Encyclopedia of Biopharmaceutical Statistics

edited by Shein-Chung Chow

StatPlus, Inc.
Yardley, and
Temple University
Philadelphia, Pennsylvania

MARCEL DEKKER, INC. NEW YORK • BASEL

Cutoff Designs*

I. INTRODUCTION

The randomized design is the preferred method for assessing the efficacy of treatments. Randomization of all subjects should be employed whenever possible. Randomization, in principle, serves at least three important purposes: (1) it avoids known and unknown biases on average; (2) it helps convince others that the trial was conducted properly; and (3) it is the basis for the statistical theory that underlies hypothesis tests and confidence intervals (1).

Randomization of all subjects has been criticized, however, because it may raise ethical concerns or practical limitations in certain situations. Ethical tensions may arise, for example, when strong a priori (though inconclusive) information favors the experimental treatment, when the disease is potentially life-threatening, and when randomization does not explicitly incorporate subjects' baseline clinical need or their willingness to incur risk (2,3). Examples that have stirred considerable debate about the ethics of the randomized design include the controversies about the release of drugs for AIDS (4), the availability of drugs for cancer treatment (5), and the use of extracorporeal membrane oxygenation (ECMO) for neonatal intensive care (6,7).

A second potential drawback of the randomized design occurs in instances when randomization is not feasible or practical. Such situations may arise in health services or outcomes research, where, for example, a health education program is to be targeted only to people who need it (8). An evaluation and comparison between managed care and usual care could be made feasible if high users of health care utilization receive managed care only and low users of health care utilization receive usual care only. A study concerned with the effect of a letter as an intervention to control health care costs could be made practical if the letter is sent only to physicians with high billed charges per subscriber, while those with lower billed charges per subscriber don't receive a letter (9). In these contexts, economic constraints and logistical barriers may dictate that an experimental intervention is neither practical nor efficient for those who don't need it or who are not the targeted candidates. Moreover, treatment allocation that reflects actual practice allows for testing the effectiveness of the intervention—its benefit in a real-life setting, as opposed to its efficacy in a controlled setting.

This entry discusses alternative design strategies that are intended to address ethical or practical concerns when it is deemed unethical or infeasible to randomize all subjects to study interventions. These design strategies are called *cutoff* designs, because they involve, at least in part, the assignment of subjects to treatments based on a cutoff score on a quantitative baseline variable that measures clinical need, severity of illness, or some other relevant measure. What follows is an overview of cutoff designs.

II. DESCRIPTION OF THE REGRESSION-DISCONTINUITY DESIGN

The most basic of cutoff designs is the regression-discontinuity design (8,10–13), in which a baseline indicator, for example, severity of illness, can be used to assign subjects to an intervention. All subjects below a cutoff point on the baseline indicator receive one treatment, while all subjects above it receive another treatment. The history of the regression-discontinuity (RD) design is found in the social sciences, specifically in

*This entry is drawn largely from JC Cappelleri. Embedding the regression-discontinuity design within the randomized design. Proceedings of the Biopharmaceutical Session of the American Statistical Association, 1997. ©Joseph C. Cappelleri.

program evaluation. It has been employed to evaluate the effects of compensatory education, being on the dean's list, a criminal justice program, a health education program on serum cholesterol, accelerated math training, and the NIH Career Development Award (14). In these scenarios randomization of subjects was not a viable alternative.

The traditional RD design is a single-cutoff quasi-experimental design that involves no random assignment. The RD design received its name from the "jump," or discontinuity, at the cutoff in the regression line of baseline and outcome (follow-up) scores that occurs when there is a treatment effect. Figure 1 depicts an RD design with a hypothetical 10-point treatment effect (reduction). All subjects with scores above 20 on the baseline assignment indicator are most in need of the intervention and hence are automatically assigned to the test (experimental) treatment, while those with scores of 20 or less (those less in need) are automatically assigned to the control treatment.

As Fig. 1 shows, the outcome scores of the test treatment group (those scoring above the cutoff) are lowered by an average of 10 points from where they would be expected in the absence of a treatment effect. The solid lines show the predicted regression lines for a 10-point effect, and the dashed lines show the expected regression lines for patients in a treatment group if they were given the other intervention instead.

The baseline assignment covariate should be measured on at least an ordinal scale; it is more desirable, though, to have a continuous (ratio-level or interval-level) baseline assignment variable. Baseline and outcome 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 group), the direction of improvement can be positive or negative for either variable, the treatment groups could have more than two levels, and the response variable can be discrete or continuous.

