Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

A Psychometric Model for

Detection of Idiographic Treatment Effect

by

ROBERT GOLDEN and TRAVIS THOMPSON

Report Number PR-76-7

October, 1976

Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

A Psychometric Model for Detection of Idiographic Treatment Effect

by

Robert Golden and Travis Thompson

Report Number: PR - 76 - 7 October, 1976

This research was supported in part by grants from the Psychiatry Research Fund.

I. Introduction

The general problem giving rise to the psychometric model discussed in this report is that of determining in an objective way the extent that a given treatment program is helping a patient. The outcome of a treatment is generally multidimensional when viewed nomothetically (i.e. across individuals), and must be evaluated in the presence of unique situational and general nuisance variables which are difficult to control and even in some cases, to identify. As a result of such complexity, several writers have concluded that therapy outcome is probably impossible to evaluate accurately (e.g. see Kiesler, 1966, Bergin and Garfield, 1971). The present model allows the use of a single and, possibly, unique criterion variable for each patient. The use of a unique criterion for each patient is one way to attempt to avoid the multidimensionality-of-outcome problem of nomothetic studies; that is, the complexity may not be due to multidimensionality within individual patients' behavior so much as it is due to individual differences in the dimensions of change across individuals.

The present model is idiographic, in that it is explicitly intended to be applied to the data of one individual at a time. Although such an approach has been proposed before (e.g. see Chassan, 1967), it is apparently often assumed that statistical models can only be properly applied to groups of individuals, since independent observations require the use of different individuals. The present report is intended to illustrate that the use of nomothetic methods is not necessary and the use of idiographic vs. nomothetic methods is not best handled by theoretical debate but rather as empirical questions which can be answered as such. In particular, idiographic methods for testing the assumptions of the present model are suggested and it is claimed that such tests can be

used to determine if the model is valid or not.

When a population of individuals is studied in the usual nomothetic way the implicit assumption is made that some actual psychological structure is the same for each individual; that is, that they vary only with regard to the parameter values of the structure. Obviously this assumption is not always warranted. Some sort of empirical evidence is necessary to prevent the nomothetic approach from being easily challenged at the outset on theoretical grounds (Lykken, 1975).

A typical nomothetic statistical model used in the comparison of, say, an experimental group with a control group is the two-factor repeated-measure analysis of variance model. In this model it is assumed that the rate of improvement (on the factor that is repeatedly measured) is the same for all individuals within a treatment group; that is, all such rate of improvement differences are due to measurement error. Now, of course, in the treatment of patients, this assumption is typically not even close to the truth. It is more often true that in a given treatment group the number of individuals that show marked improvement will be of the same order as the number that clearly show no improvement or get worse. These differences in response to a treatment are not, in the main, random measurement error but reliable individual differences. This is simply a well-known and obvious clinical fact. The analysis of variance model was originally developed for use in agronomy such as for the comparison of the growth of corn stalks given one treatment rather than another. Suffice it to say that while individual corn stalks evidently may be treated as replicates, each having the same rate of change parameter except for independent random error, an analogous situation for patients typically fails to obtain!

It is clear that the reactions of patients to a psychotropic drug, are typically so varied that nomothetic studies, even when diagnosis is carefully controlled, partialled out, etc. often yield either no differences or low, unimpressive statistically significant differences between treatment and control groups. Further, it is difficult to apply such an average or group result to a particular patient for whom the drug or treatment might be considered. Finally, even if the group averages could be meaningfully applied to a patient, there usually exists only a highly fallible diagnosis to determine which experimental drug group the patient resembles most. Most importantly, there may be no such group to which the patient is sufficiently similar so that any useful prediction of outcome for that patient can be made.

II. Development of the Model

Let us assume that we have a single global indicator of the performance of the patient, and that measurements on this indicator can be repeatedly taken, at regular intervals. For example, points can be awarded each day for various activities (such as participating in social events, getting out of bed, etc.) that have been selected in advance as indicators of the patient's improvement. The total points earned each day for these activities can be used to obtain repeated measurements of the patient's behavior. Given a number of such measurements, the major statistical problems are to determine if there is improvement or not and to estimate the amount of improvement over a number of days.

Let i index the days beginning with, say, the first day of treatment, and let x_i be the value of criterion variable for the ith day. We know that x_i is usually not even an approximately smooth function of i but fluctuates greatly day to day due to (a) measurement error, however defined

(e.g. due to sampling of behavior that is rated), (b) actual changes due to environmental sources that vary from day to day and partially determine the criterion behavior (e.g. different opportunities for social interaction), and (c) to actual changes in the patient himself. However, if there is some improvement in the criterion behavior over a period of time it is possible to write $x_i = g(i) + d_i$ where g is some monotonic increasing function of i, and d_i is the daily deviation due to measurement error, and to actual daily fluctuation in the patient and the environment. On the other hand, if g is a monotonic decreasing function of i, then we have indication that the treatment should be changed.

