Compliance as an Explanatory Variable in Clinical Trials

B. EFRON and D. FELDMAN*

The Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) measured the effectiveness of the drug cholestyramine for lowering cholesterol levels. The patients in the study were measured for compliance (the proportion of the intended dose actually taken) and for cholesterol decrease. The compliance-response regression for the Treatment group shows a smooth increasing effect of the drug in cholesterol level with increasing compliance. However, a similar, though less dramatic, compliance-response regression is seen in the Control group. This article investigates the recovery of the true dose-response curve from the Treatment and Control compliance-response curves. A simple model is proposed, analyzed, and applied to the LRC-CPPT data. Under this model, part but not all of the true dose-response curve can be estimated.

1. INTRODUCTION

Figure 1 shows results from the Stanford portion of the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT). This was a placebo-controlled double-blind randomized clinical trial concerning the efficacy of cholestyramine for lowering cholesterol level and thereby reducing coronary heart disease. Lipid Research Clinics Program (1984) describes the complete study.

The left panel depicts the Treatment group, comprising 164 men after removal of the one indicated outlier. The vertical axis is the decrease in total cholesterol level, while the horizontal axis indicates compliance (the proportion of the nominal cholestyramine dose actually taken). A "dose-response" relationship is quite evident: better compliance leads to a greater decrease in cholesterol level, as indicated by the quadratic regression curve. (More careful definitions of compliance and cholesterol differences are given in Section 3.)

The right panel of Figure 1 shows the same plot for the placebo Control group, comprising 171 men after removal of one outlier. Compliance here has been adjusted to match the compliance distribution in the Treatment group, as discussed in Section 3. Interestingly enough, there is a significant dose-response relation between compliance and cholesterol decrease in the placebo Control group, though of smaller magnitude than in the Treatment group.

It seems unfortunate that compliance was so variable in this experiment, but a good argument can be made that the situation is actually more favorable to the investigator than if 100% compliance had been enforced. Figure 1 seems to indicate a nice dose-response relationship, which would not be visible if compliance had been perfect.

The difficulty here of course is that compliance (and hence dose) has not been assigned in a randomized fashion by the investigators, as it would have been in a genuine dose-response experiment. Compliance is an uncontrolled covariate, and it may be that better compliers are better patients to begin with. In fact, the positive dose-response relationship for the Control group indicates that this is certainly the case.

This article concerns the proper interpretation of the data in Figure 1. What would we have seen if 100% compliance had been enforced? To what degree can we estimate the actual dose-response relationship? Section 2 discusses a simple model, which is applied to the LRC-CPPT data in Section 3. Some final remarks are made in Section 4.

Because of the special nature of compliance, the questions considered here are not standard problems of covariate adjustment, as discussed for example in Koch et al (1982). First of all compliance is an adjustable covariate, unlike, say, age. What we would see if 100% compliance were enforced is a different question than what we would see if all patients were 60 years old. The former question concerns *changing* everyone's compliance to 100%, while the latter presumably refers to considering only the population of 60-year-olds.

Second, unlike other adjustable covariates (such as exercise level), compliance has a different meaning in the Treatment of Control groups. Compliance determines the amount of active drug taken for Treatment group patients and also indicates something about the patient's psychological status. In the Control group, only the psychological component of compliance applies.

Third, the Treatment and Control groups should behave identically at the 0% end of the compliance scale, since there is no distinction between 0% compliance with an active drug or with a placebo. This seems to be the case in Figure 1, and is in fact the case for experiment LRC-CPPT, as standard hypothesis tests will confirm in Section 4.

The results of the LRC-CPPT focused considerable attention on compliance as an important explanatory variable in clinical trials. Cholestyramine's package labelling now includes a compliance-based dose-response curve; see Urquhart and Chevalley (1988). A recently developed electronic monitor for pill bottles makes compliance data of quality superior to that from LRC-CPPT (which was based on packet counts) available for most clinical trials. See Chevalley and Urquhart (1987).

The goal of this article, beyond analysis of the LRC-CPPT

* B. Efron is Max H. Stein Professor of Humanities and Sciences and Professor of Statistics and Biostatistics, Department of Statistics, Stanford University, Stanford, CA 94305. D. Feldman is an M.D. in Palo Alto, CA 94303. The authors are grateful to the editors and referees for an unusually helpful critique of the original manuscript, and to Dr. John Urquhart for originally suggesting the problem studied here.

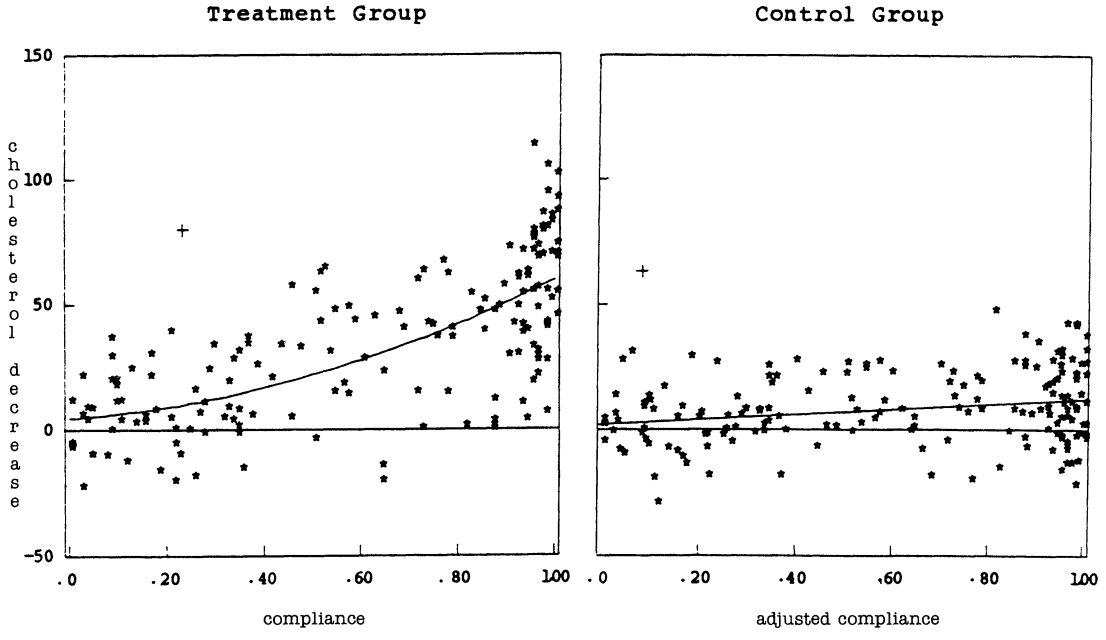

Figure 1. Stanford portion of LRC-CPPT. Left Panel: Treatment group, 164 men (after removal of outlier indicated by +); vertical axis is decrease in total cholesterol; horizontal axis is compliance (the proportion of the nominal choelstyramine dose actually taken). Better compliance leads to larger decreases in total cholesterol, as indicated by the quadratic regression curve. Right Panel: Placebo Control group, 171 men (after removal of outlier +); compliance has been adjusted to match the distribution of compliance in the Treatment group. There is a smaller, but still significant, dose-response relationship between compliance and cholesterol decrease, indicated by the linear regression line.

data, is to establish a moderately general framework for making use of compliance data in clinical trials. The theory presented here is intended to supplement rather than to replace the more conservative intent-to-treat analysis, which in its simplest form is just a two-sample test for equality between the Treatment and Control groups, ignoring covariate information. The LRC-CPPT results are too clear-cut to make such a test interesting, but in more ambiguous situations it can, because of its assumption-free basis, play a decisive role.