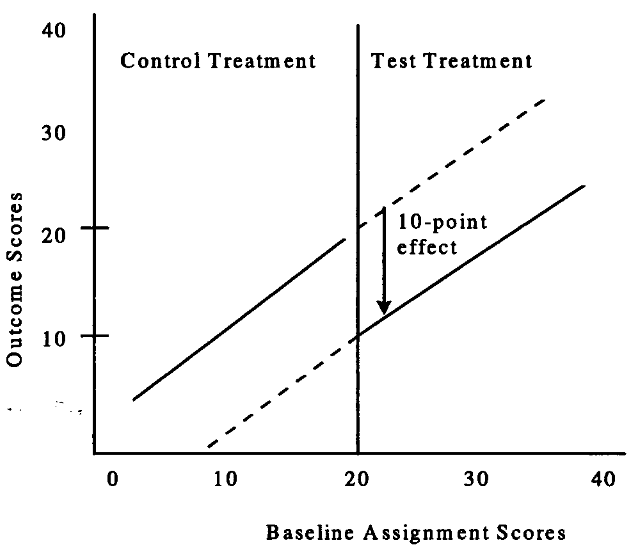

Fig. 1 Regression-discontinuity design with a 10-point treatment effect.

III. VALIDITY OF THE REGRESSION-DISCONTINUITY DESIGN

Under the assumption that the outcome-baseline functional form is correctly specified, the RD design results in an unbiased estimate of treatment effect. An unbiased estimate of treatment effect is obtained because the assignment process is known perfectly and controlled for in the analysis (10). Formal statistical derivations proving this lack of bias are found elsewhere (14–17). Like the randomized experimental (RE) design, the RD design gives a known probability of assignment to treatments. It is imperative, though, that the cutoff assignment rule be followed strictly. If subjects are misclassified, then the treatment effect is likely to be biased.

It can also be demonstrated that the estimate of treatment effect in the RD design, like the RE design, remains unbiased when random measurement error in the observed, fallibly measured baseline covariate is considered (14–17). The reason for this is that once the fallibly measured observed baseline scores are known, treatment assignment is completely determined and hence independent of anything else, including the perfectly measured true baseline scores, in the RD design. Similarly, in the RE design, treatment assignment is completely determined by a randomization scheme and hence independent of anything else.

Regression to the mean, which naturally emanates from random measurement error in the observed baseline covariate, does not therefore affect the estimate of treatment effect in both the RD design and the RE design. Figure 2 graphically shows the impact of regression to the mean, or, equivalently, random measurement error in the observed covariate, in the case of no treatment effect when the same variable is measured at baseline and follow-up. In the absence of a treatment effect, and with no other effects that might change a subject's score at follow-up, the true regression line should be a 45° line beginning at the origin. Regression to the mean causes the fitted regression line to be attentuated by an amount proportional to the reliability

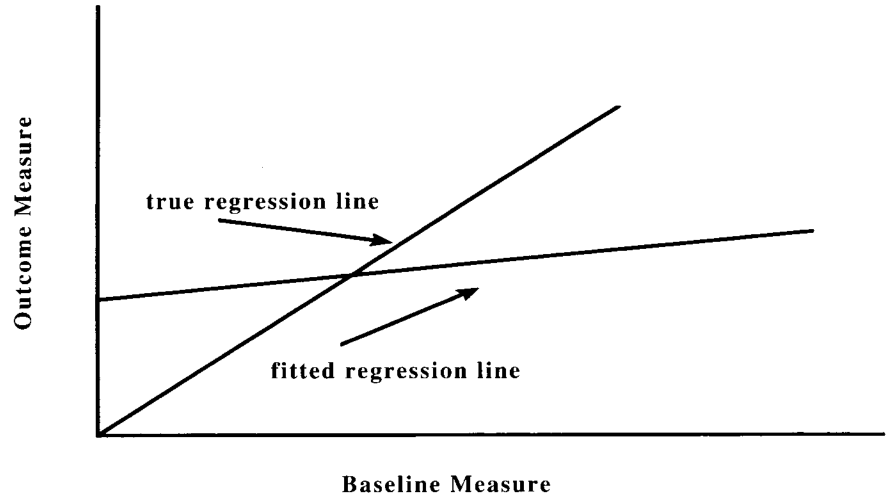

Fig. 2 Regression to the mean: randomized and regression-discontinuity designs.

coefficient of the baseline covariate; therefore, the sample regression coefficient of the baseline covariate on the outcome measure is biased, but the sample regression coefficient of the treatment effect is not (14,15).