Suppose for the moment that we restrict g(i) to be linear; that is to be of the form $a + b(i - \bar{i})$ where a and b are constants and $\bar{i}$ is the mean day of the days during which measurements were made. In order to determine the values of a and b for a given patient it is necessary to make certain assumptions about the d_i.

The well-known maximum-likelihood criterion for the formulation of this regression problem can be used if we assume that the d_i are each independent of $g(i) = a(i - \bar{i})$ for each i, that they are mutually independent, and that they are each normally distributed with a zero mean and a common variance, call it s^2, even though we only have one realization of each d_i.

It is natural to expect these assumptions to be adequate approximations so that accurate inferences can be made from the model in some instances and not so in others. This is a matter which must be somehow empirically handled, not one which can be solved by convention, fiat or a priori considerations. It is the purpose of the model 'consistency tests,' some of which are proposed below, to determine if the model fits

well enough or not. Before considering such tests let us develop the equations for parameter estimation.

The maximum-likelihood procedure gives the values of $\underline{a}$ and $\underline{b}$ which maximizes the likelihood of obtaining the given (sample) results if the above assumptions are perfectly true. It turns out, that such estimates are given by $\hat{a} = \bar{x}$, and $\hat{b} = \frac{\Sigma(x_i - \bar{x})(i - \bar{i})}{\Sigma(i - \bar{i})^2}$; also the estimate of the variance of d_i is given by $\hat{s}^2 = \frac{1}{n} \Sigma\{x_i - \hat{a} - \hat{b}(i - \bar{i})\}^2$ (for example, see Lindgren, pp.360ff). Lindgren points out that these estimates are also the least-squares estimates; that is, they are the ones that minimize $\hat{s}^2$.

In Figure 1 is a graph of the daily points earned in a behavior modification study for the first 14 days of hospitalization of a manic patient given lithium (Thompson et. al. 1976). The most striking characteristic of the graph may be the lack of day-to-day stability and/or reliability. However, this lack of reliability and stability does not prevent the indicator from having adequate validity, even though it sometimes is assumed that high reliability is always a psychometric necessity. It is of interest, for example, that blood pressure as measured in a doctor's office has relatively low reliability but evidently unsurpassed validity for many purposes. An equally outstanding characteristic of the graph in Figure 1 is that there clearly is a general trend for the daily points to increase with respect to the number of days of treatment. To help assure ourselves that this is the case it would be of interest to see if indeed there is enough information in the 14 measurements of behavior to reject the null-hypothesis that b=0. When b=0 it can be shown that the quantity $\frac{\hat{b}S_d}{\hat{s}}(n - 2)^{\frac{1}{2}}$ (where n is the number of days of treatment,

