjusted. The alert reader may see from the graphs that the asserted treatment effects are implausible, undermining the credibility of the presentation.

The solution to this problem is to adjust the data before graphing. The procedure involves two steps, adjusting the baseline and adjusting the changes. To adjust the baseline:

- Fit the regression model $y_0 = \alpha_0 + \alpha_1 x + e$
- Compute the fitted values ($\alpha_0 + \alpha_1 x$ where the parameters (α's) are estimated)
- Compute the regression residuals (observed minus fitted). All programs will do this
- Subtract $\alpha_1 x$ from each fitted value, and add $\alpha_1 x_0$ (parameters are estimated, x_0, determined as below)
- Add the residuals to the values just computed. These are the adjusted baselines.

Although it may have gotten lost in the recipe, the values computed are $\alpha_0 + \alpha_1 x_0 + e$, where the parameters are estimated and e is the computed residual. These represent values that would have been observed if each patient

had had x_0 in place of his or her treatment indicator. Often $x_0 = \frac{1}{2}$ will be sensible, but one can sometimes argue for other values, such as the mean of x (that is, the treated fraction).

The process is similar for the adjusted changes:

- Fit the regression model $y_1 - y_0 = \beta_0 + \beta_1 x + \beta_2 y_0 + e$
- Compute the fitted values of change
- Compute the regression residuals
- Subtract $\beta_2 y_0$ from each fitted value, and then add $\beta_2 y_0$
- Add the residuals to the values just computed. These are the adjusted changes.

If the baseline values were centered, then their mean y_0 will be zero. However, one can choose some other value for this constant (such as the median, for example).

These two procedures will give baseline data adjusted for treatment, as if everyone had treatment value x_0, and changes adjusted for baseline, as if everyone had baseline value y_0 (or whatever baseline constant was chosen). Thus, for the adjusted data, baseline will have the same mean in the two treatment groups, and change will correspond to the effects obtained from the conditional change model.

Adjusted endstudy values are computed by adding adjusted changes to adjusted baselines, and are adjusted both for baseline treatment group imbalance and the consequences of this imbalance on change.

It must be emphasized that the adjusted data are not to be used for statistical inference. Their only purpose is to make it possible to show tables of summary statistics or graphics, with the above effects adjusted for. The reason is that in general, adjusted data have less variability than the original data (this is an inevitable consequence of removing variability due to the adjustment variable), and so effect estimates based on adjusted data will be artificially precise. It may be worth mentioning that the conditional change analysis easily lends itself to adjustment for other variables that might be thought to influence the outcome. Gender, age, and comorbidities are logical contenders. Data can be adjusted for these as well. The only part that becomes more complex is that the fourth step (as given earlier) must be done for each adjustment variable. One must then be careful to report the constant value substituted for each variable, to avoid misinterpretation of the adjusted tables or graphics.

Table 4. Differences between Conditional Change model and *t*-test

Conditional Change model	*t*-Test
Primary reasons:	
1. Less artifact of regression to the mean	1. Artifact of regression to the mean
2. Lessen baseline differences of two groups	2. Subject to baseline differences between two groups
3. Lower SDE (more precise)	3. Higher SDE
Cautions:	
1. Baseline outliers producing bias	1. Baseline outliers producing variability
2. Assumes single regression line for both groups (should be investigated)	2. Assumes no regression line in either group
3. Dangerous to use mid-study value as baseline for late change	3. Same issue

SDE = standard deviation of the estimate.

parable between the groups, and for an understandable reason. The second (midstudy → endstudy) analysis will thus try to remove real effects, some of which are just regression to the mean (the high values in the treated group tending to drop) and could actually result in the treatment appearing to do worse in this second analysis, even though continued treatment is continuing to benefit the patients. The conclusion is that for the analysis of change, one should do the baseline → midstudy and baseline → endstudy analyses, because the conditional change approach is equally valid in these cases and questionable in the midstudy → endstudy analysis.

Conclusion

There is an ethical principle in biostatistics, which says that the most powerful appropriate analysis should be used in evaluating the results of biomedical studies. This emerges not just from general scientific principles but also out of respect for the human participants who allowed themselves to be used in a medical experiment. For a rather long time now, this ethical principle has not been followed as well as one would like, in studies where the issue is to compare mean changes between two treatment groups, one of the most common of all clinical trial designs.

The benefits of the conditional change approach were clearly put forward by Plewis more than twenty years ago. The failure to heed Plewis's advice may stem from the virtual absence of supportive articles, even in the statistics literature. Indeed, an otherwise excellent statistical article[2] on the topic made misstatements about the conditional change model that were rectified two years later, but only in a letter to the editor.[3] It is understandable that a vacuum in the literature would open the way to statistics textbooks' continuing to promote a weak analysis. The purpose of this article has been to try to change the situation. ❖

References
1. Plewis I. Analysing change: measurement and explanation using longitudinal data. Chichester, UK: John Wiley & Sons; 1985.
2. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for designs. Stat Med 1992 Sep 30;11(13):1685–704.
3. Senn S. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1994 Jan 30;13(2):197–8.

A Peculiar and Perpetual Error

It is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than by negatives, whereas it ought duly and regularly to be impartial; nay, in establishing any true axiom, the negative instance is the most powerful.

— Aphorisms, *Sir Francis Bacon, 1561-1626, English philosopher, statesman, scientist, lawyer, jurist, and author*