The RE design is robust in giving unbiased estimates of treatment effect when the true functional form between the baseline covariate and the outcome measure is not correctly specified. On the other hand, the RD design is not robust here. The most critical step in obtaining an unbiased estimate of effect in the RD design lies in modeling this true functional form correctly. The true functional form, however, is not known in the RD design because of missing data. As shown in Fig. 1, which assumes a linear functional form, the extrapolated regression line of the control group (dashed line, right) if this group's subjects were given the test treatment instead is assumed to continue in the same linear way as its fitted line (solid line, left). The extrapolated regression line of the test-treated group (dashed line, left) if this group's subjects were given the control treatment is assumed to continue in the same way as its fitted line (solid line, right). There is no way to know prospectively whether the form or the slope of the lines in the region of missing data will be the same as that in the region of observed data.

IV. SPECIFYING THE FUNCTIONAL FORM

One suggestion for helping to arrive at the correct functional form is to use a polynomial backward-elimination regression approach (18). Another suggestion uses empirical Bayesian methods to overcome situations when the outcome and baseline relationship may not be linear, as when true baseline scores are not normally distributed (19,20). A third approach, which can be used with either of the other two approaches, is to fit a regression line over a wider range of the baseline-outcome distribution, resulting in less extrapolation and hence a more valid fit. This last approach can be achieved by combining the RD design with the RE design, resulting in a cutoff design with randomization.

V. COMBINING REGRESSION-DISCONTINUITY AND RANDOMIZED EXPERIMENTAL DESIGNS

A regression-discontinuity design can be described as a cutoff design without randomization. This design can also be coupled with a randomized design. For instance, patients who score within the middle range of scores on a baseline severity-of-illness indicator (e.g., those moderately ill) are randomized to either one of two treatments, while patients who score below a given cutoff value on this indicator (e.g., those most ill) are automatically assigned to the novel treatment and patients who score above another, higher cutoff value (e.g., those least ill) are assigned to the control treatment. Another type of cutoff design, for instance, would have subjects below the single cutoff point (e.g., the most ill) randomized to either treatment, while those above it (e.g., the least ill) are automatically assigned to control treatment. These are only two possible design variations that combine cutoff assignment and random assignment. Other variations are mentioned elsewhere (21,22).

Combining the RE design and the RD design may give advantages over either design alone (21–23). Relative to the RE design, this hybrid design may be better suited to address ethical or practical concerns, may result in a larger eligible and diverse sample, and may address better the effectiveness (as opposed to the efficacy) of interventions in particular circumstances. Compared with the RD design, RD-RE design has enhanced validity and improved statistical power.

VI. ILLUSTRATION: COCAINE PROJECT

To illustrate the combined design, we describe a cocaine project, conducted at the University of California at San Francisco, that applied the RD-RE design instead of the completely randomized design, which was considered neither ethical nor feasible (24). The study included about 500 patients with cocaine addiction. The objective of the study was to ascertain whether inpatient (intensive) rehabilitation showed better improvement, and by how much, over outpatient rehabilitation. The baseline assignment covariate was based on a weighted composite of four scales: (1) employment and legal status, (2) family relationship and recovery, (3) alcohol and drug history, and (4) psychological status. Higher scores indicated more clinical need for the more intensive (inpatient) rehabilitation. The primary outcome variable was the same variable measured at follow-up.

Figure 3 portrays how patients may be allocated into inpatient or outpatient rehabilitation in this setting. All patients who score above 60—those most severely ill or most in need—are automatically assigned to inpatient rehabilitation; all patients who score below 40—those least ill or least in need—are automatically assigned to outpatient rehabilitation; and patients who score between 40 and 60, inclusive—those moderately ill or in need—are randomized to either inpatient rehabilitation or outpatient rehabilitation. Note that it is this cutoff interval of randomization that distinguishes the RD-RE design from the RD design, which instead has a cutoff point(s) with no randomization.