2. DOSE–RESPONSE AND COMPLIANCE–RESPONSE

We are interested in recovering the true dose-response curve from the compliance-response curves shown in Figure 1. This section discusses a simple model relating dose, compliance, and response. We will see that it is impossible to recover the full dose-response curve in general, but that we can always estimate certain important aspects of it. In the special situation where the true dose-response curve is linear, it can be estimated in its entirety.

Our theory addresses certain hypothetical questions, such as what results would have been seen if all patients complied perfectly. We will use Rubin's theory of causality, as nicely described in Holland (1986), to phrase such questions and their answers clearly. See also Holland (1988). Rubin's theory begins with a population U of possible patients, u representing an individual member of U. The set of patients actually enrolled in the study, those represented in Figure 1, is obtained by some sampling scheme from U which need not be specified for the purposes here.

Let $Y_0(u)$ represent patient u's placebo-response, in our case the cholesterol decrease experienced by patient u if zero amount of the drug is taken. $Y_0(u)$ is defined for all patients u in U, but only observed for patients in the Control group of the study, or for those in the Treatment group who had zero compliance. Let $z(u)$ be patient u's *compliance*, measured as the proportion of nominal dose actually taken (the horizontal axis in Figure 1); $z(u)$ is observed for all patients in the study. Both $Y_0(u)$ and $z(u)$ are considered to be inherent properties of the patients, what Holland calls "attributes," which may or may not be observed depending on the sampling and treatment assignment schemes employed. (See Remark B.)

We can imagine giving each patient u every possible dose X of the active drug. Here X is measured in the same units as z, as the proportion of nominal dose, so $0 \le X \le 1$. The model we will use relates $Y_X(u)$, the response (cholestral decrease) of patient u, to the amount X of the active drug and the placebo response $Y_0(u)$, as follows:

$$Y_X(u) = G_X + (1 + H_X)Y_0(u) + e_X(u). \tag{2.1}$$

G_X and H_X are continuous functions of X, with

$$G_0 = H_0 = 0; \tag{2.2}$$

$e_X(u)$ is a disturbance term satisfying

$$e_0(u) = 0 \quad \text{and} \quad E\{e_X \mid z\} = 0. \tag{2.3}$$

Model (2.1)–(2.3) is further discussed later (see Remark A).

The true dose-response curve $\delta(X)$ is defined as the average net effect on response of giving all patients amount

X of the active drug rather than amount zero,

$$\delta(X) \equiv E\{Y_X - Y_0\}. \tag{2.4}$$

Expectations and conditional expectations are defined as the appropriate averages over U, so $E\{Y_X - Y_0\} = \Sigma_{u\in U}(Y_X(u) - Y_0(u))/N$, N being the number of elements in U. Expression (2.4) is what Rubin and Holland call the *average causal effect* of X on response. From (2.1)–(2.3) we calculate

$$\delta(X) = G_X + H_X C_0, \tag{2.5}$$

where

$$C_0 \equiv E\{Y_0\}, \tag{2.6}$$

the average placebo-response over all of U.

In a true dose-response experiment, the statistician gets to observe $Y_X(u)$ for different values of X on randomly selected patients u. This makes estimation of $\delta(X)$ straightforward. The situation is more complicated here. Two more definitions are necessary to define what is actually being observed in Figure 1.

Let $s(u)$ indicate patient u's group assignment,

$$\begin{aligned} s(u) &= 0 \quad \text{if} \quad u \quad \text{assigned to Control group} \\ &= 1 \quad \text{if} \quad u \quad \text{assigned to Treatment group.} \end{aligned} \tag{2.7}$$

In a completely randomized experiment like LRC-CPPT, the statistician determines $s(u)$ by independent flips of a fair coin. [In practice $s(u)$ is determined only for patients u selected into the study, but we can think of it as defined over all of U.] Also let $x(u)$ be the amount of active drug actually taken by patient u, so

$$x(u) = s(u) \cdot z(u); \tag{2.8}$$

$x(u) = 0$ in the Control group, and $x(u) = z(u)$ in the Treatment group.

The observed response for patient u is

$$y(u) = G_{x(u)} + (1 + H_{x(u)})Y_0(u) + e_{x(u)}(u) \tag{2.9}$$

according to (2.1), (2.8). This is the quantity plotted along the vertical axis in Figure 1. Let $y_C(u)$ indicate a Control group response, that is $y(u)$ for a patient u having $s(u) = 0$, and likewise write $y_T(u)$ for a Treatment group response. The previous definitions give

$$y_C(u) = Y_0(u) \tag{2.10}$$

and

$$y_T(u) = G_{z(u)} + (1 + H_{z(u)})Y_0(u) + e_{z(u)}(u). \tag{2.11}$$

The compliance-response regression functions $C(z) \equiv E\{y_C \mid z\} = E\{y \mid s = 0, z\}$ and $T(z) \equiv E\{y_T \mid z\} = E\{y \mid s = 1, z\}$ are obtained from (2.10), (2.11),

$$C(z) = E\{Y_0 \mid z\} \tag{2.12}$$

and

$$T(z) = G_z + (1 + H_z)E\{Y_0 \mid z\}. \tag{2.13}$$

These are the functions estimated by the linear and quadratic curves in Figure 1. We will be particularly interested in the difference between the two regressions, the *observed difference* $D(z)$,

$$\begin{aligned} D(z) \equiv T(z) - C(z) &= G_z + H_z E\{Y_0 \mid z\} \\ &= G_z + H_z C(z). \end{aligned} \tag{2.14}$$

The observed difference $D(z)$ is an obvious first guess for the true dose-response function $\delta(z)$. Comparing (2.14) with (2.5), (2.6) gives a simple but important result:

Lemma. Under model (2.1)–(2.3), the observed difference $D(z)$ and the true dose-response $\delta(z)$ satisfy

$$D(z) - \delta(z) = H_z \cdot \{C(z) - C_0\}. \tag{2.15}$$

In practice we can estimate $C(z)$ and $D(z)$ from the compliance-response data, as in Figure 1. But we want to estimate the true dose-response function $\delta(z)$. Relation (2.15) shows the limitations on estimating $\delta(z)$:

1. For any given compliance-response regressions $C(z)$ and $D(z)$, there is a family of possible dose-response functions $\delta(z)$,

$$\delta(z) = D(z) - H_z \cdot \{C(z) - C_0\}, \tag{2.16}$$

corresponding to different choices of H_X in (2.1). In general, $\delta(z)$ is not completely identifiable under model (2.1)–(2.3).