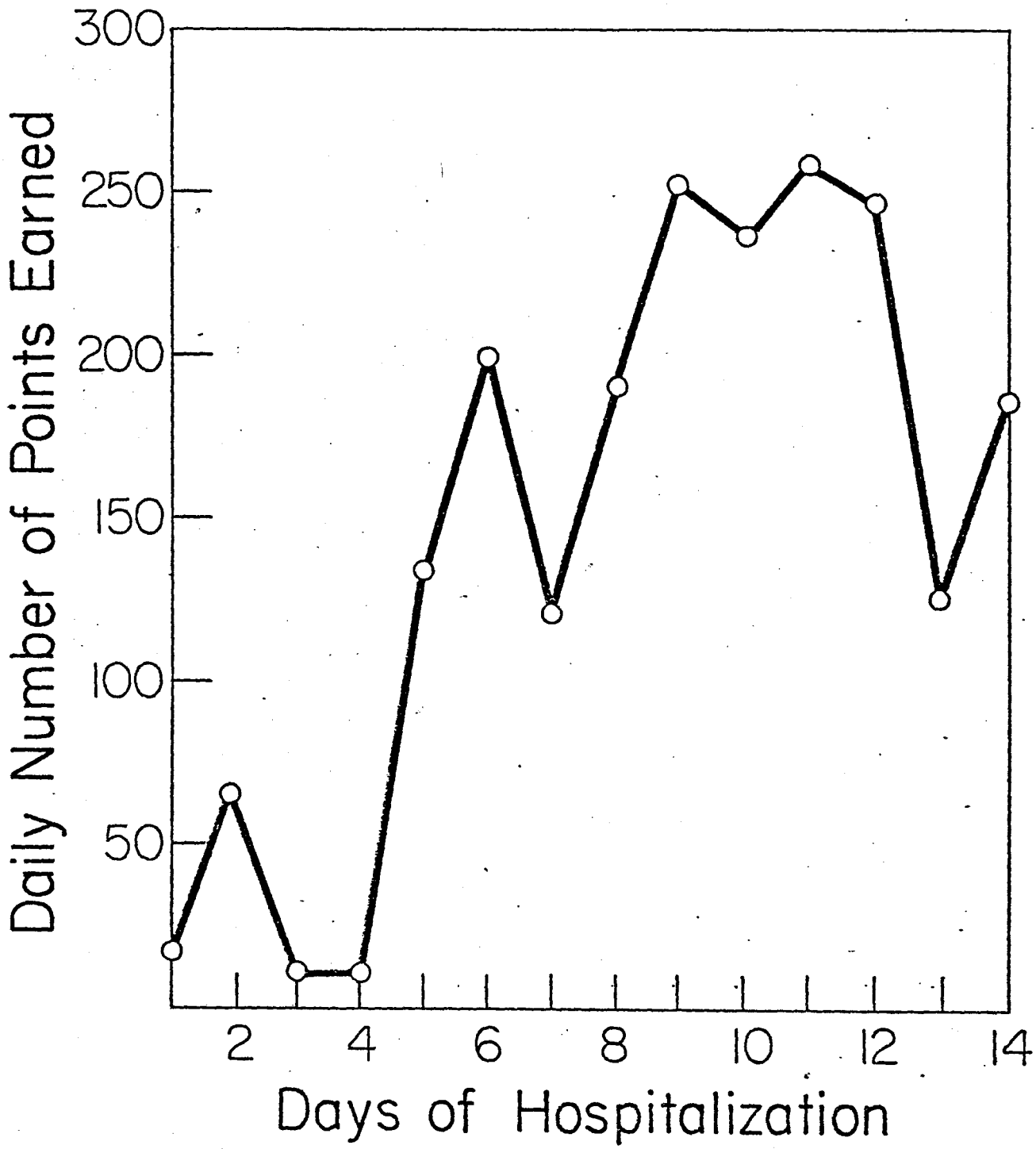

Figure 1. Number of points earned for each day of hospitalization for a lithium treated manic patient.

and S_d is the standard deviation of treatment day values) is distributed as t with n - 2 degrees of freedom. It might be noted there that if the treatment-evaluation days are contiguous; that is, i has the values 1, 2, 3, ..., n then it can be shown that

$$S_d = \left[\frac{n^2 - 1}{12}\right]^{\frac{1}{2}}.$$

The value of t for the example is 3.87 which for 12 degrees of freedom is significant at the .005 level.

We can also test the hypothesis $b = b_o$ since $\frac{(\hat{b} - b_o)S_d}{\hat{s}}(n - 2)^{\frac{1}{2}}$ is also distributed as t with n - 2 degrees of freedom if $b = b_o$. If we require that there be at least 5 points per day average increase, then the null hypothesis b = 5 is rejected for the illustrative example at $\alpha = .01$; to do so for b = 10 would evidently require further treatment measurements.

We can estimate the total improvement due to the general trend component even though it exists in the presence of day-to-day instability and unreliability. To do this we consider c = g(n) - g(1) which reduces to $c = \hat{b}(n - 1)$ for the linear case. If we use $\hat{c} = \hat{b}(n - 1)$ we see that $\hat{c}$ is a random variable which gets its randomness through day-to-day instability and measurement unreliability. It follows that an estimate of a $1 - \alpha$ confidence interval for c is given by

$$\left[\hat{b}(n - 1) - t_{(1 - \frac{\alpha}{2})}k,\ \hat{b}(n - 1) + t_{\frac{\alpha}{2}}k\right]$$

where

$$k \doteq \left[\frac{12}{n}\right]^{\frac{1}{2}}\hat{s}.$$

For the example, since $\hat{b} = 16.30$, $\hat{s} = 58.80$ and n = 14 we obtain (180,243) as a 95% confidence interval for c. Also we can determine the

probability that $c > c_o$, some predetermined minimum number of points required for sufficient improvement. For example, if $c_o = 100$, the probability we would observe a value for $\hat{c}$ as high as obtained in the example is less than .01.

One generalization of the model above consists simply of using two regression lines, one for the first p days, and one for the remaining n - p days. As will be shown, there is evidence that for manic patients treated with lithium the general linear trend $a + b(i - \bar{i})$ has a different a - value for the first week or so of hospitalization than for the remaining period of hospitalization, the former a - value being considerably lower than the latter one. It is expected that two regression lines will often fit treatment evaluation data better than a single line especially if there is a relatively discrete point at which the treatment begins to take effect.

It could also happen that the two b - values (the slopes) are quite different but for present purposes we will require the two b - values to be the same. This is, of course, another idealization or model assumption which must be eventually checked like the others.

Thus we have

$$x_i = a_1 + b(i - \bar{i}_1) + d_i \text{ for } i \leq p,$$

and

$$x_i = a_r + b(i - \bar{i}_r) + d_i \text{ for } i > p,$$

where 1 denotes the left or first regression line and r denotes the right or second regression line. As before we assume

A_1: The d_i are each distributed normally with zero means and variance s^2 for all i,

A_2: the d_i are mutually independent, and

A_3: the d_i are independent of i.

We can determine the value of p, the point of 'regression discontinuity', which provides for the best model fit. Consider some value of i less than p, call it k, then the expected value of the mean of the x_i's for all $i \leq k$ is given by

$$E(L_k) = \{\sum_{i=1}^{k} a_1 + b(i - \bar{i}_1)\}/k$$

where $L_k = \sum_{i=1}^{k} x_i/k$. It follows that

$$E(L_k) = \frac{b}{2}(k + 1) + a_1 - \frac{bp}{2}(1 + p).$$

Similarly for the x_i values where $p \geq i > k$ we have

$$E(U_k) = \left[\sum_{i=k+1}^{p} a_1 + b(i - \bar{i}_1) + \sum_{i=p+1}^{n} \{a_r + b(i - \bar{i}_r)\}\right]/(n - k)$$

where $U_k = \sum_{i=k+1}^{n} x_i/(p - k)$. It follows that

$$E(U_k) = a_1(\frac{p - k}{n - k}) + a_r(\frac{n - p}{n - k})$$

$$+ b\left[(p - k)(k + 1 + p) - (k + 1 + p) + (n - p)(p + 1 + n)\right]/2(n - p)$$

and

$$D_k = E(U_k) - E(L_k) = \frac{nb}{2} + (a_r - \bar{i}_r b - a_1 + \bar{i}_1 b)(\frac{n - p}{n - k}).$$

Since $\bar{i}_1 - \bar{i}_r = \frac{-n}{2}$, we have

$$D_k = \frac{nb}{2} + (a_r - a_1 - \frac{nb}{2})(\frac{n - p}{n - k})$$

which has a maximum at k = p. In a similar manner it can be shown that for $k \geq p$, D_k has a maximum at k = p. Thus, D_k has a local maximum at p under the assumptions of the model and the condition $a_1 < a_r$. The expected value of maximum is given by $E(D_p) = a_r - a_1$. The curve of D_k as a function of k for the illustrative example is given in Figure 2; it is

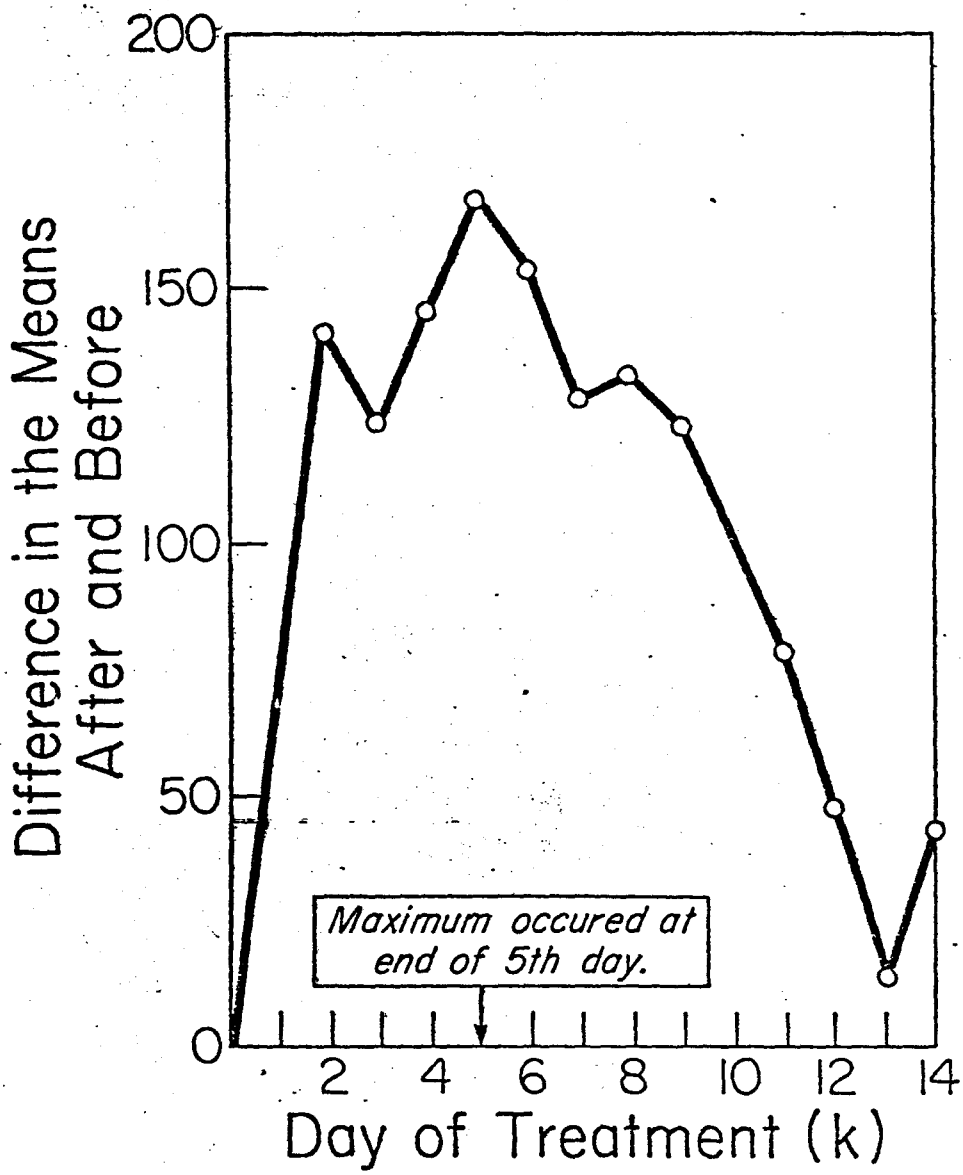

Figure 2. The mean daily points after day k less the mean daily points before day k for the illustrative example.

noted that there exists a well-defined local maximum at day 5.

A pooled estimate of b is given by

$$\hat{b} = \frac{\sum_{i \le p} (x_i - \bar{x}_l)(i - \bar{i}_l) + \sum_{i > p} (x_i - \bar{x}_r)(i - \bar{i}_r)}{\sum_{i \le p} (i - \bar{i}_l)^2 + \sum_{i > p} (i - \bar{i}_r)^2}$$

For the example, the value of $\hat{b}$ is 3.5. We estimate $\hat{a}_l$ and $\hat{a}_r$ as in the single regression line model; for the example we obtain $\hat{a}_l = 47.0$ and $\hat{a}_r = 200.9$. We can now determine the d_i values for each regression line (See Table 1); from these values we obtain $\hat{s}_l = 45.9$ and $\hat{s}_r = 48.7$ which gives the unbiased pooled estimate $\hat{s} = 51.5$.

Under the above assumption of homoscedasticity, it can be shown that $\hat{a}_l$ is normally distributed with mean a_l; likewise it is shown that $\hat{a}_r$ is normally distributed with mean a_r. It follows that under the null hypothesis where $a_l = a_r$ that $\frac{\hat{a}_r - \hat{a}_l}{\hat{s}(n/n_l n_r)^{\frac{1}{2}}}$ is distributed as t with n - 4 degrees of freedom. For the example $\hat{a}_r - \hat{a}_l = 153.9$ which results in a t-value of 7.63 (10 df) which is significant at $\alpha = .001$. It would appear that such a test is one of the most sensitive tests for the detection of any overall improvement for such a treatment study.

If we think of the discontinuity as being caused by a relatively sudden onset of a treatment effect, it would be more impressive to show that the two regression lines are not coincident. The lines are coincident if

$$a_l + b(p - \bar{i}_l) = a_r + b(p - \bar{i}_r).$$

The expected jump at p is given by

$$D = a_r + b(p - \bar{i}_r) - a_l - b(p - \bar{i}_l), \text{ or}$$

$$D = a_r - a_l + b(\bar{i}_l - \bar{i}_r);$$

Table 1

The Two-Line Model Estimates of the General Trend Component $E(x_i)$ and the Daily Fluctuation $x_i - E(x_i)$ for the Illustrative Example

Day	x_i	$E(x_i)$	$x_i - E(x_i)$
1	15	40	-25
2	66	43	23
3	10	47	-37
4	10	51	-41
5	134	54	80
6	200	187	13
7	120	190	-70
8	190	194	-4
9	252	197	55
10	237	201	36
11	259	204	55
12	238	208	30
13	127	211	-84
14	185	215	-30

thus,

$$D = a_r - a_1 - \frac{nb}{2};$$

and for the example, $\hat{D} = 129.4$. Also $\text{var}(\hat{D}) = \text{var}(\hat{a}_r) + \text{var}(\hat{a}_1) + \frac{n^2}{4}\text{var}(\hat{b}) = \frac{\hat{s}^2}{n_r} + \frac{\hat{s}^2}{n_1} + \frac{n^2}{4}\left(\frac{12\hat{s}^2}{n(n^2 - 1)}\right)$; for the example, $\text{var}(\hat{D})$ has the value $(37.4)^2$. Since $\hat{D}$ is approximately normally distributed we can test the null hypothesis $D = 0$, by using $z = \frac{\hat{D}}{\text{var}(\hat{D})^{\frac{1}{2}}}$; for the example $z = 3.46$ which is significant at the .001 level.

III. Testing the Model

Many other indices of improvement (such as the total change) can be estimated for use along with confidence intervals and hypothesis tests. However, it would seem that the use of the jump at the point of model discontinuity should lead to interesting construct validity results when the model holds. It should be noted that for real phenomena there probably is seldom an actual discontinuity point, as required by the model, but, rather, a point of inflection. Nevertheless, the appeal of this model exists to the extent that the discontinuity or inflection point is known to be associated with the onset of treatment effect, for then the regression line prior to p can evidently be used as a baseline for the evaluation of the treatment effect for the individual.

The model was used in one small study where direct observation of patient behavior for evaluating the effects of lithium in treating acutely disturbed manic patients, was used to determine the role of reinforcement procedures in combination with lithium treatment in effecting behavior change. The subjects were five patients admitted to a twelve bed acute and short term adult psychiatry unit of the University of Minnesota Hospital. Two were female, aged 20 and 22 and three male, aged 22, 30

and 39 years of age. All had prior histories of manic episodes and two reported histories of severe depression, including scuicide attempts.

The service on which the patients were treated functions as a token economy in which a variety of adaptive, constructive activities earn points, and destructive (to self or others) or grossly inappropriate behaviors lose points. Each patient has an individualized treatment plan written in the form of a contract which is approved and signed by the patient, the medical officer, primary nurse and the psychologist assigned to the case. Points earned on any given day are exchangeable for supplementary goods and services (e.g. edibles, T.V. rental, outings with staff) the following day. The data of one patient in the study was used for the example above. The point of discontinuity was found to correspond with the point at which a therapeutic blood level (1.2 meq/liter in acutely psychotic and 0.7 - 0.8 meq/liter in subacute patients) was obtained for each of the 5 patients (see Figure 3). While the procedures had little effect in improving the behavior of manic patients prior to p, it appeared that behavior began to improve after p. Evidently, lithium made the manic patient more sensitive to the environmental reinforcement contingencies, since previous studies indicate that without such behavioral therapy, 10 to 14 days are required before there is improvement. Most important, however, is that the correspondence of p with lithium blood level provides some favorable construct validity for the discontinuity characteristic of the model. Suffice it to say here that the other 4 manic patients showed model estimated improvement similar to the patient in the above example.

On the other hand, the data of 5 depressed patients treated with amytriptyline did not show an inflection point upon inspection of the raw data. Further, the "differences in the means" curve showed no

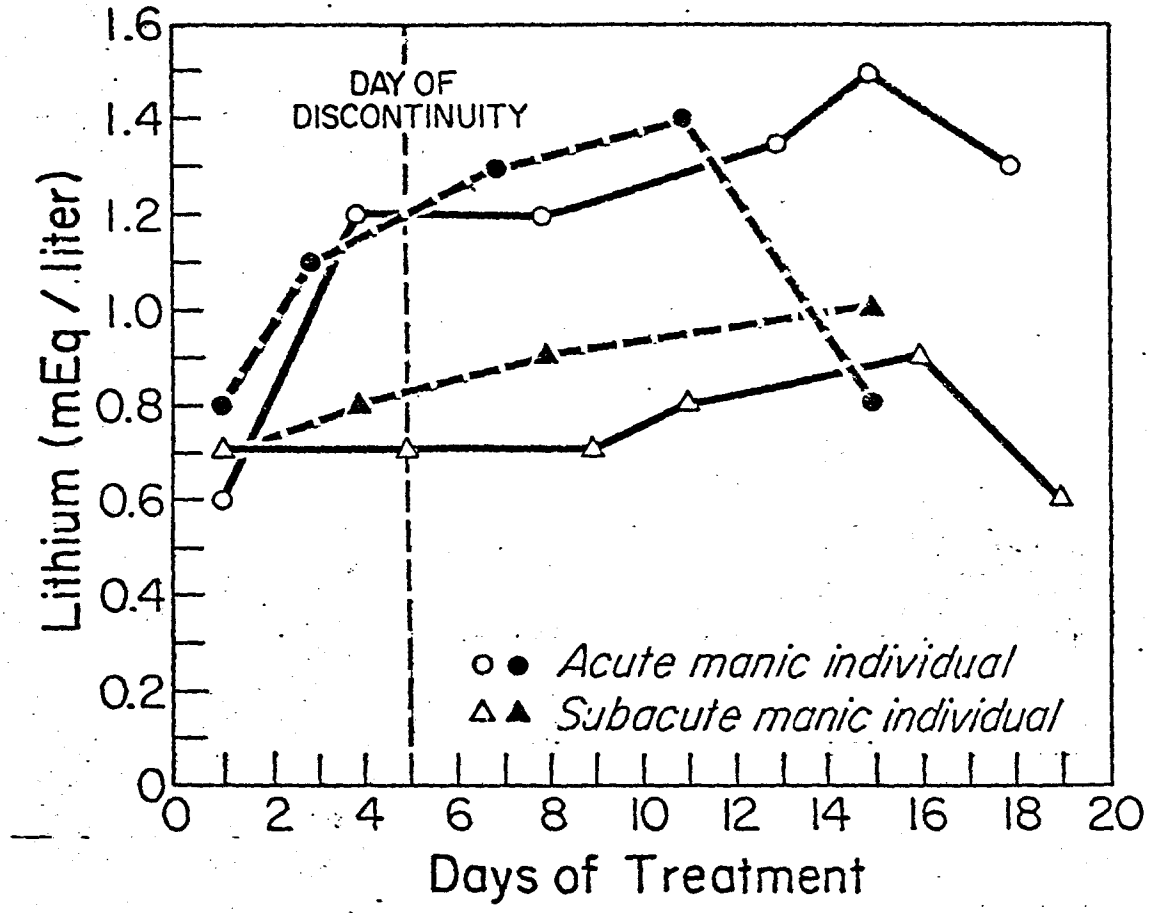

Figure 3. The lithuim blood level as a function of the days of treatment for four manic patients.

evidence of a local maximum as was the case for each of 5 manic patients. This small study, thus, shows that we can expect the two-line model to be useful for certain kinds of treatments and patients but not others.

Any of the model based parameters of improvement such as the slope of the regression line, the total change, and the difference in the regression line means can be subjected to construct validity tests of either a nomothetic or an idiographic nature or both. For example, the relationship of such a parameter with changes on a personality inventory, physiological or biochemical measures can be determined within or across individuals. If these relationships are strong enough, then the problem of testing the assumptions of the model has been taken care of in a straightforward and simple manner. It is often the case, however, that such extra-model measures, such as lithium blood level, are not known or available. One reason for such a state of affairs is the typical multi-dimensionality of the change due to the treatment; that is,not any one measure is a clear criterion. Often, the best that is available is a single criterion of change (such as the point earning behavior). In such a situation we can, nevertheless, still test the model with regard to the assumptions upon which it is based. Here we are concerned with detection of spurious results such as finding a statistically significant improvement for an individual when, in fact, there was no improvement at all. In such an instance the spurious result is caused, for example, by excessive violation of the normality, homoscedasticity or independence assumptions or because of measurement error.

Consider the assumption concerning the independence of the x_i's. If the contiguous x_i's are somewhat dependent then there might result longer than expected runs of the same sign of the d_i values. How long

a run must be to be considered excessive can be determined by Monte Carlo study as described below. The normality of the d_i's should result in close enough agreement between the number of d_i values between $\pm$ one sigma unit of the actual d_i distribution and the theoretical one for example. The linearity of the x_i values, if violated, could again result in excessively long runs of the d_i signs, the d_i's being too highly correlated with i (time) or a superior fit of a non-linear function.

Undoubtedly better tests than these simple ones can be developed. The proposed tests are only offered as illustrative examples of candidates to be evaluated by further analytical work or by the Monte Carlo method. The development of useful consistency tests can be a complex mathematical problem and is beyond the scope of the present report.

Suppose the x_i values are determined by a random number generator such as a multivariate normal one described in Golden et al.(1974). In this method the means, variances and correlations of the x_i's are specified such that artificial data can be generated so that there is systematic departure from the linearity and independence assumptions. (Other data generation methods such as described in Golden et al. (ibid) can be used to generate non-normal x_i values). Monte Carlo study of the model and consistency tests would consist of repeated generation of data samples, analysis of each by the model, comparison of the model results with the known true values, and the evaluation of the accuracy of the consistency tests in the detection of both spurious and inaccurate results. The results of such a study would then allow us to be confident that assumption departures are not too deviant for some data and that they are too deviant for other data.

The passage of powerful consistency tests does not imply that model-based parameter of improvement has adequate or predictive construct validity. The criterion chosen, could, for example, be irrelevant for clinical purposes, even though the treatment resulted in a genuine change on the criterion. Thus passage of internal consistency testing of the model is intended to be a necessary but not sufficient condition for construct or predictive validity to obtain. It is conjectured, however, that many results, otherwise consistent and coherent will be correctly found through consistency testing to be spurious or grossly inaccurate and therefore, totally misleading.

While the failure to obtain construct or predictive validity from a model based parameter can be due to (a) the lack of validity of the criterion measure, or (b) the lack of verisimilitude of the model used to detect or estimate the change, it can also be due to (c) the lack of reliability in the detection of estimation of change. The greater the unreliability of measurement of the x_i values, the larger the d_i values and, for a given improvement rate, the lower the proportion of the variance of the x_i values accounted for by the regression function. Most importantly, unreliability will allow for a greater chance of spuriousness. Again, the Monte Carlo study discussed above could evidently be designed to determine what the reliability must be (say, as a function of regression slope) in order for one to be confident of the model results. In brief, unreliability causes loosening of the model fit which is exactly what consistency tests are intended to detect.

As seen in the point earning example above, the reliability of the criterion measure can be quite low in drug therapy treatment. Lindgren (ibid) has pointed out that to make the variance of b as small as possible,

the optimal arrangement of n measurements is to put most at the beginning and at the end of each regression line interval. It follows that to make the hypothesis tests developed above for the two-line model more powerful it would be advantageous to make 2 or 3 measurements at i = 1, p, p + 1, and n, where we must estimate p prior to the experiment. Evidently this slight change can have a dramatic effect on the power of the tests, enough so that an inflection at day 5 could probably be detected by the end of day 6 or 7, rather than a week or so later.

The problems in lack of reliability in an improvement parameter are well known (e.g. see Harris, 1963). Whereas this literature has focused on change scores calculated from just two scores, the present method does allow for the use of any number of scores and presumably the greater this number, the greater the reliability of improvement parameter; again Monte Carlo study is required.

With all of these uncertainities with regard to unreliability it would seem to be highly useful to make two measurements at each point, to analyze the two sets of data separately and to compare the results. This was done in one unpublished study where two equally face valid criteria were used on a sample of 44 psychiatric patients for 21 weeks (one measurement/week). One measure was supervisor rating of out-of-hospital work performance (w) and the other was nurses rating of behavior in the hospital (h). The single line regression model showed that there was significant criteria and that the fit of the model was evidently very good in each case as the correlation between the observed and model estimates of the x_i values was nearly always above .9. However, the correlation across patients of the two improvement parameter values (the slopes or b - values for each measure) was .08. Thus, it follows

that either w and h are measures of unrelated traits, or w and h are low in reliability (for such purposes), or b_w and b_h are low in reliability, or b_w and b_h are measures of unrelated traits. Factor analysis (varimax rotation) of the 42 measures (w and h combined) showed that one factor accounted for 75.3% of the common variance with each of the 42 measures correlating .68 to .91 with the factor. Thus, it would appear that both w and h reliably measure a common trait; evidently this trait is highly related to general mental status. Thus, we may speculate that the changes measured on w and h are likely to be perceived changes of mental status. The two change scores did not correlate highly with average criterion status (.20 and .03) as would be expected since they correlate so lowly with each other. This line of reasoning then leads to the conclusion that the change parameters failed to correlate highly because of the lack of relationship between the (perceived) improvement in mental status while at work and while on the ward. That is, the multidimensionality of improvement problem is probably the culprit as in many former studies, not the unreliability of measurement. Hence, such double-measurement studies using the present model are best done by using pairs of truly replicate measures supported by Monte Carlo study. The study also further illustrates that because of the multidimensionality problem, it is often difficult to achieve construct validity thereby making it necessary to rely heavily, even sometimes solely, on consistency testing or intra-model validity procedures and supporting Monte Carlo study.

IV. A Summation

(a) Regression analysis can be applied to each individual's data separately, but assumption departure should be assessed by consistency

testing not just by a priori considerations.

(b) Consistency tests when Monte Carlo tested should help prevent being mislead by spurious or inaccurate results. The Monte Carlo method can be used to evaluate the consistency tests, estimation accuracy, effects of unreliability, the optimal spacing of measurements and the like.

(c) Idiographic analysis may lead to more useful results than nomothetic analysis; this question can only be answered by empirical study. Such analysis is accomplished by use of a model for the individual or a unique criterion for the individual or both.

(d) The results of the point earning study are sufficiently encouraging to warrant further study of the two-line regression model for lithium treated manic patients.

(e) The model developed here is just an illustrative example, it could be modified to accomodate, say, dichotomous data, intersecting regression lines at the day at which treatment changes (such as starting, stopping, or changing the drug), non-linear trends; more than one criterion and the like.

(f) Finally, it is claimed that clinical intuition and empirical results strongly indicate that the commonly used nomothetic analysis of variance models typically lack adequate verisimilitude and thereby provide inaccurate, spurious, and uninteresting results in the evaluation of treatment effects. It is clear that specially developed psychometric models are required for the evaluation of treatment effects on an individual's behavior; the present two-line regression model is one attempt in this kind of model development.

References

Bergin, A.E. and Garfield, S.L. Handbook of psychotherapy and behavior change. New York: Wiley, 1971.

Chassan, J.B. Research design in clinical psychology and psychiatry. New York: Appelton-Century-Crofts, 1967.

Golden, R.R., Tyan, S.H., and Meehl, P.E. Detecting latent clinical taxa, IX: A Monte Carlo method for testing taxometric theories. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota (Report No. PR-74-7), Minneapolis, 1974.

Harris, C. (ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Kiesler, D.J. Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 1966, 65, 110-136.

Lindgren, B. Statistical theory. New York: Macmillan, 1962.

Lykken, D.T. Parametric versus structural variation in psychological theory and research. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota (Report No. PR-75-3), Minneapolis, 1975.

Thompson, T., Heston, L. and Golden, R. A behaviorally-based method for evaluating lithium effects in manic patients. In press.