Like Fig. 1, Fig. 3 has solid lines representing the predicted regression lines and dashed lines representing the extrapolated regression lines, showing a constant improvement from inpatient rehabilitation over outpatient rehabilitation. An analysis of covariance model, with the baseline assignment measure and the treatment group variable as predictors, would be a correct model

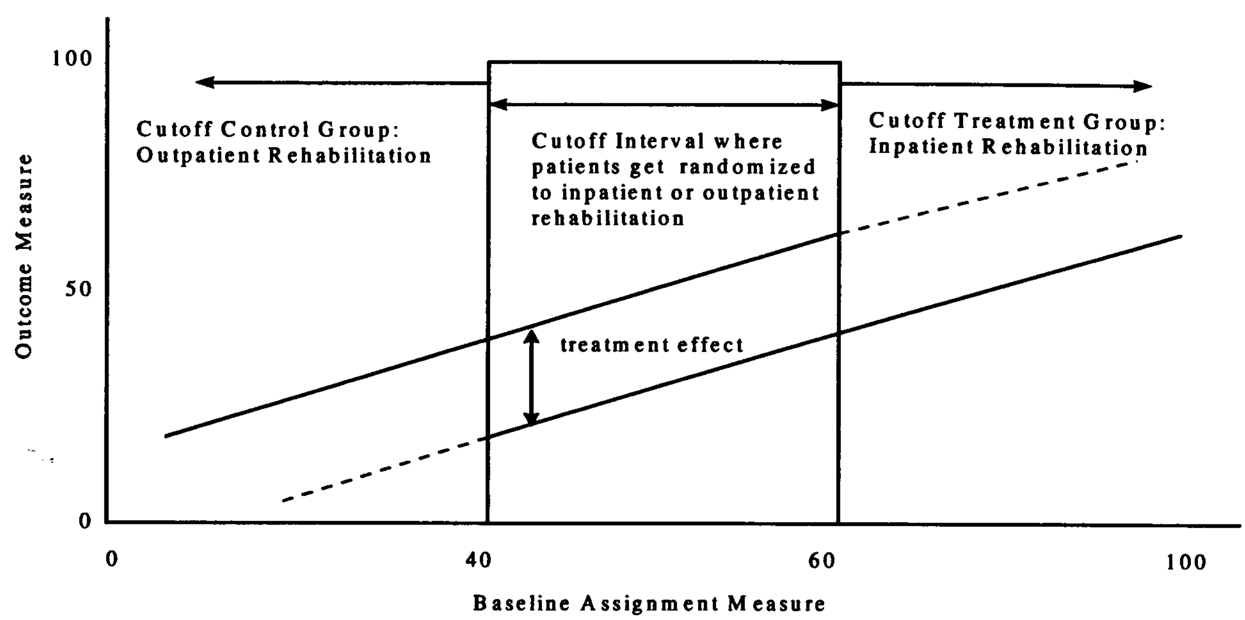

Fig. 3 Illustration of a combined randomized and regression-discontinuity design.

to fit the fitted lines in Figs. 1 and 3. An analysis of variance model, which excludes the baseline assignment variable, should be not fit, for it would result in a biased estimate of treatment effect. While linear relations are highlighted in these two figures, cutoff designs are not restricted to a linear baseline-outcome relationship; higher-order terms (e.g., quadratic or cubic terms), transformations on baseline or outcome variables, and interaction terms may also be fitted.

In a simulation study, several RD-RE design variations, of which the basic design in Fig. 3 is the simplest, were evaluated and compared among themselves, along with the traditional RD design and the traditional RE design (21,22). An unbiased main treatment effect was found for all these designs.

Figure 4 shows one of the more advanced RD-RE designs that may be useful for accommodating varying amounts of resources. One cutoff interval has its bounds at 45 and 55; the other cutoff interval has its bounds at 40 and 60. Both intervals are symmetric around 50. Because the two intervals have different widths, they include different numbers of randomized patients, with the wider interval containing more randomized subjects. As subjects accrue into a study, investigators of a clinical site may favor one interval of randomization over the other in order to address the cost implications of having a shortage or surplus of hospital beds for inpatient rehabilitation. Or one interval may be preferred because it is more commensurate with a hospital's level of resources and expertise with respect to a given treatment.

VII. MODELING AND ANALYZING CUTOFF DESIGNS

The RD-RE combination can be modeled and analyzed with the polynomial backward-elimination approach suggested in Sec. IV for the RD design. Specifically, the initial model equation is

$$\begin{aligned} y = {} & \text{bint} + (\text{btrt})*z + (\text{b}x\text{cut})*\text{xcut} \\ & + (\text{b}x\text{cut2})*(x\text{cut})^2 + (\text{b}x\text{cut3})*(x\text{cut})^3 \\ & + (\text{blinint})*(z*x\text{cut}) + (\text{blinquad})*(z*x\text{cut}^2) \\ & + (\text{blincub})*(z*x\text{cut}^3) + \text{error} \end{aligned}$$

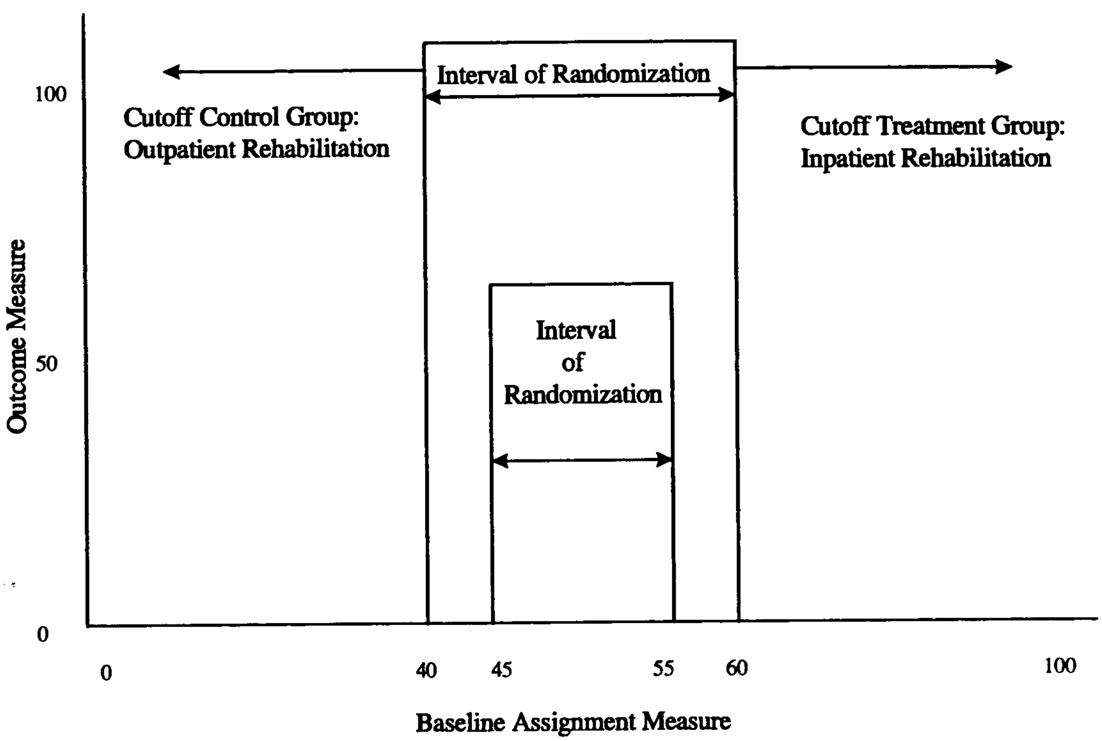

Fig. 4 Randomized and regression-discontinuity design with two cutoff intervals.

where

y = outcome measure
$x$cut = baseline assignment covariate minus a baseline value at which to measure the treatment effect (e.g., the middle value in a cutoff interval in an RD-RE design or the cutoff value itself in an RD design)
z = binary treatment group variable
bint = intercept estimator
btrt = treatment effect estimator
b$x$cut = linear slope estimator
blincut = linear interaction estimator
error = sample regression error term

The other regression coefficients are the coefficients for powers of "$x$cut" higher than 1 and for their corresponding higher-order interactions. The same set of assumptions that apply to linear regression (for continuous responses) and to logistic regression (for discrete responses) also apply here.

The modeling strategy first tests the significance of each regression coefficient separately, beginning with the higher-order interaction terms (i.e., the cubic interaction is tested first, followed by the quadratic interaction, and then linear interaction); interaction terms are tested before main effect terms. All significant terms and their lower-order counterparts are retained. The baseline covariate term and the treatment group variable are always kept in the final model.

VIII. RELATIVE SAMPLE SIZES NEEDED IN CUTOFF DESIGNS

The simulation study mentioned in Sec. VI also showed that, everything else the same, more randomization resulted in lower standard errors of the treatment effect estimate and therefore increased precision. It can be shown that the amount of this precision is completely determined by the multicollinearity or correlation (R) between the baseline assignment covariate and the treatment group variable as expressed by the variance inflation factor (VIF) (15,25):

$$\text{VIF} = \frac{1}{1 - R^2}$$

Suppose that there is a binary treatment group variable and a normally distributed baseline covariate. Table 1 provides the correlation between these two variables (R) and the accompanying variance inflation factor (VIF) in symmetric cutoff designs with varying amounts of randomization and with 50% of the subjects within the interval assigned randomly to either treatment. The VIF can be interpreted as the design effect of how many more subjects are needed in a given cutoff design relative to the completely randomized design in order to achieve the same level of statistical power, everything else the same.

Table 1 shows that, to achieve the same level of statistical power as the RE design, 2.75 times more subjects are needed in an RD design; 2.48 times more subjects are needed in an RD-RE design with 20% of all subjects randomized (i.e., 20% randomization); 1.96 times more subjects are needed in an RD-RE design with 40% randomization; 1.46 times more subjects are needed in an RD-RE design with 60% randomization; and 1.14 times more subjects are needed in an RD-RE design with 80% randomization. Derivations for the efficiency of such a cutoff design using an analogous approach, which gives the same results, are published elsewhere (26).

IX. RECENT CRITIQUES

Cutoff designs are certainly not without limitations. As mentioned earlier, an unbiased estimate of treatment effect requires that the functional relationship between outcome and baseline covariate be correctly modeled. Finklestein et al. (19,20) proposed a mathematical and statistical foundation, illustrated with examples, for how to analyze the RD design and to draw valid statistical conclusions about treatment efficacy. The authors discuss and illustrate their empirical Bayes methodology, which, they mention, can be used in a variety of circumstances, as a way to overcome restrictive assumptions about the functional form between outcome and baseline covariate.

Another reservation with cutoff designs is that they preclude any serious attempt at complete blinding of treatment, making them similar to nonrandomized designs in this regard. A further drawback of cutoff designs is that they are less efficient (precise) than completely randomized designs in terms of their estimates of treatment effects. According to Senn (27), the considerable excess of patients treated with the inferior treatment in cutoff designs (especially the RD design) relative to the RE design is likely to undermine the ethical argument that favors cutoff designs. While it is also true that more patients will receive the superior treatment in cutoff designs, regardless of which treat-

Table 1 Correlations and Variance Inflation Factors for Designs with Varying Amounts of Randomization

Percentage of all subjects within interval of randomization	Correlation coefficient[a]	Variance inflation factor
0 (regression-discontinuity design)	0.79	2.75
20	0.77	2.48
40	0.70	1.96
60	0.56	1.46
80	0.35	1.14
100 (randomized design)	0.00	1.00

[a]Expected correlation between a binary treatment variable and normally distributed baseline covariate.

ment it is, researchers are urged to consider Senn's position (27) before abandoning randomization as a perceived ethical problem in a clinical trial.

X. CONCLUSIONS

Randomization should be employed whenever possible. Cutoff designs should not replace the completely randomized design in the majority of circumstances, usually involving a drug intervention, when no appreciable logistical barriers preclude all subjects from being randomized to interventions. Cutoff designs are an alternative design when circumstances in health services research or outcomes research warrant that randomization of all subjects cannot be undertaken, for whatever reason. Cutoff designs are much more likely to be relevant and appropriate in studies on program evaluation that involve educational or behavioral interventions than in traditional phase 3 studies on drug interventions, but cutoff designs may have potential in phase 2 therapeutic trials as well. When compared with nonrandomized designs, the regression-discontinuity design (a cutoff design with no randomization) is an attractive alternative. When some subjects can be randomized, coupling the regression-discontinuity design with the randomized design is even a more attractive alternative than the regression-discontinuity design.

REFERENCES

1. S Green. Controlled Clin Trials 3:189–198, 1982.
2. KF Schaffner. J Med Philos 11:297–404, 1986.
3. LS Parker, RM Arnold, A Meisel, LA Siminoff, LH Roth. Clin Res 3:537–544, 1990.
4. E Marshall. Science 245:346–347, 1989.
5. JL Marx. Science 245:345–346, 1989.
6. JH Ware. Stat Sci 4:298–340 (with discussion), 1989.
7. RD Truog. Clin Res 38:537–544, 1992.
8. WMK Trochim. In: L Sechrest, P Perrin, J Bunker, eds. Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Rockville, MD: Agency for Health Care Policy and Research, U.S. Public Health Service, 1990, pp 119–130.
9. SV Williams. In: L Sechrest, P Perrin, J Bunker, eds. Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data. