Cutoff Assignment Strategies for Enhancing Randomized Clinical Trials

William M.K. Trochim, PhD, and
Joseph C. Cappelleri, PhD

Cornell University, Ithaca, New York 14853

ABSTRACT: The randomized clinical trial (RCT) is the preferred method for assessing the efficacy of treatments. Recent ethical and logistical criticisms suggest that new variations of the traditional RCT are needed. Some of these criticisms may be addressed with new hybrid designs that combine random assignment with assignment by one or more cutoff values on a baseline variable (e.g., severity of illness). In a simple version of such a "cutoff-based" RTC, persons scoring below a cutoff score on a baseline measure (i.e., the least severely ill) are automatically assigned to the control-treated group, those scoring above a second, higher cutoff (i.e., the most ill) are automatically assigned to the test-treated group, and those scoring in the interval between the cutoff scores (i.e., the moderately ill) are randomly assigned to either group. Depending on the baseline score, the patient is assigned to treatment either randomly or by the need-based, clinically related baseline score. Six cutoff-based design variations are studied via simulations and compared with the traditional RCT and the single-cutoff (i.e., regression-discontinuity) design. All variations yield unbiased estimates of the treatment effect but estimates differ in efficiency, with the RCT being most efficient and the single-cutoff design being least efficient. Secondary analyses of data from the Cross-National Collaborative Study of the Effects of Alprazolam (Xanax) on panic are conducted for each variation by selectively discarding cases from the original dataset to simulate cutoff-based assignment. The results confirm the simulations and illustrate how cutoff-based designs might look with real data.

CUTOFF ASSIGNMENT STRATEGIES FOR ENHANCING RANDOMIZED CLINICAL TRIALS (RCTs)

Randomized clinical trials (RCTs) are consistently recognized as the preferred method for assessing the effects of treatments [1,2]. Miké [1, p. 132] states that "statisticians stand firmly for the principles of randomization and the randomized clinical trial" and that "among clinical investigators there is a much broader scope of response, although the majority strongly support the need for randomization." Green [3, p. 192] cites three major advantages of randomized trials: "(1) bias in assigning treatments is avoided, whether the bias be conscious or unconscious—and much of it is unconscious; (2) prognostic factors, both known and unknown, tend to be balanced across

Address reprint requests to: William M.K. Trochim, Department of Human Service Studies, Room N136C MVR Hall, Cornell University, Ithaca, NY 14853.

0197-2456/92/$5.00

Controlled Clinical Trials 13:190–212 (1992)

655 Avenue of the Americas, New York, New York 10010

treatment groups; and (3) randomization guarantees validity of statistical tests of significance used to compare treatments." In a typical RCT, persons who have met screening criteria are randomly assigned to either the test treatment under investigation or to a control-treated group. Usually, some baseline measurement is gathered and, following treatment, outcome measures are recorded. Random assignment assures initially probabilistically equivalent groups, enabling outcome differences to be attributed to treatment rather than other factors [3, pp. 190,192].

In recent years, criticisms of RCTs have been raised, most related to the ethics and timing of clinical trials. Critics argue that RCTs often take too long to complete—patients may be denied a potentially beneficial treatment while awaiting RCT results. Recent controversies regarding the ethics of implementing RCTs in extracorporeal membrane oxygenation (ECMO) in neonatal intensive care, AIDS, and cancer [4–6] focus on the long time span typically required to complete the study. Miké states: "National debates concerning RCTs—when to terminate the trial and make treatments available for non-research patients—are prominent also in other major areas, including life-threatening diseases such as AIDS and cancer" [4, p. 154]. Critics [7,8] also question the ethics of RCTs that were conducted even though the test treatment was already thought to be effective. RCT proponents argue that a treatment cannot be considered effective with any reasonable degree of scientific credibility until an RCT is conducted [9]. Philosophers have also questioned the ethics of RCTs. Marquis states that "randomized clinical trials as presently conducted are unethical" [10]. At least one entire journal issue has been devoted to examining ethical issues in clinical trials from a philosophical perspective [11]. Miké defends the RCT, arguing that many of the ethical concerns of philosophers rest on misunderstandings of the statistical principles of randomization and significance testing [12]. Ethical concerns have also been raised by clinicians [1]. Ethical and timing issues in clinical trials include when to stop an ongoing trial; how to deal with diseases that progress slowly or treatments that require long time periods; and how to accumulate sufficient sample sizes for diseases that occur infrequently. Issues like these have led to a number of enhancements to the traditional RCT including methods to increase rates of subject recruitment, reduce the time required to complete trials by use of multiple sites, combine phases of traditional multiphase strategies, and develop models for stopping an RCT early once a clear trend in efficacy has been established [9]. Still, there appears to be a consensus that RCT methodology needs to be developed further to better address some of these concerns [13].

This article describes a strategy for augmenting traditional RCTs using cutoff values on baseline variables in conjunction with random assignment. Throughout this article, designs that use baseline cutoff values for treatment assignment are termed *cutoff-based designs*, and any such designs that also use randomized assignment across part of the baseline range are labeled *cutoff-based RCTs*. The RCT must remain the method of choice when efficacy questions are investigated. But when some of the concerns described above exist, cutoff-based RCTs may have advantages over the traditional RCT. This article presents the rationale for cutoff-based RCTs, computer simulations of several

important variations, and illustrative secondary analyses showing how different cutoff-based variations work on data from an actual RCT study.

REVIEW OF CUTOFF-BASED ASSIGNMENT

Cutoff assignment strategies can be traced to Campbell's work [14,15] on the "regression-discontinuity" (RD) design, a single-cutoff quasi-experimental design that involves no random assignment. The RD design got its name from the "jump" or discontinuity at the cutoff in the regression line of baseline and follow-up scores that occurs when there is a treatment effect. The design deliberately creates treatment groups that are "nonequivalent" on baseline measures. But this nonequivalence should not be confused with what arises in uncontrolled (e.g., historical) studies because in the RD design the cause of the nonequivalence is known (the cutoff assignment rule) and can be adjusted for in the analysis. The design does assume an equivalence between groups in the null case—an *equivalence of baseline-outcome functional form*, not of baseline level. The null case expectation is that there are no discontinuities in the baseline–outcome relationship that coincide with the cutoff placement.

A hypothetical RD design is shown in Figures 1a and b based on simulated data with the cutoff at a baseline score of 50. Figure 1a shows the null case—the bivariate distribution expected if the test treatment doesn't work (or is never administered). The solid line through the distribution is the linear regression of the outcome measure onto the baseline measure. Figure 1b shows a hypothetical treatment effect of 5 scale units; the outcome scores of the treatment group (those scoring above the cutoff) are lowered by an average of 5 points from where they would be in the null case. The dashed line shows the expected regression function in the null case, while the solid lines show the observed regression for the case of a 5-point treatment effect.

The single-cutoff design has many variations [16,17]. Baseline and outcome measures may be the same or different, the cutoff can be placed anywhere along the baseline measure (as long as there are sufficient numbers in the control-treated group), positive–negative directionality can be in either direction for either variable (e.g., in the figures, a "positive" treatment effect is evidenced in a *lowering* of outcome values relative to the null case expectation), and the outcome measure can be dichotomous or continuous (a logistic regression analysis would be performed on a binary outcome measure).

During the 1960s and 1970s investigations of this single-cutoff RD design [18–25] demonstrated that it yielded unbiased estimates of the treatment effect and explored alternative statistical analyses for several special cases. The design was used extensively in the mid-1970s to evaluate compensatory education programs [26,27] and methodologists began to address common implementation difficulties [16,26]. There have been few uses of the RD design outside of education, most notably in criminal justice [28,29] and the evaluation of the NIH Career Development Awards [30]. Trochim [16] synthesized the work on the RD design, describing major design variations, analytical options, and implementation issues.

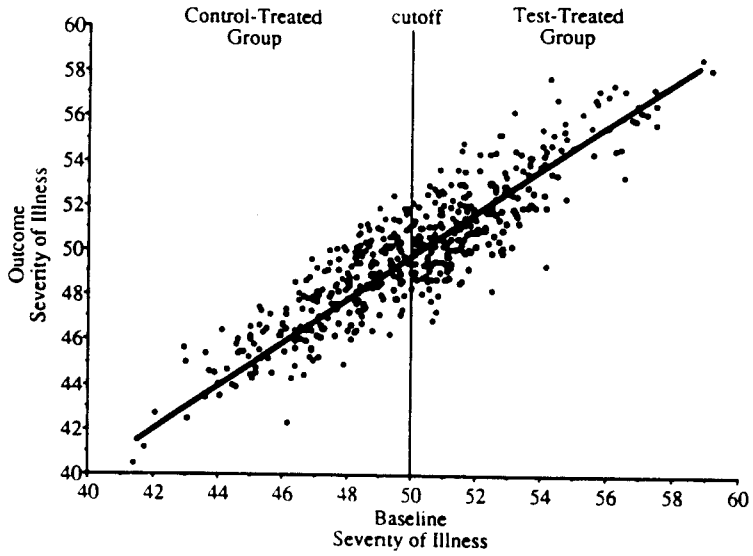

Figure 1a A simulated regression–discontinuity cutoff design with no treatment effect.

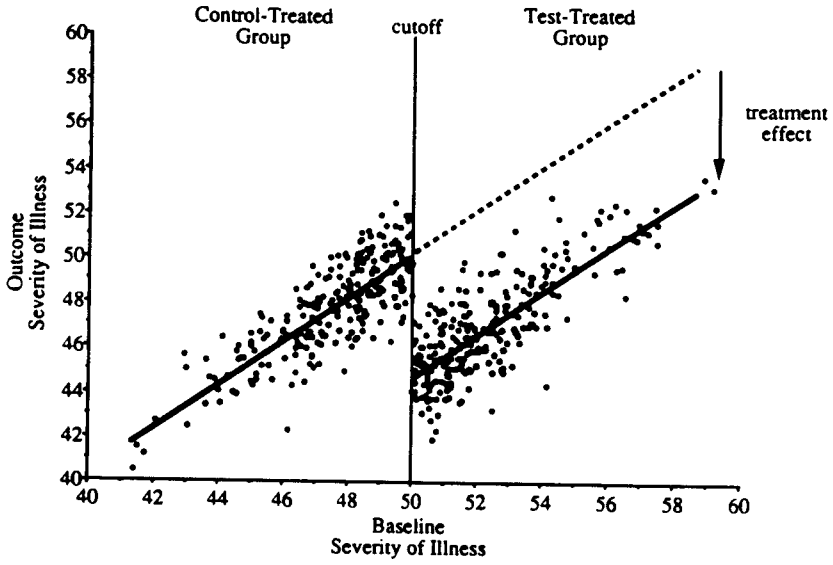

Figure 1b A simulated regression–discontinuity cutoff design with a 5-point treatment effect.

Coupling cutoff and randomized assignment is an RD variation discussed in Campbell's early expositions [31]. Boruch [2] described a simple one-interval and a more complex multi-interval cutoff-based RCT. Boruch describes two studies [32,33] that illustrate the feasibility of implementing cutoff-based as-

signment in medical research—in both studies, persons were assigned to conditions based on whether they were critically ill, marginally ill (random assignment of these cases), or negligibly ill. Rubin [24,25] shows that the treatment effect estimate is unbiased for both the single-cutoff RD design and the cutoff-based RCT variation provided the correct baseline–outcome functional form is specified. Cappelleri [34] examines the theoretical and empirical properties of cutoff-based designs, with and without randomization, in the context of controlled clinical trials.

Recently, cutoff-based designs have received more attention from medical researchers. Trochim [17] presents an overview of potential applications of cutoff designs in health settings. Fineberg [35] describes the single-cutoff design as "an intriguing method for unbiased assignment . . . well worth further theoretical and empirical work," Mosteller [36] suggests that "the method needs to be tried out more," and Luft [37] describes it as a "worthy addition to the set of tools available to the health services researcher." Williams [38] has a more mixed reaction, pointing out that the single-cutoff design should not be considered a replacement for randomized experiments and that there may be serious losses in statistical efficiency when moving from an RCT to an RD strategy. Robbins and Zhang [39] provide statistical arguments that under certain assumptions one can obtain an unbiased estimate of the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug, essentially the single-cutoff RD design. Cutoff designs have not been utilized in health or medical research settings, although there is at least one example of a single-interval cutoff-based RCT currently underway to study the relative efficacy of inpatient vs outpatient treatments for cocaine dependency [40].

The statistical analysis of cutoff-based RCTs depends on the same set of assumptions used for multiple-regression models in general, with the same set of remedial measures for violations in the assumptions. To model the baseline–outcome relationship correctly, Rubin [24] recommends using strong a priori information, Boruch [2] a "dry run" approach, Trochim [16,17] a polynomial regression approach, and Mohr [41] recommends using double-baseline measures. Undoubtedly, the best alternative is to combine cutoff assignment with as much randomization as possible.

Cutoff-based designs have not been without critics (see Refs. 16 and 42 for detailed discussions). Williams [38] questioned the applicability of the RD design that did not use any random assignment. He suggests that the design, rather than improving the ethical situation, may actually be less ethical than an RCT that requires smaller samples:

> More patients will have to be included in an RD design than in a randomized clinical trial. If the drug is eventually found safe and effective, more patients will have been denied optimal care in an RD design than in a randomized clinical trial. If the drug is found to have unacceptable side effects for the level of effectiveness, more patients will have been exposed to the risk of side effects in an RD design than in a randomized clinical trial. Either way, more patients will be given the wrong therapy in an RD design than in a randomized clinical trial. (p. 148)

Williams's point is that *after the study* more people will have been assigned

to the less effective treatment in an RD design than in an RCT. Williams's arithmetic is technically correct—because more persons are needed in both treatments in a cutoff-based design than in a traditional RCT, there have to be more people receiving the ultimately less efficacious (i.e., "wrong" in his terms) treatment, whichever that is. But Williams's critique doesn't consider *which patients* are subject to the "wrong" treatment under different possible outcomes—some are, by the nature of the level of their illness, in greater need of treatment and more willing to assume potential risks—a point that clinicians and patients often try to bring home to statisticians and methodologists. Cutoff designs explicitly take this clinical and ethical reality into account in a way that does not lead to biased estimates of treatment effect. Furthermore, cutoff designs, when coupled with random assignment, offer the potential of at least some of the benefits of both.

Other critics [43] claimed that random measurement error on the baseline variable biases estimates of the (main) treatment effect in the RD cutoff design. They state, "Even if we know perfectly the selection process and explicitly and correctly incorporate it into our evaluation model, we can still expect bias because, as a practical matter, the independent variables are always observed imperfectly" (p. 177). Their argument is incorrect, and they subsequently retracted it [44]. Assignment in cutoff-based designs is by *observed* baseline score, not by underlying construct (whatever that may be), and it is perfect knowledge of the assignment rule that enables an unbiased treatment effect estimate in both RD and cutoff-based RCTs, even in the presence of random measurement error. Logical, statistical, and simulation evidence [21,24,45–48] shows that RD and cutoff-based RCTs behave just like RCTs when there is measurement error in the covariates.

Cutoff-based RCTs constitute a middle ground between cutoff-based designs that use no random assignment (i.e., RD) and the traditional RCT that uses no cutoffs in assignment. Like all cutoff-based designs (including those with no random assignment) they allow a treatment to be given to those most in need or most willing to assume risk. At the same time they capitalize in part on the greater statistical efficiency of the traditional (no-cutoff) RCT. The major benefit is the ability to assign patients based on objective, clinically relevant factors. The major negative trade-off is that cutoff-based RCTs always require larger samples than a traditional RCT. Cappelleri [34] discusses statistical power and efficiency for a variety of cutoff-based designs, with and without some randomization, and gives sample size estimation procedures.

Cutoff-based designs have several potential advantages in addition to ethical ones. For instance, Luft [37] suggested that (nonrandom) RD designs, although less statistically powerful, may enable larger sample sizes:

> Thus, a well designed RCT often considers for randomization only those people in the midrange of some pretreatment [baseline] criteria, such as patients classified as mild hypertensives rather than normotensives or those with high blood pressure. This narrow focus may necessitate a longer period to accrue enough subjects or require complex multicenter collaborations. Furthermore, the narrow range of subjects makes it difficult to determine covariates that may enhance or

reduce the treatment effect. The RD approach, in contrast, allows the inclusion of a much broader range of subjects, possibly counteracting the reduced power of the design compared with the RCT with a lower cost per subject. Much more work is needed to examine the costs of achieving equally credible results using alternative designs under various situations. (p. 141)

In most RCTs, even after eligible patients have been selected, there are dropouts after enrollment. If attrition is related to the treatment, bias is likely to result. Even small percentages of treatment-related dropouts can degrade the initial equivalence between groups and lead to biased estimates of treatment effect. In theory, RCTs may be expected to have greater treatment-related attrition because the treatment is not assigned on clinically sensible grounds. Patients who are less sick may be more likely to drop out due to deleterious side effects of the new treatment. Cutoff-based designs may reduce this dropout rate because treatments are assigned based on clinically related need. Analyses by "intent to treat" can help eliminate some of this bias only if outcome information can be obtained on the dropouts. Even so, intent-to-treat analyses are inherently conservative and yield estimates that are lower in statistical efficiency, thus reducing some of the efficiency advantage of the traditional RCT. In a similar vein, when an illness is extreme or prior treatments have failed, the sickest patients may be expected to seek the new experimental treatment even when they are assigned to the control condition. This appears to be happening in current RCTs of AZT treatment of AIDS [49]. Green et al. [13] write about the need to speed up the completion of the RCT because of the impatience of the subjects and the possibility that physicians may remove severely ill patients from the control-treated group of the trial in order to address their clinical needs. Cutoff-based RCTs may help to alleviate some of this treatment switching without giving up randomization altogether.

Cutoff-based designs may also allow the RCT strategy to be extended into earlier (i.e., phase II or phase I) clinical trial stages. Green et al. [13] point out that in earlier phases it is common practice to use the sickest patients on whom other treatments have failed. There is a need to investigate strategies that allow increased sample sizes in phase I and II trials. Cutoff-based designs accomplish this with greater validity than do nonrandomized alternatives. Similarly, combining phases may speed up completion of trials. A cutoff-based RCT may be a more feasible phase II–III combination than the traditional RCT approach alone.

SIMULATIONS OF CUTOFF-BASED RCTS

Simulation Models

Simulations are conducted for six cutoff-based RCT variations that might be useful in medical and health research settings. Simulations are also conducted of the traditional RCT and the single-cutoff RD designs to compare them with the cutoff-based RCT variations. The simulations are not meant to suggest that these are the only or even the best ways to implement cutoff-based RCTs. Instead, they show potentially useful alternatives that illustrate some of the key principles of combining these design features in medical

research contexts. The reader may extrapolate from these designs to others of relevance.

All of the simulation models share some important properties and limitations. They are all of pretest–posttest designs, i.e., the same variable is assumed to be measured before and after treatment. They are also all two-group test treatment vs control treatment designs. In the six cutoff-based RCT models 25% of all cases are randomly assigned, the overall probability of random assignment to treatment is 50%, and the overall rate of assignment to treatment (i.e., across both randomized and cutoff-based groups) is also 50%. Multiple-site or multiple-stage designs assume that there are no site or time effects. Only linear first-order relations are modeled and no interaction effects are constructed. In all simulations, a simple true score model is assumed. Let the subscript i denote the ith observation. In each simulation, a true baseline score (T_i) value for each case is generated such that:

$$T_i \sim \text{NID}\ (\mu = 50, \sigma^2 = 9)$$

where NID represents normally and independently distributed. Then, separate error measures are created for the observed baseline (x_i) and outcome (y_i) measures such that:

$$e_x \sim \text{NID}\ (\mu = 0, \sigma^2 = 1)$$

$$e_y \sim \text{NID}\ (\mu = 0, \sigma^2 = 1)$$

In accordance with the simple true score model, the baseline variable actually used for assignment is then constructed:

$$x_i = T_i + e_{x_i}$$

The treatment group dummy-coded (0, 1) assignment variable z_i is constructed differently for each design as described below. Finally, the outcome measure is constructed using the formula:

$$y_i = T_i + (-5)z_i + e_{y_i}$$

Given these specifications, the mean for the baseline measure is 50 and the standard deviation is 3.1625. The null case outcome measure would also have this distribution. The reliability is 0.90 for the observed baseline measure (x_i) relative to the "true," error-free baseline measure (T_i):

$$\text{rel} = \frac{\text{var}(T_i)}{\text{var}(T_i) + \text{var}(e_{x_i})}$$

$$= \frac{9}{9 + 1}$$

$$= 0.9$$

In the null case, the x_i and y_i variables are assumed to be continuous normal measures of the severity of illness level, where higher scores indicate greater illness. The simulations create a "positive" or efficacious outcome through a 5-point reduction of severity of illness for the treated cases. For each of the eight models, five simulations are run for sample sizes of 100, 200, 300, 400,

and 500 cases (i.e., patients). Each simulation setup is run 1000 times. Thus there are 8 (models) × 5 (sample sizes) × 1000 runs = 40,000 separate simulation runs in the study.

Table 1 shows cutoff values for each model in both z-score and baseline score units. For designs with multiple cutoff intervals the total percentage of cases falling within each interval is given; the percentage of cases assigned to treatment is shown in the last column. If the line in the table describes an *interval* (i.e., a cutoff range is specified), the last column shows the percentage of cases *within the interval* assigned to treatment. For RCT and RD designs the percentage shown is across all cases. Descriptions and simulation specifications for each of the eight models are as follows.

Model 1: The Randomized Clinical Trial (RCT). This model is a traditional RCT with 50/50 probability of being assigned to either group. The assignment variable z_i is constructed by generating a normal variable a_i such that

$$a_i \sim \text{NID}(0, 1)$$

and then dichotomizing this variable at the mean of zero:

Table 1 Cutoff Specifications, Expected Percentage of Cases Within Intervals, and Expected Percentage of Cases (Within Interval) Assigned to Treatment for the Eight Simulation Models

	Model	Z-Score Cutoff(s)	Baseline Cutoff(s)	% in Interval	% in Treatment
1.	The randomized clinical trial (RCT)	—	—	—	50
2.	Single-cutoff regression–discontinuity (RD)	0	50	—	50
3.	Single-cutoff interval	−0.318 to +0.318	48.99440 to 51.00560	25	50
4.	Different proportions in the randomization interval	−0.318 to +0.318	48.99440 to 51.00560	5	25
		−0.318 to +0.318	48.99440 to 51.00560	5	33
		−0.318 to +0.318	48.99440 to 51.00560	5	50
		−0.318 to +0.318	48.99440 to 51.00560	5	66
		−0.318 to +0.318	48.99440 to 51.00560	5	75
5.	Increasing probability of random assignment	−0.318 to −0.189	48.99440 to 49.40233	5	25
		−0.189 to −0.062	49.40233 to 49.80394	5	33
		−0.062 to +0.062	49.80394 to 50.19606	5	50
		+0.062 to +0.189	50.19606 to 51.59767	5	66
		+0.189 to +0.318	50.59767 to 51.00560	5	75
6.	Different interval widths	−0.419 to +0.419	48.67501 to 51.32499	32.5	50
		−0.351 to +0.351	48.89004 to 51.10996	27.5	50
		−0.285 to +0.285	49.09875 to 50.90125	22.5	50
		−0.221 to +0.221	49.30114 to 50.69886	17.5	50
7.	Different interval centers	−1.688 to −0.688	44.66208 to 47.82435	20	50
		−1.228 to −0.228	46.11672 to 49.27900	30	50
		+0.228 to +1.228	46.11672 to 49.27900	30	50
		+0.688 to +1.688	52.17565 to 55.33792	20	50
8.	Single cutoff with a "model check" randomization interval	0.318	51.0056	—	50
		−1.462 to −0.462	45.37675 to 48.53903	25	50

$z_i = 1$ if $a_i > 0$

$= 0$ otherwise

where a value of 1 indicates assignment to the treatment group and 0 indicates assignment to the placebo-treated group.

Model 2: The Regression–Discontinuity Design (RD). This model assumes that there is a single cutoff at the mean of the normally distributed baseline variable that is used for assignment. Since the baseline measure has a mean of 50, the assignment variable is constructed so that:

$z_i = 1$ if $x_i \geq 50$

$= 0$ if $x_i < 50$

That is, patients with a severity of illness score of 50 or greater are assigned to the new treatment; those with less than 50 are assigned to the placebo-treated group.

Model 3: Single-Cutoff Interval. Both of the prior models are likely to meet with some resistance in clinical settings. The random assignment in model 1 may be rejected where it is seen as denying a potentially efficacious treatment to needy patients. Physicians may disagree with the cutoff assignment of "close call" patients in model 2. This design is a simple cutoff-based RCT compromise that combines cutoff and random assignment. In model 3, two cutoff values, arranged symmetrically around the baseline mean, define the randomization interval. All cases scoring above that interval on the baseline measure are assigned to the new treatment, while those below the interval are assigned to the control-treated condition. The design is shown in Figure 2a.

Model 4: Different Proportions in the Randomization Interval. The cutoff interval in model 3 can be constructed so that a specific percentage of all persons will be assigned to each treatment (e.g., 50/50), but there is no guarantee that the assumptions of the cutoff choices will hold up over time. If the average baseline illness levels get higher or lower, disproportionately more persons may be assigned to one treatment than the other. For hospitals, this can lead to cost inefficiencies and census problems—at times there may be too many eligible cases or empty beds. In an RCT, once the probability of assignment is set, the only thing that affects the number of cases in each group is the overall rate at which eligible patients come into the system, but the proportions in each group will approximate the probability of assignment. Cutoff-based designs are thus more likely to lead to unexpected disproportionate numbers in treatment groups. One way to address this problem for model 3 is to alter the probabilities of random assignment during the course of the trial. If there are too few test treatment cases during a specified time period, the random assignment proportions can be altered for the next period to increase the ratio of test-treated to control-treated patients, e.g., from a 50/50 ratio to a 66/33 ratio. Changing the probabilities of assignment for the randomized cases should not lead to bias but may decrease efficiency of estimates of treatment effect. Model 4 (identical in cutoff structure to model 3) examines this adaptive random assignment probability scenario. A second

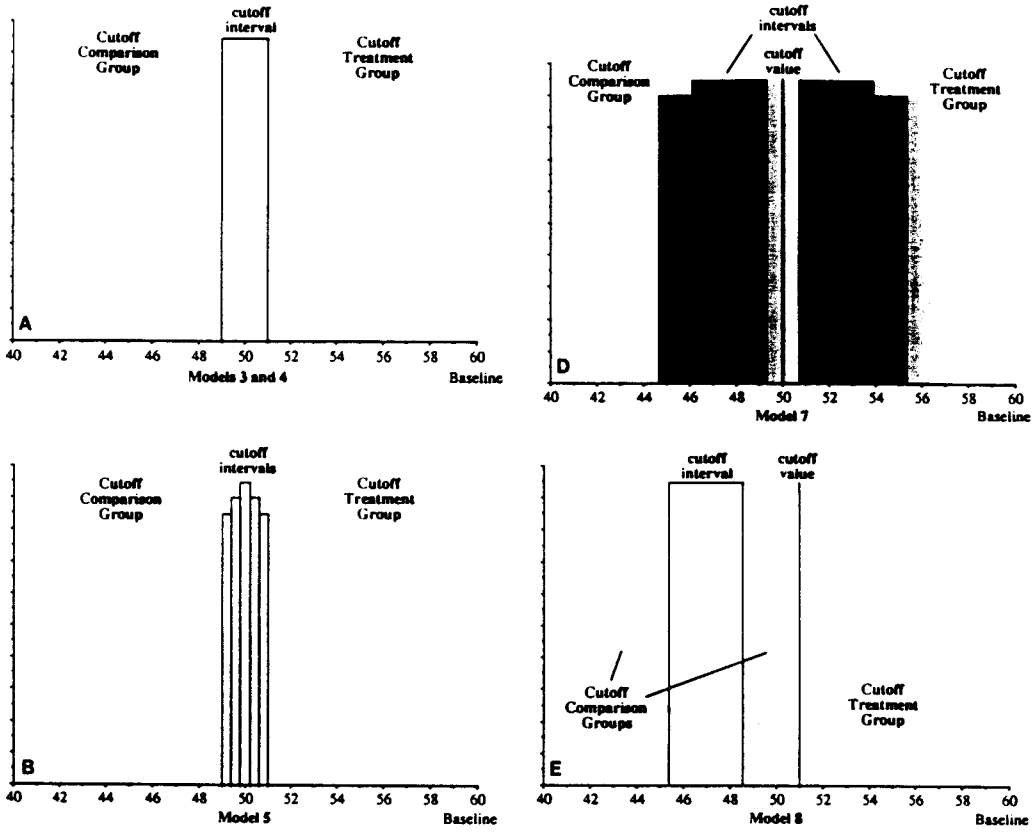

Figure 2a Model 3 (the single-cutoff interval design) and model 4 (the cutoff-based RCT design with different random assignment proportions in the randomization interval).
Figure 2b Model 5—The cutoff-based RCT design with increasing probabilities of random assignment in the randomization interval.
Figure 2c Model 6—The cutoff-based RCT design with different randomization interval widths.
Figure 2d Model 7—The cutoff-based RCT design with different randomization interval center.
Figure 2e Model 8—The single-cutoff RD design with a "model check" randomization interval.

application of model 4 occurs when there are multiple test sites, each with its own census needs. Some hospitals may have more beds or treatment facilities available. To maximize efficient use of resources, a different probability of within-interval random assignment to treatment may be used at each site. The simulations test a broad range of within-interval assignment probabilities—five different proportions of random assignment to treatment (25, 33, 50, 66, and 75%). This is likely to exceed any realistic range of probabilities. For the single-site adjustments over time scenario, this simulates an annual

adjustment to the assignment proportion over the course of a 5-year project. For the multisite scenario, this illustrates the case of five sites, each having a different probability of random assignment. This model has the same two cutoffs as model 3, but within this interval, cases are randomly divided into five subgroups each with different probabilities or percentages of random assignment to treatment (namely, 25, 33, 50, 66, and 75%).

Model 5: Increasing Probability of Random Assignment. When cutoffs are used, physicians and patients may be concerned that "close call" cases are incorrectly assigned. Model 3 attempts to address this by introducing a randomization interval, but even so, concerns may be raised about the abrupt change in probability of assignment that occurs at the interval cutoff boundaries. These sharp jumps can be reduced by having a more gradually changing probability of assignment within the randomization interval. Model 5 is identical to model 3 except that the interval is subdivided into five adjacent subintervals with gradually increasing probabilities of assignment across them. Treatment is given to all those who are most severely ill (above the interval), with probability of assignment to treatment *declining* as severity of illness declines, reaching a probability of zero for those patients below the randomization interval. This model is shown in Figure 2b.

Model 6: Different Interval Widths (at Different Sites or Times). Random assignment to treatment is used for scientific purposes, not necessarily because it is clinically sensible. Over the course of a clinical trial, it may be possible and desirable to alter the proportion of randomly assigned cases if preliminary results indicate that the treatment appears to be efficacious. As evidence for the efficacy of the new treatment increases, the width of the interval might periodically be reduced to allow more sick patients to receive the apparently more efficacious treatment. In the extreme, this variation moves from a traditional RCT to a single-cutoff RD by periodically reducing the range of random assignment. In model 6 (shown in Figure 2c) four different random assignment interval widths are used. The study might begin with the widest interval and reduce the width after periodic assessment. Model 6 also describes a multisite study where different width random assignment intervals are negotiated at each site. Some sites, for instance, might readily accept the legitimacy of random assignment across all or much of the baseline measure range, while others may be more reluctant. Because each of the four randomization intervals in model 6 have different widths, they include different percentages of the total number of cases.

Model 7: Different Interval Centers (at Different Sites or Times). Another way to adjust the proportions assigned to each treatment is to move the random assignment interval up or down on the baseline measure at different times in the study or to place the intervals at different baseline measure locations for different sites. Model 7 (shown in Figure 2d) uses four random assignment intervals placed at different points along the baseline continuum. This might be used if there are multiple treatment sites that have different admission rates and/or numbers of treatment slots. For instance, one facility may have more surgeons who can perform a novel treatment and thus might be able

to have a higher proportion of cases assigned to that condition. Each site could have a different randomization interval location to allow a site-specific sample size that better meets their administrative needs. In the single-site variation, this design might help when the availability of resources or patient demand for treatment changes over time or the numbers of eligible patients presenting is subject to fluctuation. Model 7 takes the mean (or median) baseline value of 50, which is the average of the four interval centers, as the point at which the main effect is estimated.

Model 8: Single Cutoff with a "Model Check" Randomization Interval. This design is the single-cutoff RD design with a randomization interval located at a separate place from the cutoff value. Its major purpose is to provide empirical verification for the assumed regression model—cases in the interval allow us to verify regression line extrapolations from cutoff-based groups. In other randomization interval designs, such as models 3 through 7, it is legitimate to be concerned with whether the projection of the placebo-treated patient group line into the region of the treated group is an accurate reflection of its null case expectation. The cutoff interval in model 8 makes it possible to "check" or verify that the regression line projections from the single-cutoff design have some empirical basis. In these simulations the interval is in the placebo-treated group range, although placing the interval in the treatment group range would not change the conclusions. This model, shown in Figure 2e, is the only true asymmetrical model of the eight. The treatment effect is estimated at the median of the randomization interval, which is 46.95.

There are many other promising cutoff-based RCT variations that warrant investigation in future studies. Patients above a cutoff on severity of illness could be randomly assigned to the test treatment with all cases below the cutoff placed in the placebo-treated group. This below-cutoff group might consist of persons whose symptoms are not severe enough to be included in the traditional RCT, but who could be included under a cutoff-based design primarily to improve statistical power. This design could also be reversed, with all persons above a cutoff automatically receiving the treatment and all those below randomly assigned. Or, there could be an RCT interval in the center of a baseline distribution with all persons both above and below this interval assigned to the treatment. Only a relatively small subsample of the population would be denied the treatment in this variation. Or, cutoff strategies could be used in a staged or adaptive design. In the initial stage, all persons would be randomly assigned to treatment. If preliminary results indicate that the test treatment may be efficacious, a cutoff could be used to assign all new cases above the cutoff to the test treatment, while those below would be randomly assigned. As long as each intermediate analysis continues to point in the direction of efficacy, the cutoff value above which new persons would all be assigned to the test treatment can continue to be lowered. The study begins as a traditional RCT, moves in the intermediate stages to a cutoff-based RCT, and ultimately administers the test treatment to all eligible participants (full clinical adoption). Each of these variations has merit and should be studied further for their statistical and methodological properties and their ability to address situation-specific ethical and logistical concerns.

Statistical Analysis Model

All simulation models are analyzed using the same ANCOVA-based statistical model. Given the baseline measure x_i and outcome measure y_i, the general model used can be stated as follows:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 \tilde{x}_i + \hat{\beta}_2 z_i + e_i$$

where:

$\tilde{x}_i$ = baseline measure for individual i minus the value m (i.e., $\tilde{x}_i = x_i - m$)

where

for model 1 $m = 0$

for models 2, 3, 4, 5, 6, and 7, $m = 50$ (the median of the baseline scor

for model 8, $m = 46.95$ (the median of the randomization interval)

y_i = outcome measure for individual i

z_i = treatment group variable (1 if test-treated participant; 0 if control-treated participant)

$\hat{\beta}_0$ = intercept estimator

$\hat{\beta}_1$ = linear slope estimator

$\hat{\beta}_2$ = treatment effect estimator

e_i = sample regression disturbance term

The major null hypothesis of interest

H_0: $\beta_2 = 0$ (i.e., the treatment effect parameter is zero)

is tested against the alternative hypothesis that

H_1: $\beta_2 \neq 0$

The only difference in the analysis applied to each model is in the $\tilde{x}_i$ term. The value m is subtracted from the baseline measure in order to control the point at which the treatment effect estimate is made. In the single-cutoff RD design, for instance, the cutoff value is subtracted from each baseline score. This sets the cutoff equal to $\tilde{x}_i = 0$, i.e., it equates the cutoff and intercept. Since the key coefficient of interest $\hat{\beta}_2$ is estimated at the intercept, after the transformation, it is actually being estimated at the cutoff point. In fact, this is technically not necessary in these simulations since there is no interaction effect. The transformation is included here primarily because with real data—where they may be interaction effects or nonlinearities—it is important to be careful about where the main treatment effect is estimated [46]. The reader is encouraged to consult the technical literature [16,17,34] for more details regarding statistical models for cutoff-based designs.

Simulation Results

The key results for all models are shown in Table 2. The table shows average $\hat{\beta}_2$ values (i.e., treatment effect) and the average of their standard errors (SEs),

Table 2 Average Gain ($\hat{\beta}_2$) and Average SE (in parentheses) for Each Model for Five Different Sample Sizes (Each Average Based on N = 1000 Simulation Runs)

Model	N = 100	200	300	400	500
1. The randomized clinical trial (RCT)	−4.9904 (0.2783)	−5.0031 (0.1949)	−4.9985 (0.1591)	−4.9985 (0.1382)	−5.0015 (0.1236)
2. Single-cutoff regression–discontinuity (RD)	−5.0227 (0.4616)	−5.0145 (0.3246)	−4.9914 (0.2643)	−5.0064 (0.2292)	−5.0042 (0.2045)
3. Single-cutoff interval	−4.9916 (0.4273)	−4.9992 (0.3012)	−4.9879 (0.2446)	−4.9951 (0.2117)	−5.0028 (0.1900)
4. Different proportions in the randomization interval	−4.9854 (0.4280)	−4.9835 (0.3006)	−5.0029 (0.2449)	−5.0016 (0.2122)	−4.988 (0.1893)
5. Increasing probability of random assignment	−4.9922 (0.4433)	−4.9971 (0.3092)	−5.0122 (0.2532)	−4.9917 (0.2185)	−4.9857 (0.1958)
6. Different interval widths	−5.0149 (0.4264)	−4.9987 (0.2995)	−5.0053 (0.2444)	−4.9931 (0.2113)	−5.0018 (0.1887)
7. Different interval centers	−4.9949 (0.3186)	−4.9998 (0.2248)	−5.002 (0.1834)	−5.0025 (0.1587)	−4.9933 (0.1419)
8. Single cutoff with a "model check" randomization interval	−5.0082 (0.3302)	−4.9957 (0.2323)	−4.9922 (0.1891)	−4.9995 (0.1642)	−4.9961 (0.1462)

each average based on 1000 simulation runs. All models yield unbiased estimates of the true treatment effect of −5 outcome score units. The major difference is in efficiency and statistical power. Of course, as sample sizes are increased, SEs are reduced. But more interesting is the relative SE level across designs shown for the eight models in Figure 3. The single-cutoff RD and the traditional (fully randomized) RCT provide the upper and lower SE boundaries as expected. The designs cluster into three groups in terms of efficiency. The single-cutoff RD, the three designs having the same overall interval width (i.e., models 3, 4, and 5), and the design with varied interval widths (model 6) are comparably low in efficiency (i.e., higher SEs). The designs that have randomization across a wider baseline score range—by varying the interval center (model 7) or by placing the interval apart from a single cutoff (model 8)—are comparable to each other and have greater efficiency than the previous set even though the same percentage of the total cases is randomly assigned. Finally, the traditional RCT stands alone as the most efficient. There seems to be some efficiency gain when randomization intervals span a wider baseline score range, even when the same percentage of all cases is randomly assigned. Cappelleri [34] discusses power analysis and sample size estimation for cutoff-based designs with and without some randomization.

ILLUSTRATIVE SECONDARY ANALYSES

Although no cutoff-based RCTs have been implemented in medical research contexts, it is possible to use existing RCT data to construct such designs post hoc to examine how they perform in estimating treatment effects on real data. To do this, cutoff values are arbitrarily selected in previously

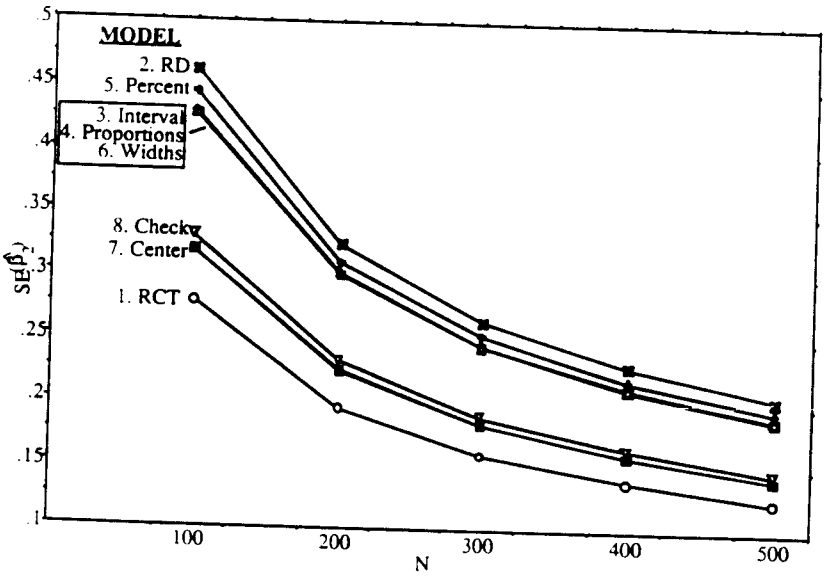

Figure 3 Average standard errors of treatment effect for eight models across five sample sizes.

conducted RCTs and used to discard data to simulate a cutoff-based design structure.

Data from the Cross-National Collaborative Study [50–54] of the Effect of Alprazolam (Xanax) on panic is used to illustrate the cutoff-based models described earlier. This was a two-phase, multinational, multicenter RCT to evaluate the drug treatment of panic disorder and associated agoraphobia. It is not the purpose of this reanalysis to assess efficacy across the range of possible outcome measures or even to develop a definitive substantive test of the efficacy hypothesis for these data. The major purpose of these analyses is to see how well treatment effect estimates from cutoff-based RCTs (created from the original data) compare with the traditional RCT (which uses data from the entire sample) in terms of bias and efficiency.

The reanalysis reported here is limited to the baseline and first study week. Only one of the many measures in the original study was used here—the Sheehan Clinician Rated Anxiety Scale (CRAS) [55], which consists of average clinician's ratings of 35 items (e.g., "spells of imbalance," "tires easily," "tension/nervousness/anxiety") rated on a 0 (Absent) to 4 (Very Severe) scale. All cutoff placements are done in z-score units and are designed to be comparable to the cutoff placements in the simulations. Cutoff specifications for the eight models are given in Table 3.

RESULTS OF ILLUSTRATIVE SECONDARY ANALYSES

The original baseline-outcome distribution for the CRAS variable is shown in Figure 4. Visual inspection of the figure indicates a likely positive treatment effect, i.e., a *reduction* in the average CRAS score for the Xanax patients relative

Table 3 Cutoff Specifications, Expected Percentage of Cases Within Intervals, and Expected Percentage of Cases (Within Interval) Assigned to Treatment for the Xanax Reanalysis

	Model	Z-Score Cutoff(s)	Baseline Cutoff(s)	% in Interval	% in Treatment
1.	The randomized clinical trial (RCT)	—	—	—	50
2.	Single-cutoff regression–discontinuity (RD)	0	1.233	—	50
3.	Single-cutoff interval	−0.318 to +0.318	1.069 to 1.397	25	50
4.	Different proportions in the randomization interval	−0.318 to +0.318	1.069 to 1.397	5	25
		−0.318 to +0.318	1.069 to 1.397	5	33
		−0.318 to +0.318	1.069 to 1.397	5	50
		−0.318 to +0.318	1.069 to 1.397	5	66
		−0.318 to +0.318	1.069 to 1.397	5	75
5.	Increasing probability of random assignment	−0.318 to −0.189	1.06885 to 1.13547	5	25
		−0.189 to −0.062	1.13547 to 1.20106	5	33
		−0.062 to +0.062	1.20106 to 1.26510	5	50
		+0.062 to +0.189	1.26510 to 1.33069	5	66
		+0.189 to +0.318	1.33069 to 1.39732	5	75
6.	Different interval widths	−0.419 to +0.419	1.01669 to 1.44948	32.5	50
		−0.351 to +0.351	1.05181 to 1.41436	27.5	50
		−0.285 to +0.285	1.08589 to 1.38027	22.5	50
		−0.221 to +0.221	1.11895 to 1.34722	17.5	50
7.	Different interval centers	−1.688 to −0.688	0.36131 to 0.87776	20	50
		−1.228 to −0.228	0.59888 to 1.11533	30	50
		+0.228 to +1.228	1.35084 to 1.86730	30	50
		+0.688 to +1.688	1.58841 to 2.10487	20	50
8.	Single cutoff with a "model check" randomization interval	−0.318	1.06885	—	50
		+0.462 to +1.462	1.47169 to 1.98815	25	50

to placebo-treated patients. For all models, all cutoff values are based on a baseline CRAS score with a mean of 1.233 and a standard deviation of 0.5165. For instance, for the single-cutoff RD design, a cutoff is constructed at the mean, i.e., at a *z* value of 0 or a CRAS value of 1.233. All placebo-treated cases falling above the mean and all Xanax-treated cases falling below the mean are discarded, yielding the baseline–outcome distribution shown in Figure 5. It is visually apparent that there is a "jump" or discontinuity in the baseline-outcome distribution at the cutoff point that is indicative of a treatment effect. The other models are created in an analogous manner.

The same ANCOVA regression model described for the simulations is applied to all of the models. The scatter plot of the outcome–baseline CRAS measure for all placebo-treated patients in the original RCT clearly depicts a linear relationship. Linearity is confirmed by an *F* test for lack of fit and by a residual analysis. Therefore, the specified ANCOVA model is deemed appropriate for the cutoff-based models. The results are shown for the eight models in Table 4. Results are given for two types of traditional RCTs. The

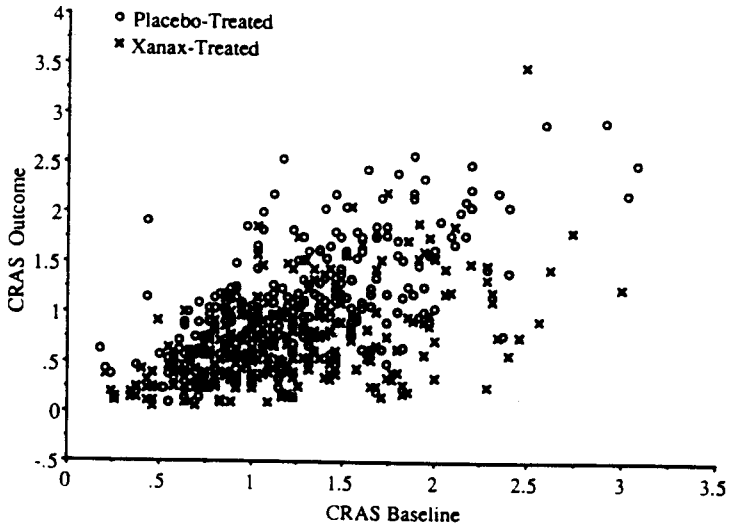

Figure 4 Baseline-outcome distribution (entire sample) of Clinician Rated Anxiety Scale from the Xanax Study.

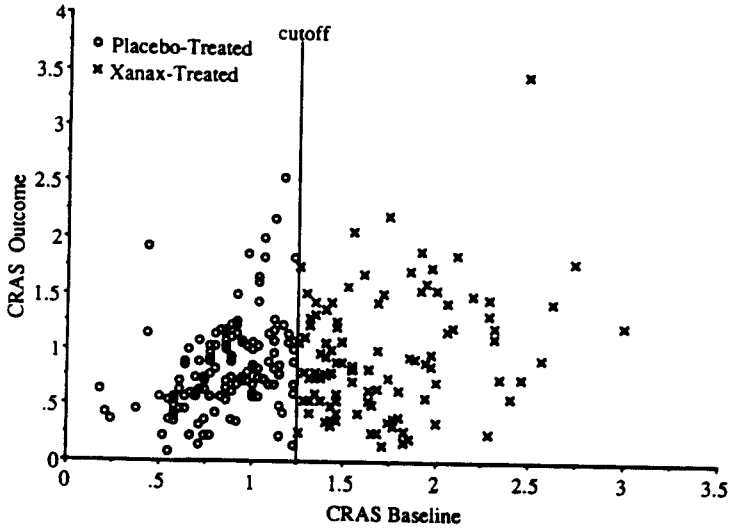

Figure 5 Baseline-outcome distribution of Clinician Rated Anxiety Scale from the Xanax Study using a single-cutoff RD design.

first uses the full-sample data from all patients who have both baseline and outcome measurements. The second (model 1) uses only half of the sample data (randomly discarding the other half), so that this RCT will have a sample size comparable to the seven cutoff-based models. The treatment effect estimate for the full-sample RCT is -0.368, with a standard error of 0.037. The major question of interest is how well the estimates from the eight models approach this benchmark. For all models the estimated effect falls within one

Table 4 Average Estimates of Treatment Efficacy ($\hat{\beta}_2$), Standard Errors, and t Values for the Full-Sample RCT and the Eight Models for the CRAS Variable from the Xanax Study

Model	$\hat{\beta}_2$	SE($\hat{\beta}_2$)	t Value
The randomized clinical trial (RCT)—entire sample	−0.368	0.037	−9.824
1. The randomized clinical trial (RCT)—random subsample	−0.370	0.053	−6.935
2. Single-cutoff regression–discontinuity (RD)	−0.294	0.103	−2.861
3. Single-cutoff interval	−0.330	0.089	−3.714
4. Different proportions in the randomization interval	−0.329	0.089	−3.707
5. Increasing probability of random assignment	−0.360	0.086	−4.191
6. Different interval widths	−0.362	0.083	−4.362
7. Different interval centers	−0.359	0.062	−5.802
8. Single cutoff with a "model check" randomization interval	−0.341	0.064	−5.293

standard error unit of the full-sample RCT model. As in the simulations, the major differences across models are related to their relative efficiency—the RCT design yields the most efficient estimate while the single-cutoff RD yields the worst. In both simulations and real data analyses, models 7 and 8 yield estimates nearly as efficient as the RCT design with about as many cases. These cutoff-based RCTs appear not to lose much in efficiency and, consequently, can use sample sizes nearly as small as the traditional RCT while achieving comparable statistical power. However, these models may be more difficult to implement or justify to nonspecialists than other cutoff-based RCTs and the traditional RCT.

The reanalysis shows that if cutoff-based RCT designs (or a single-cutoff RD design) had been used instead of the traditional RCT, unbiased and statistically significant estimates of treatment efficacy could have been obtained.

CONCLUSIONS

This study has some notable limitations. First, very little is known about how cutoff-based designs work in real-world clinical trials. They may be more difficult than the traditional RCT to communicate for informed consent, may not be as readily accepted by patients and physicians, and may require more data management effort from physicians and research staff. Second, the simulations did not include site or time effects. Work is currently underway to

examine more complex models. Third, only a limited number of design variations were included. Field tests of cutoff-based RCTs in real-world medical contexts are needed to examine these and other important implementation issues.

Randomized clinical trials are the best designs for assessing the efficacy of treatment. Cutoff-based RCT variations attempt to balance the need for scientific rigor with the pressures to provide treatments based on clinical factors. They include random assignment to treatment, but only for part of the enrolled population. They provide for assignment of patients to treatment based on clinically established need, but only for some of those enrolled. Assuming that the correct analytical model is applied, they yield unbiased estimates of treatment efficacy. They have lower statistical efficiency and power than a traditional RCT, but in some contexts they may better address the ethical concerns of some patients and clinicians. Cutoff-based RCTs should only be mounted after careful review of ethical issues. When there are strong ethical concerns about denying a test treatment to those most in need or giving it to those less in need, cutoff-based RCTs should be considered.

REFERENCES

1. Miké V: Clinical studies in cancer: A historical perspective. In Statistics in Medical Research: Methods and Issues with Applications in Cancer Research. Miké V, Stanley KE, Eds. New York, John Wiley and Sons, pp 111–155, 1982
2. Boruch R: Coupling randomized experiments and approximations to experiments in social program evaluation. Socio Meth Res 4:31–53, 1975
3. Green S: Patient heterogeneity and the need for randomized clinical trials. Controlled Clin Trials 3:189–198, 1982
4. Miké V: Suspended judgment: Ethics, evidence, and uncertainty. Controlled Clin Trials 11:153–156, 1990
5. Marshall E: Quick release of AIDS drugs. Science 245:346–347, 1989
6. Marx JL: Drug availability is an issue for cancer patients, too. Science 245:345–346, 1989
7. Beecher H: Ethics and clinical research. N Engl J Med 274:1354–1360, 1966
8. Veatch R, Sollitto S: Human experimentation: The ethical questions persist. Hastings Center Per 3:1–3, 1973
9. Gail MH: Monitoring and stopping clinical trials. In Statistics in Medical Research: Methods and Issues with Applications in Cancer Research. Miké V, Stanley KE, Eds. New York, John Wiley and Sons, pp 455–486, 1982
10. Marquis D: Leaving therapy to chance. The Hastings Center Per 13:40–47, 1983
11. Schaffner KF: Ethical issues in the use of clinical controls. J Med Phil 11:297–404, 1986
12. Miké V: Philosophers assess randomized clinical trials: The need for dialogue. Controlled Clin Trials 10:244–253, 1989
13. Green S, Ellenberg S, Finkelstein D, Forsythe L, Freedman L, Freeman K, Lefkopoulou M, Schoenfeld D, Smith R: Issues in the design of drug trials for AIDS. Controlled Clin Trials 11:80–87, 1990
14. Campbell DT, Stanley JC: Experimental and quasi-experimental designs for research. Chicago, Rand McNally, 1966
15. Cook TD, Campbell DT: Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston, Houghton-Mifflin, 1979

16. Trochim WMK: Research Design for Program Evaluation: The Regression-Discontinuity Approach. Beverly Hills, Sage, 1984
17. Trochim WMK: The regression-discontinuity design. In Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Sechrest L, Perrin P, Bunker J, Eds. Washington, DC, US Dept HSS, Public Health Service, Agency for Health Care Policy and Research, May 1990
18. Barnow BS, Cain GG, Goldberger AS: Issues in the analysis of selectivity bias. In Evaluation Studies Review Annual. Stromsdorfer EW, Farkas G, Eds. Beverly Hills, Sage, 1980
19. Sacks J, Ylvisaker D: Linear estimation for approximately linear models. Discussion paper #9, Center for Statistics and Probability, Evanston, IL, Northwestern University, October 1976
20. Spiegelman C: Estimating the effect of a large scale pretest-posttest social program. Proceedings of the Social Statistics Section, Am Stat Assoc, pp 370–373, 1979
21. Reichardt CS: The statistical analysis of data from nonequivalent group designs. In Quasi-Experimentation: Design and Analysis Issues for Field Settings. Cook TD, Campbell DT, Eds. Boston, Houghton-Mifflin, pp 147–206, 1979
22. Sween JA: The experimental regression design: An inquiry into the feasibility of nonrandom treatment allocation. Unpublished doctoral dissertation, Evanston, IL, Northwestern University, 1971
23. Sween JA: Regression-discontinuity: Statistical tests of significance when units are allocated to treatments on the basis of quantitative eligibility. Unpublished manuscript, Evanston, IL, Northwestern University, April 1977
24. Rubin DB: Assignment to groups on the basis of a covariate. J Educ Stat (2):1–26, 1977
25. Rubin DB: Bayesian inference for causal effects: The role of randomization. Ann Stat 6:34–58, 1978
26. Tallmadge GK, Wood CT: User's guide: ESEA title I evaluation and reporting system. Mountain View, CA, RMC Research Corp, 1978
27. Trochim WMK: Methodologically-based discrepancies in compensatory education evaluations. Eval Rev 6(4):443–480, 1982
28. Berk RA, Rauma D: Capitalizing on nonrandom assignment to treatment: A regression discontinuity of a crime control program. J Am Stat Assoc 78:21–28, 1983
29. Lipsey MW, Cordray DS, Berger DE: Evaluation of a juvenile diversion program: Using multiple lines of evidence. Eval Rev 6(3):283–306, 1981
30. Carter GM, Winkler JD, Biddle AK: An evaluation of the NIH research career development Award. Santa Monica, CA, Rand, 1987
31. Campbell DT: Reforms as experiments. Am Psychologist 24(4):409–429, 1969
32. Mather HG, Pearson NG, Read KL, Shaw DB, Steed GR, Thorne MG, Jones S, Guerrier CJ, Eraut CD, McHugh PM, Chowdhury NR, Jaffaray MH: Acute myocardial infarction: Home and hospital treatment. Br Med J 3:334–338, 1971
33. Sackett DL: End result analyses in a randomized trial of nurse practitioners. Research Memorandum, Burlington Study Group, McMaster University Medical Center, Hamilton, Ontario, Canada, 1973
34. Cappelleri JC: Cutoff-based designs in comparison and combination with randomized clinical trials. Unpublished doctoral dissertation, Ithaca, New York, Cornell Univ, 1991
35. Fineberg HV: The quest for causality in health services research. In Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Sechrest L, Perrin P, Bunker J, Eds. Washington, DC, US Dept HSS, Public Health Service, Agency for Health Care Policy and Research, May 1990
36. Mosteller F: Improving research methodology: An overview. In Research Meth-

odology: Strengthening Causal Interpretations of Nonexperimental Data. Sechrest L, Perrin P, Bunker J, Eds. Washington, DC, US Dept HSS, Public Health Service, Agency for Health Care Policy and Research, May 1990

37. Luft HS: The applicability of the regression discontinuity design in health services research. In Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Sechrest L, Perrin P, Bunker J, Eds. Washington, DC, US Dept HSS, Public Health Service, Agency for Health Care Policy and Research, May 1990
38. Williams SV: Regression discontinuity design in health research. In Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Sechrest L, Perrin P, Bunker J, Eds. Washington, DC, US Dept HSS, Public Health Service, Agency for Health Care Policy and Research, May 1990
39. Robbins H, Zhang CH: Estimating the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug. Proc Natl Acad Sci USA 86:3003–3005, 1989
40. Havassey B: Efficacy of cocaine treatments: A collaborative study. NIDA Grant Number DA05582, UCSF, San Francisco, 1988
41. Mohr LB: Impact Analysis for Program Evaluation. Chicago, Dorsey Press, 1988
42. Judd CM, Kenny DA: Estimating the Effects of Social Interventions. Cambridge, MA, Cambridge University Press, 1981
43. Stanley TD, Robinson A: Sifting statistical significance from the artifact of regression discontinuity design. Eval Rev 14:166–181, 1990
44. Stanley TD, Robinson A: "Second best" evaluation design: Regression discontinuity or abbreviated time series? Paper presented at the Annual Conference of the American Evaluation Assoc, Washington, DC, October 1990
45. Cappelleri JC, Trochim WMK, Stanley TD, Reichardt CS: Random measurement error doesn't bias the treatment effect estimate in the regression-discontinuity design. I. The case of no interaction. Eval Rev 15(4):395–419, 1991
46. Trochim WMK, Cappelleri JC, Reichardt CS: Random measurement error doesn't bias the treatment effect estimate in the regression-discontinuity design. II. When an interaction effect is present. Eval Rev 15(5):571–604, 1991
47. Goldberger AS: Selection bias in evaluating treatment effects: Some formal illustrations. Discussion paper #123. Madison, WI, Institute for Research on Poverty, 1972
48. Goldberger AS: Selection bias in evaluating treatment effects: The case of interaction. Discussion paper #129. Madison, WI, Institute for Research on Poverty, 1972
49. Kolata G: A critique of pure reason, a passion to survive. *New York Times*, p. E4, October 21, 1990
50. Klerman GL, Coleman JH, Purpura MD: The design and conduct of the Upjohn cross-national collaborative panic study. Psychopharmacology Bull 22(1):59–64, 1986
51. Klerman GL: Overview of the cross-national collaborative panic study. Arch Gen Psychiat 45:407–412, 1988
52. Ballenger JC, Burrows GD, DuPont RL, Lesser IM, Noyes R, Pecknold JC, Rifkin A, Swinson RP: Alprazolam in panic disorder and agoraphobia: Results from a multi-center trial. I. Efficacy in short-term treatment. Arch Gen Psychiat 45:413–422, 1988
53. Noyes R, DuPont RL, Pecknold JC, Rifkin A, Rubin RT, Swinson RP, Ballenger JC, Burrows GD: Alprazolam in panic disorder and agoraphobia: Results from a multi-center trial. II. Patient acceptance, side effects and safety. Arch Gen Psychiat 45:423–428, 1988
54. Pecknold JC, Swinson RP, Kuch K, Lewis CP: Alprazolam in panic disorder and

agoraphobia: Results from a multi-center trial. III. Discontinuation effects. Arch Gen Psychiat 45:429–436, 1988

55. Sheehan DV, Ballenger JC, Jacobsen G: Treatment of endogenous anxiety with phobic, hysterical and hypochondriachal symptoms. Arch Gen Psychiat 37:51–59, 1980

Controlled Clinical Trials

DESIGN, METHODS, AND ANALYSIS

Official Journal of the Society for Clinical Trials

Volume 13, Number 3, 1992

Contents

Controlled Clinical Trials is cited in Behavioral Medicine Abstracts, Biological Abstracts, Current Contents (Life Sciences), Excerpta Medica, Index Medicus, and Science Citation Index