2. If we assume $H_X \equiv 0$, then $\delta(z) \equiv D(z)$. Setting $H_X \equiv 0$ amounts to assuming no interaction between X and $Y_0(u)$ (see Remark A), so the no-interaction assumption makes $\delta(z)$ directly estimable from $D(z)$.

3. Let z_0 be a value of the compliance z such that

$$C(z_0) = C_0 = E\{Y_0\}. \tag{2.17a}$$

Then

$$\delta(z_0) = D(z_0), \tag{2.17b}$$

no matter what the interaction H_X may be. In other words, $\delta(z)$ and $D(z)$ always intersect at $z = z_0$.

4. If $C(z)$ is linear, then the intersection value z_0 is given by

$$z_0 = E\{z\}, \tag{2.18}$$

since a linear regression passes through the point $(z_0, C(z_0))$.

5. The curves $\delta(z)$ and $D(z)$ always intersect at two points in the plane, namely (0, 0) and $(z_0, D(z_0))$ (see Remark C). Therefore, if we assume that $\delta(z)$ is linear, it can be fully estimated from the compliance-response data via the estimation of z_0 and $D(z_0)$.

These points are illustrated by the quadratic model shown in Figure 2, a simple example of model (2.1)–(2.3), which will be used in Section 3 as part of the analysis of the LRC-CPPT data. The quadratic model assumes $C(z)$ and H_z linear, and $D(z)$ quadratic,

$$\begin{aligned} C(z) = c_0 + c_1 z, \qquad D(z) = d_1 z + d_2 z^2, \\ H_z = h_1 z. \end{aligned} \tag{2.19}$$

[This is equivalent to model (2.1), (2.3) with $G_z = g_1 z + g_2 z^2$, $H_z = h_1 z$, and $E\{Y_0 \mid z\} \equiv c_0 + c_1 z$, where $(g_1, g_2) = (d_1 - h_1 c_0, d_2 - h_1 c_1)$.] Since $C(z)$ is linear, (2.18) gives

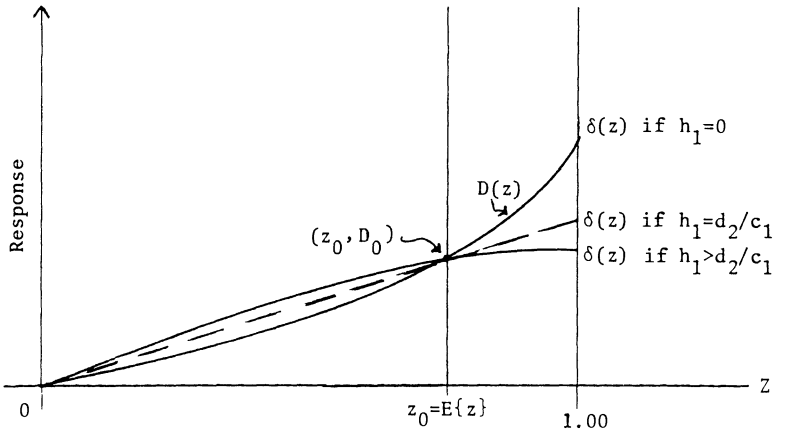

Figure 2. The quadratic model (2.19), drawn with $d_2 > 0$. Three possibilities for $\delta(z)$ are pictured: $\delta(z) = D(z)$, if $h_1 = 0$; $\delta(z)$ the dashed straight line through (0, 0) and (z_0, D_0), if $h_1 = d_2/c_1$; $\delta(z)$ bending downward, if $h_1 > d_2/c_1$. In the last case, $\delta(1.00)$ is less than the linear extrapolate D_0/z_0.

$z_0 = E\{z\}$ and $C_0 = c_0 + c_1E\{z\}$. The true dose-response curve is quadratic, according to (2.16),

$$\delta(z) = (d_1 + h_1c_1z_0)z + (d_2 - h_1c_1)z^2. \qquad (2.20)$$

Notice that $\delta(z_0) = d_1z_0 + d_2z_0^2 = D(z_0)$, in accordance with (2.17).

Figure 2 is drawn with $d_2 > 0$, so $D(z)$ bends upward. Three different possibilities for $\delta(z)$ are shown, all passing through (0, 0) and (z_0, D_0) as they must, where

$$D_0 \equiv D(z_0) = d_1z_0 + d_2z_0^2. \qquad (2.21)$$

If $h_1 = 0$, the no-interaction case, then $\delta(z) = D(z)$; if $h_1 = d_2/c_1$, then $\delta(z)$ is linear and so must equal the dashed line; if $h_1 > d_2/c_1$, then (2.20) shows that $\delta(z)$ bends downward. This last case could be worrisome. It implies that the estimate of the full-dose response $\delta(1.00)$ obtained by linear extrapolation through (0, 0) and (z_0, D_0), say,

$$\delta_{\text{LIN}}(1.00) \equiv D_0/z_0 \qquad (2.22)$$

is an overestimate.

Remark A. In model (2.1)–(2.3), requirements (2.2) and (2.3) are really definitions. They ensure that $Y_0(u)$ is actually the placebo-response, and that the "disturbance term" is just that, and not part of regressions (2.13), (2.14). The crucial part of the model is (2.1), which can be written as $Y_X(u) = \mu(X, Y_0(u)) + e_X(u)$, where $\mu(x, y_0)$ is the function of two variables

$$\mu(x, y_0) = G_x + (1 + H_x)y_0. \qquad (2.23)$$

If $H_x \equiv 0$, then $\mu(x, y_0)$ is purely additive in x and y_0, so that there is no interaction between dose X and placebo effect Y_0 in the determination of Y_X. For general H_x we calculate

$$\mu(x, y_0) - \mu(x, 0) - \mu(0, y_0) + \mu(0, 0) = H_xy_0. \qquad (2.24)$$

We see that model (2.1) allows "one degree of freedom for interaction between X and Y_0," similar to Tukey's model for interactions in a two-way ANOVA layout; see Section 4.8 of Scheffé (1959).

Remark B. We have assumed that $Y_0(u)$ and $z(u)$ are inherent attributes of patient u, so that they are not affected by u's assignment to the Treatment or Control group. [This plays a crucial role at (2.12), (2.13), where $E\{Y_0 \mid z\}$ is taken to be the same function in both equations.] We will call this the *perfect blind* assumption: that (Y_0, z) has the same joint distribution in both the Treatment and Control group, that is, (Y_0, z) is independent of the assignment function $s(u)$, (2.17).

However, Section 3 shows that perfect blindness does not hold in LRC-CPPT. The Control group enjoys substantially greater compliance than does the Treatment group. A simple correction is applied in Section 3, with some theoretical support coming in Remark H of Section 4.

Remark C. Perfect blindness implies that $C(0) = T(0)$, that is, 0% compliers have the same outcome in both the Treatment and Control group. In other words, the observed difference $D(z)$ must satisfy the

$$\textit{zero constraint}: \quad D(0) = 0. \qquad (2.25)$$

[Notice that $D(0)$ is not the same as $D_0 = D(z_0)$.] Without perfect blindness we could have $C(0) \neq T(0)$, since the zero compliers might have different distributions of Y_0 in the two groups. Notice that model (2.19) obeys the zero constraint.

Remark D. The theory presented above only considers the expectations in the two groups, namely the regression functions $C(z)$ and $T(z)$. By also considering the variance functions $\text{var}\{y_T \mid z\}$ and $\text{var}\{y_C \mid z\}$, we can reduce, though not eliminate, the unidentifiability problem for $\delta(z)$. We calculate

$$\text{var}\{y_C \mid z\} = \text{var}\{Y_0 \mid z\}$$

and

$$\text{var}\{y_T \mid z\} = (1 + H_z)^2 \,\text{var}\{Y_0 \mid z\} + \text{var}\{e_z \mid z\} \qquad (2.26)$$

from (2.10), (2.11). [This second calculation requires a stronger assumption than (2.3), $E\{e_X \mid Y_0, z\} = 0$ being sufficient.] This gives an upper bound for the interaction term,

$$H_z \leq (\text{var}\{y_T \mid z\}/\text{var}\{y_C \mid z\})^{1/2} - 1. \qquad (2.27)$$

For the quadratic model (2.19), we can express (2.27) as $c_1 h_1 - d_2 \le Q(z)$,

$$Q(z) \equiv \frac{c_1}{z}\left[\left(\frac{\text{var}\{y_T \mid z\}}{\text{var}\{y_C \mid z\}}\right)^{1/2} - 1\right] - d_2. \quad (2.28)$$

The function $Q(z)$ is estimable from the data in Figure 1. In Section 3 we will use (2.28) to show that h_1 is probably less than d_2/c_1 for the LRC-CPPT study, so that $\delta(1)$ is probably greater than the linear extrapolate in Figure 2.

3. THE LRC–CPPT DATA

The dose-compliance-response model decribed in Section 2 will now be used to analyze the LRC-CPPT data, as it appears in Figure 1. We begin with a more careful description of the variables involved. See Lipid Research Clinic Program (1984) for further details.

1. The patients were men aged 35 to 59 years with high initial cholesterol levels (total plasma cholesterol level greater than 265).
2. Two baseline cholesterol measurements were taken for each patient, one before and one after a low-cholesterol diet was suggested to them.
3. Patients were subsequently observed at two-month intervals, for a period averaging 7.3 years.
4. The nominal dose was six 4-gram packets per day of cholestyramine. This was reduced for some patients who could not tolerate 24 grams per day.

The quantity y, labelled cholesterol decrease in Figure 1, is

y = .25 · (prediet baseline cholesterol)
+ .75 · (postdiet baseline cholesterol)
− (average of all subsequent cholesterol readings). (3.1)

The weights .25, .75 were chosen as nearly optimal on the basis of a preliminary regression analysis; see Remark F. Patients returned unused packets of cholestyramine or placebo at each visit. Compliance was the proportion of assigned packets not returned, averaged over all visits. There are some obvious weaknesses of this compliance measure compared with the newly available electronic compliance monitoring, so the strength of the regression relationships in Figure 1 is perhaps surprising.

Perfect Blind Assumption. Figure 3 compares compliance in the two groups. It is obvious that compliance was better for the Controls, violating the perfect blind assumption that (Y_0, z) has the same joint distribution in both groups. As a simple corrective, each Control compliance z_C was mapped into $\tilde{z}_C = m(z_C)$, where m was defined in terms of $\hat{F}_T$ and $\hat{F}_C$, the cumulative distribution functions (cdf) for the Treatment and Control groups respectively,

$$\tilde{z}_C = m(z_C) \equiv \hat{F}_T^{-1}\hat{F}_C(z_C). \quad (3.2)$$

Mapping (3.2) makes the 171 $\tilde{z}_C$ values for the Control group have nearly the same empirical distribution as the

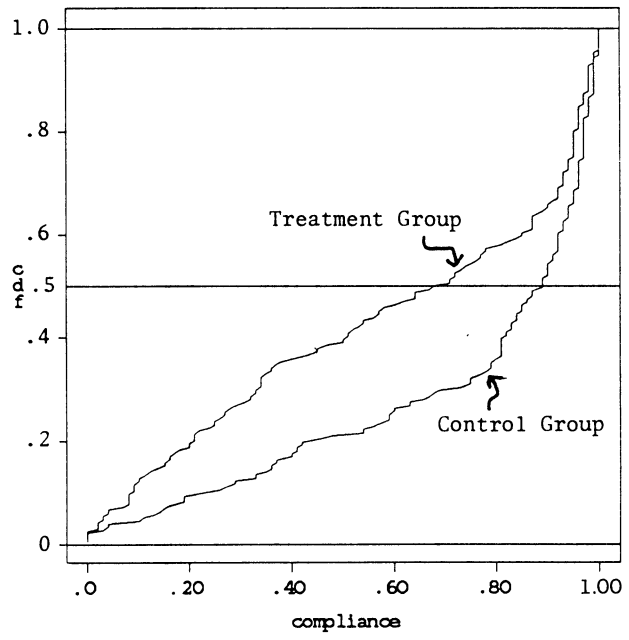

Figure 3. Compliance cdf's in the Treatment and Control groups. Compliance was significantly higher in the Control group. The average compliance for the Treatment group was z_0 = .601.

164 z_T values for the Treatment group. In particular, the average value of the $\tilde{z}_C$ is nearly the same as the z_T average, z_0 = .601.

In all of our analyses, what is called "compliance" for the Control group is actually the adjusted compliance, (3.2). Accepting result (2.15) amounts to believing that $E\{Y_{0T} \mid z_T = z\} = E\{Y_{0C} \mid \tilde{z}_C = z\}$, in particular for $z = z_0$. Here (Y_{0T}, z_T) is (Y_0, z) for a random patient put into the Treatment group, and likewise $(Y_{0C}, \tilde{z}_C)$ is $(Y_0, m(z))$ for a random patient put into the Control group. See Remark G.

Variance Structure. Figure 4 displays regression percentiles for the Treatment and Control data obtained by the method of asymmetric least squares; see Efron (1988). These are estimates of the conditional percentiles of cholesterol difference y given compliance z. The Treatment group clearly shows increasing variability of y given z as we move toward the high end of the compliance scale: the percentiles are about 2.11 times as far apart of z = 1.00 as at z = 0. The Control group percentiles, except for the lowest one, show this same effect though less dramatically: the percentiles are about 1.35 times as far apart at z = 1.00 as at z = 0.

We will use the following values for the conditional variances in the Treatment and Control groups, $v_{Ti} \equiv \text{var}\{y_{Ti} \mid z_i\}$ and $v_{Ci} \equiv \text{var}\{y_{Ci} \mid z_i\}$,

$$v_{Ti} = 471.52 + 485.00(z_i - z_0)$$
$$v_{Ci} = 198.09 + 128.33(z_i - z_0), \quad (3.3)$$

where z_0 = .601, the average compliance. (For v_{Ci}, z_i is actually the adjusted value $\tilde{z}_i$, as it will be from now on unless noted otherwise.)

The numerical coefficients in (3.3) were obtained by first running an ordinary least squares regression of cholesterol

decrease versus compliance (quadratic for the Treatment group, linear for the Control group), calculating the usual residuals r_i, and then running an ordinary linear regression of r_i^2 versus compliance. Thus, $471.52 = \Sigma_{i=1}^{164} r_{Ti}^2/164$ for r_{Ti} the ith residual in the Treatment group (almost the usual estimate of σ^2) and, likewise, $198.09 = \Sigma_{i=1}^{171} r_{Ci}^2/171$. According to (3.3), the variance ratios are

$$\frac{\text{var}\{y_T \mid z = 1.00\}}{\text{var}\{y_T \mid z = 0\}} = 1.92^2$$

and

$$\frac{\text{var}\{y_C \mid z = 1.00\}}{\text{var}\{y_C \mid z = 0\}} = 1.44^2, \tag{3.4}$$

agreeing reasonably well with the ratios obtained from Figure 4. Higher-order regressions for the variance structure gave no significant improvements over (3.3) for either the Treatment or Control group. Remark I concerns a pleasant property of the linear variance structure (3.3).

Here is an analysis of the LRC-CPPT data using the quadratic model (2.19). Other models will be considered later. The quadratic model has $T(z)$ quadratic, say $T(z) = t_0 + t_1 z + t_2 z^2$, and $C(z) = c_0 + c_1 z$ linear. The zero constraint (2.25) forces t_0 to equal c_0. (Empirical support for the zero constraint is given in Remark G.) Table 1 shows the Gauss–Markov estimates and their standard errors obtained from this model, assuming variances (3.3). The estimates are presented in terms of regressions centered at z_0:

$$C(z) = C_0 + C_1(z - z_0),$$
$$C_1 = c_1,\ C_0 = c_0 + c_1 z_0$$
$$T(z) = T_0 + T_1(z - z_0) + T_2(z - z_0)^2,$$
$$T_2 = t_2,\ T_1 = t_1 + 2t_2 z_0,\ T_0 = t_0 + t_1 z_0 + t_2 z_0^2$$
$$D(z) = D_0 + D_1(z - z_0) + D_2(z - z_0)^2,$$
$$D_2 = d_2,\ D_1 = d_1 + 2d_2 z_0,\ D_0 = d_1 z_0 + d_2 z_0^2 \tag{3.5}$$

The intersection point in Figure 2 is estimated to be

$$(z_0, \hat{D}_0) = (.601, 20.79). \tag{3.6}$$

In other words, we estimate that the true dose-response curve $\delta(z) = 20.79 \pm 2.89$ at $z = z_0 = .601$. If $\delta(z)$ is assumed to be linear (the dashed line in Figure 1), then the perfect compliance response is estimated to be

$$\hat{\delta}_{\text{LIN}}(1.00) = 34.50 \pm 4.81 \tag{3.7}$$

compared with $\hat{D}(1.00) = 46.50 \pm 3.98$, the observed difference of the compliance-response regression at $z = 1.00$.

The upper bound (2.28) for $c_1 h_1 - d_2$ is estimated to attain its smallest value, -23.24, at $z = 1.00$ [taking $\hat{d}_2 = 29.86$, $\hat{c}_1 = 10.03$, and variance functions (3.3)]. A delta-method error analysis gives the upper bound as -23.24 ± 14.96. Since $-23.24/14.96 = -1.55$, we have reasonable though not overwhelming evidence that $c_2 h_1 - d_2 < 0$, or, equivalently, that $h_1 < d_2/c_1$. But, $h_1 < d_2/c_1$ implies that $\delta(1.00)$ is greater than the linear extrapolate $\delta_{\text{LIN}}(1.00)$, (2.22). In other words, $\hat{\delta}_{\text{LIN}}(1.00) = 34.58 \pm 4.81$ is a reasonable lower bound estimate for $\delta(1.00)$.

Model Selection. The usual model selection methods point strongly toward a linear regression for the Control group data. The coefficient for the linear term is significantly nonzero ($t_{169} = 3.27$), while adding a quadratic term does not significantly reduce the squared error ($F_{1,168} = 0.59$). The C_p statistic comparing different degree regression polynomials is minimized by the linear case:

degree :	0	1	2	3	4
C_p :	15.6	4.7	6.2	7.6	8.5

(3.8)

[C_p here is the unbiased estimate of $\Sigma_1^{171}(\mu_{Ci} - \hat{\mu}_{Ci})^2/v_{Ci}$, with v_{Ci} as in (3.3), $\mu_{Ci} = E\{y_{Ci} \mid z_{Ci}\}$, and $\hat{\mu}_{Ci}$ the regression estimate of μ_{Ci}; see Mallows (1974).]

The situation is not nearly as clear-cut for the Treatment

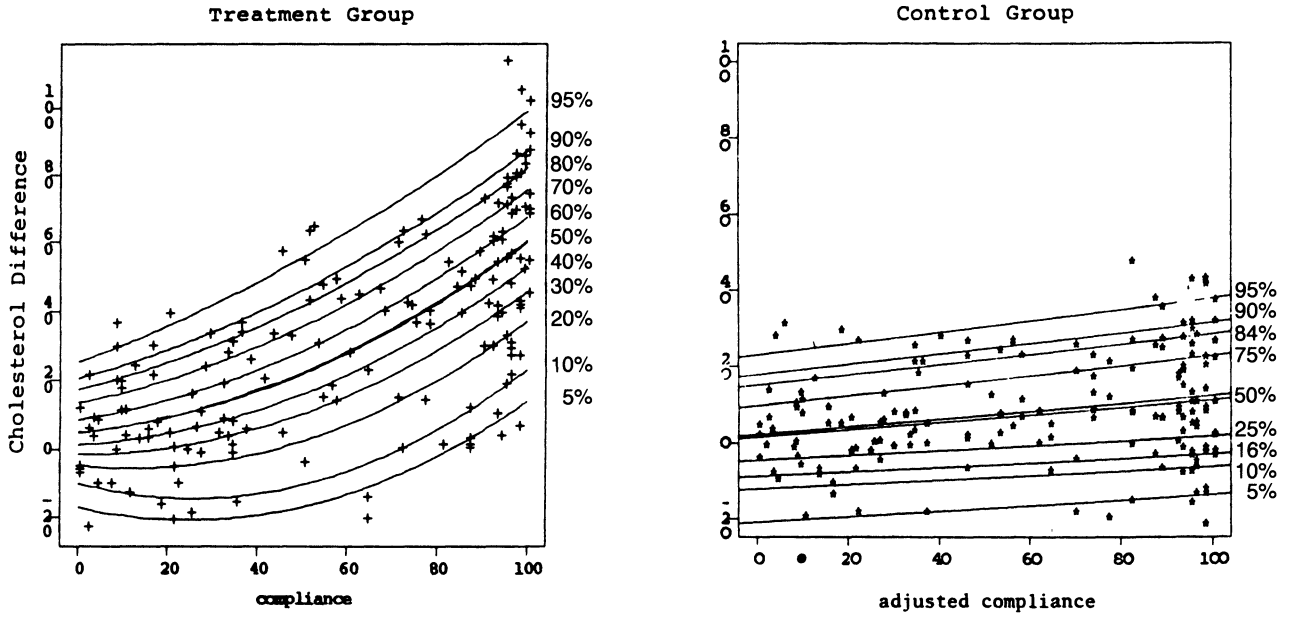

Figure 4. Regression percentiles for the data of Figure 1 estimated by the method of asymmetric least squares. The percentile curves were taken to be quadratic for the Treatment group and linear for the Control group. Both groups show increased variability of cholesterol difference with increased compliance, the effect being larger in the Treatment group.

Table 1. Estimated Coefficients and Standard Errors for C(z), T(z), and D(z), Using Data in Figure 1 for the Quadratic Model (2.19), variances (3.3), in centered form (3.5). This model satisfies the zero constraint (2.25)

	C(z)		T(z)			D(z)		
	C_0	C_1	T_0	T_1	T_2	$D_0 = T_0 - C_0$	$D_1 = T_1 - C_1$	$D_2 = T_2$
est:	8.29	10.03	29.08	62.57	29.86	20.79	52.54	29.86
se:	1.07	2.71	2.61	5.64	14.83	2.89	6.25	14.83

group. The C_p statistics are now

$$\begin{array}{lccccccc} \text{degree}: & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ C_p: & 167.9 & 6.2 & 4.7 & 5.4 & 3.2 & -1.5, & -1.9, \end{array} \quad (3.9)$$

showing a local but not global minimum at the quadratic case. Table 2 displays 11 different point estimates of $T(z_0)$, along with their estimated standard errors. The estimates range from 26.25 to 35.23. Two of the regressions are constrained to follow the zero constraint $\hat{T}(0) = \hat{C}(0)$.

Without pretending that this problem has a neat solution, the authors prefer either the constrained quadratic regression, which gave result (3.6), or the constrained quartic regression, which gives about the same answer,

$$\hat{D}_0 = 20.05 \pm 3.76. \quad (3.10)$$

Here are the reasons:

1. The zero constraint is supported by the data; see Remark F.
2. The quartic regression globally resembles "supsmu$_{+25}$," a nonparametric running linear regression smoother described in Efron (1988, sec. 6), while eliminating the local irregularities in supsmu$_{.25}$. Figure 5 compares the unconstrained quadratic and quartic regressions with supsmu$_{.25}$, which is inherently unconstrained. A bootstrap analysis showed that for no value of z was the absolute difference between the quartic regression and supsmu$_{.25}$ ever more than 1.33 bootstrap standard errors away from 0; for z in the crucial region between 50 and 70, the difference never exceeded 0.79 standard errors. (The choice .25 for the span parameter, which means that the regression at each z_i was determined by the 25% of the data points nearest in z value to z_i, was more important here than the choice of supsmu. Other local smoothers gave similar results for the same choice of span.)
3. The quintic regression is noticeably nonmonotone near $z = z_0$, actually declining for z between 68 and 82. It is also about 50% further away from supsmu$_{.25}$ than is the quartic regression.
4. The constrained quadratic regression enjoys a substantially smaller standard error. With standard errors as shown in Table 2, the constrained quadratic would have to be biased by more than 3.4 cholesterol units in order to exceed the expected mean squared error of the quintic.
5. Much of the regression difficulty here comes from the upper limit $z = 1.00$, which bunches up points at the right endpoints of Figure 1. Figure 6 shows the data replotted with the horizontal axis transformed to normalized compliance, defined as

$$z_i^N \equiv \Phi^{-1}\left(\frac{i - .5}{n}\right), \qquad i = 1, 2, \ldots, n, \quad (3.11)$$

for the ith largest compliance in each group ($n = 164$ for the Treatment group and $n = 171$ for the Control group).

Using this scale eliminated the ambiguity in the Treatment group regressions; the cubic plot was now clearly the best. The value $z_0 = .601$ nearly equals the 76th ordered value of compliance in the Treatment group, so, for com-

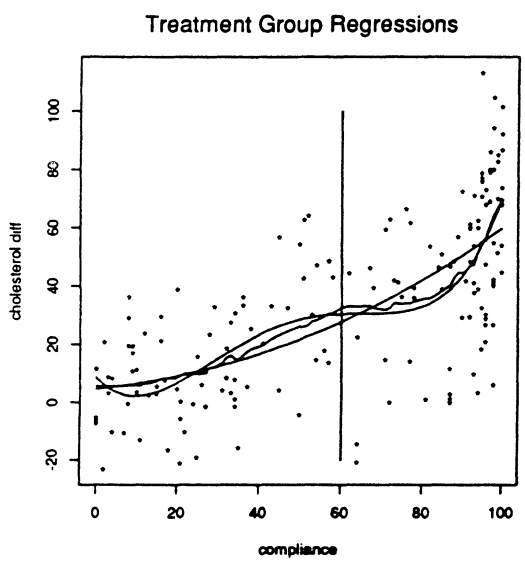

Figure 5. Comparison of unconstrained quadratic and quartic regressions for the Treatment group data, and also supsmu$_{25}$, the running linear regression smoother; the quartic regression follows supsmu$_{25}$ closely for $z > 20$.

Table 2. Estimates of $T(z_0)$, $z_0 = .601$, for Various Models Fit to the Treatment Group Data of Figure 1; Also Estimated Standard Errors; Regressions 4a and 6a Constrained to Agree with Control Group Linear Regression at $z = 0$; supsmu$_{+25}$ is a running linear regression smoother with span .25; for supsmu$_{+25}$ obtained by 100 bootstrap replications, resampling pairs (z_i, y_i)

Estimator	Estimate	se
1. $\bar{y}_T$	32.81	2.34
2. $\bar{y}_T(z \in [40, 80])$	33.40	3.78
3. Linear regression	32.81	1.69
4. Quadratic	28.64	2.80
4a. Quadratic (zero-constrained)	29.08	2.61
5. Cubic	26.25	3.47
5a. Cubic (normalized compliance)	29.23	2.02
6. Quartic	28.91	3.71
6a. Quartic (zero-constrained)	28.31	3.63
7. Quintic	35.23	4.32
7a. supsmu$_{+25}$	32.56	3.67

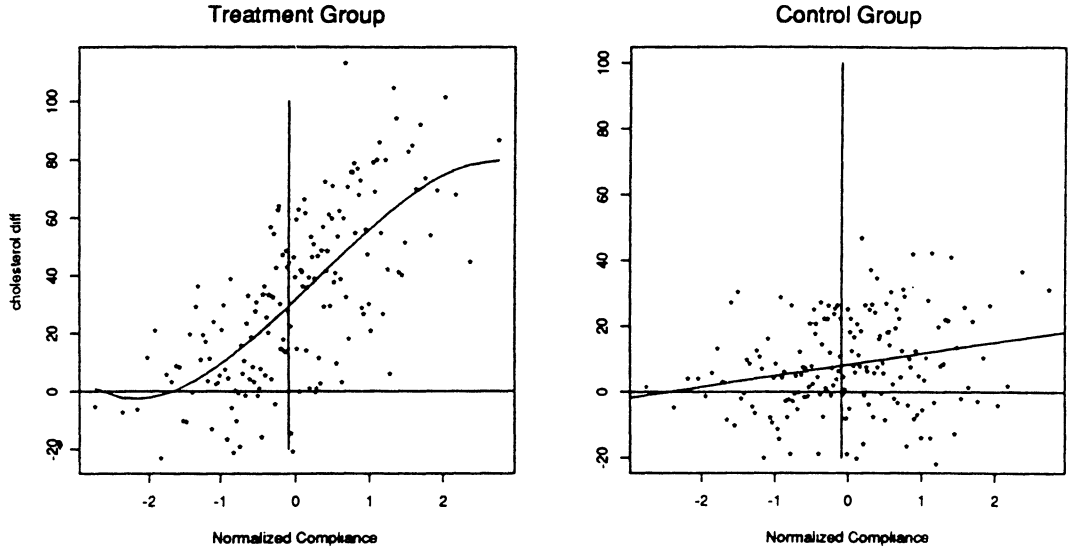

Figure 6. Treatment and Control group data of Figure 1; the horizontal axis is now normalized compliance, (3.11); the vertical axis is plotted at $\Phi^{-1}(75.5/164)$, corresponding to z_0 for unnormalized compliance. The cubic regression for the Treatment group fits better than any of the regressions using unnormalized compliance.

parison purposes, $z_0^N = \Phi^{-1}[(76 - .5)/164]$ was used as the transformed point of interest. The cubic regression, using a linear variance structure, then gave the value $\hat{T}(z_0^N) = 29.23 \pm 2.02$ appearing in Table 2. Similarly $\hat{C}(z_0) = 7.86 \pm 1.07$, giving $\hat{D}_0 = 21.37 \pm 2.28$ for this analysis, not much different than (3.6) or (3.10) (except for a smaller standard error due to a better fit on the normalized compliance scale).

Summary. The zero-constrained quadratic model estimates $\hat{\delta}(.601) = 20.79 \pm 2.89$, extrapolating linearly to the full-dose value $\hat{\delta}_{\text{LIN}}(1.00) = 34.50 \pm 4.81$; even without assuming $\delta(z)$ linear, this last estimate is a reasonable lower bound for $\delta(1.00)$; if we are willing to assume no interaction between compliance and placebo-response ($H_X \equiv 0$), then the full-dose response is estimated to be $\hat{\delta}(1.00) = \hat{D}(1.00) = 46.50 \pm 3.98$. These results are not changed much by assuming a quartic, rather than quadratic, model for the Treatment group regression.

4. SOME FINAL REMARKS

Remark D. In practice, model (2.1)–(2.3) is more suggestive than prescriptive. The statistician uses the compliance-response data, Figure 1, in the obvious way: regressions are fit to the Treatment and Control groups separately (after equalizing the two compliance scales, if necessary), perhaps honoring the constraint that they intersect at zero; the observed difference between the two regression curves $D(z)$ is a reasonable first guess at the true dose-response curve $\delta(z)$. It will be a good guess if the statistician is willing to accept the hypothesis of no interaction between compliance and placebo-response. If not, or perhaps in any case, the model focuses attention on the estimation of (z_0, D_0), the point where $\delta(z)$ intersects $D(z)$, irrespective of interactions. The intersection point entirely determines $\delta(z)$ if $\delta(z)$ is assumed linear. Variance considerations can help bound $\delta(z)$ for compliance values away from z_0, even in the presence of interactions and not assuming linearity.

Remark E. In addition to the compliance and response data, we might observe a covariate vector $w(u)$ for each patient u in the study. Adding a covariate term to model (2.1),

$$Y_X(u) = w(u)\alpha + G_X + (1 + H_X)Y_0(u) + e_X(u), \quad (4.1)$$

where α is an unknown parameter vector, gives

$$\tilde{C}(w, z) \equiv E\{y_C \mid w, z\} = w\alpha + C(z) \quad (4.2)$$

and

$$\tilde{T}(w, z) \equiv E\{y_T \mid w, z\} = w\alpha + T(z). \quad (4.3)$$

Here $C(z)$ and $T(z)$ are the functions in (2.12), (2.13). The Lemma (2.15) now becomes

$$\tilde{D}(Ew, z) - \delta(z) = H_z\{\tilde{C}(Ew, z) - \tilde{C}_0\}, \quad (4.4)$$

where $\tilde{D}(w, z) \equiv \tilde{T}(w, z) - \tilde{C}(w, z)$, $\tilde{C}_0 \equiv Ey_C$, and $Ew \equiv \Sigma_U w(u)/N$. Result (4.4) is applied to the estimation of $\delta(z)$ as before, except that the Control and Treatment group regressions now includes the $w\alpha$ term. Ew is estimated by $\bar{w}$, the vector average of $w(u)$ for the study patients.

Remark F. Another way to handle covariate information is to simply replace the actual response observation, say $\bar{y}(u)$, with a corrected version $y(u) \equiv \bar{y}(u) - w(u)\hat{\alpha}$, where $\hat{\alpha}$ is some reasonable guess for α, and then to proceed exactly as before. If, for example, $w(u)$ is a baseline measurement of patient u's response, we might just take $y(u) = \bar{y}(u) - w(u)$.

In the LRC-CPPT there were two baseline measurements, prediet and postdiet, say $w_1(u)$ and $w_2(u)$. The "reasonable guess," $\hat{\alpha} = (.25, .75)'$, was obtained from preliminary regressions of $\bar{y}$ (the average subsequent cholesterol reading) on w_1 and w_2. These gave optimal linear combinations $.34w_1 + .66w_2$ in the Control group and $.04w_1 + .74w_2$ in the Treatment group, with R^2 values .828 and .387 respectively. The compromise linear combination $.25w_1 + .75w_2$ gave R^2 values .826 and .381 respectively.

Table 3. Comparison of the 30 Patients in the Treatment Group Having Compliance $z < .20$ with the 32 Patients in the Control Group Having Adjusted Compliance $\tilde{z} < .20$; the Means in the Two Groups are not Significantly Different According to a Two-Sample t Test

	#{<20}	$\bar{y}$	$\hat{\sigma}$	t-Statistic
Treatment Group	30	6.90(4.13)*	14.56	1.33(0.55)*
Control Group	32	2.17	13.29	

*Adjustment of −2.77 made to $\bar{y}$ on the basis of the quadratic model.

Remark G. We can directly check the zero-constraint (2.25), that $T(0) = C(0)$. Table 3 concerns those patients in Figure 1 having compliance or adjusted compliance less than .20. The means and standard deviations are seen to be quite similar for the two groups.

The two-sample t statistic is nonsignificant and becomes more so after the $\bar{y}$ difference between the Treatment group and the Control group is adjusted for the fact that the z_i's used in Table 3 were not actually 0. The adjustment was done on the basis of the quadratic model (2.19), with estimated coefficients as shown in Table 1. [The average (adjusted) compliance was .08 in both groups of Table 3.] The agreement shown in Table 3, besides validating the zero-constraint for the LRC-CPPT data, offers some mild empirical encouragement for the stronger statement that transformation (3.2) restores the perfect blind assumption.

Remark H. Here is a theoretical argument supporting the use of mapping (3.2) to restore the perfect blind assumption. In using (3.2) we are really assuming that

$$E\{Y_{0T} \mid z_T = z_T^{(\alpha)}\} = E\{Y_{0C} \mid z_C = z_C^{(\alpha)}\}, \tag{4.5}$$

where $z_T^{(\alpha)}$ is the αth percentile of the z_T distribution, and similarly for $z_C^{(\alpha)}$ [reverting to the notation following (3.2), so z_C is again the original, unadjusted Control compliance]. Relation (4.5) will be true if $z_T = m(z_C)$ for some monotonic function m; in other words, if each patient's Treatment compliance is a monotone deterministic function of his Control compliance: $(Y_{0T}, z_T) = (Y_{0C}, m(z_C))$.

Relation (4.5) will not be true in general if z_T and z_C are correlated but not deterministically related. As an example, let $Z_C \equiv \Phi^{-1}F_C(z_C)$ and $Z_T \equiv \Phi^{-1}F_C(z_T)$, where Φ is the standard normal cdf and so $Z_C \sim N(0, 1)$ (ignoring the possible discreteness of F_C). Suppose that, on this scale, the mapping from Z_C to Z_T is

$$Z_T = m(Z_C + \epsilon), \tag{4.6}$$

for some monotonic function m, where $\epsilon \sim N(0, \tau^2)$ independently of (Y_{0C}, z_C); that is, $(Y_{0T}, Z_T) = (Y_{0C}, m(Z_C + \epsilon))$. Also assume that $E\{Y_{0C} \mid Z_C\}$ is linear in Z_C, as seems to be at least approximately the case for the LRC-CPPT data.

Under these assumptions it is easy to show that

$$E\{Y_{0T} \mid z_T = z_T^{(\alpha)}\} = \rho E\{Y_{0C} \mid z_C = z_C^{(\alpha)}\} \tag{4.7}$$

where $\rho = 1/\sqrt{1 + \tau^2}$ is the correlation between Z_T and Z_C. For example, if $\tau^2 = 1$, so that the disturbance ϵ has the same variance as Z_C in (4.2), then $\rho = .71$.

Ignoring the fact that (4.7) rather than (4.5) is true gives an overestimate of the placebo-factor contribution $E\{Y_{0T} \mid z_T\}$ to the Treatment group regression (2.13). This would seem to suggest an underestimation of the dose-response curve δ. However, most of the theory in Section 2 uses (4.5) only at $z = z_0 \doteq z_C^{(.5)}$; and at $\alpha = .5$, all of the regressions in (4.5), (4.7) equal 0, so that there is no difference between (4.5) and (4.7). This reassuring result can be obtained in broader circumstances than (4.6), but won't be pursued further here.

Remark I. It is easy to prove the following theorem: given n pairs of points (z_i, y_i) with a linear variance structure $\text{var}\{y_i \mid z_i\} = a_0 + a_1 z_i$, the weighted Gauss–Markov linear regression line for y on z passes through the unweighted mean value point $(\bar{z}, \bar{y}) = (\Sigma(z_i/n), \Sigma(y_i/n))$. In other words, assuming a linear variance structure, as we did at (3.3), does not affect the value of the estimated linear regression at $z = \bar{z}$. Nearly the same result is true for higher-order polynomial regressions. The variance structure (3.3) has little effect on estimates (3.6) or (3.10), no matter what values we use for the coefficients in (3.3), though it does affect the estimated standard errors.

Remark J. The placebo-response effect is small compared to the true dose-response in LRC-CPPT. This isn't always the case. In the clofibrate study, Table 1 of Coronary Drug Research Project Group (1980), evidence is presented for a placebo effect almost as large as the Treatment effect! The theory presented here will estimate true dose-response δ near zero in such a case.

REFERENCES

Chevalley, C. and Urquhart, J. (1987), "Medication Event Monitoring to Project Impact of Patients' Dosing Errors on Clinical Trial Results," Abstracts of the 8th Conference, International Society of Clinical Biostatistics, Gothenburg: Sweden.

Coronary Drug Research Group (1980), "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project," *New England Journal of Medicine,* 302, 1038–1041.

Efron, B. (1988), (in press) "Regression Percentiles Using Asymmetric Squared Error Loss," *Statistica Sinica.*

Holland, P. (1986), "Statistics and Causal Inference" (with discussion), *Journal of the American Statistical Association,* 81, 945–970.

Holland, P. (1988), "Causal Inference, Path Analysis, and Recursive Structural Equations Models," *Sociological Methodology 1988,* C. Clogg.

Koch, G., Amari, I., Davis, G., and Gillings, D. (1982), "A Review of Some Statistical Methods for Covariance Analysis of Categorical Data," *Biometrika,* 38, 563–595.

Lipid Research Clinic Program (1984), "The Lipid Research Clinics Coronary Primary Prevention Trial Results, Parts I and II," *Journal of the American Medical Association,* 251, 351–374.

Mallows, C. (1973), "Some Comments on C_p," *Technometrics,* 15, 661–675.

Scheffé, H. (1959), *The Analysis of Variance,* New York: John Wiley.

Urquhart, J., and Chevalley, C. (1988), "Impact of Unrecognized Dosing Errors on the Cost and Effectiveness of Phamaceuticals," *Drug Information Journal,* 22, 363–378.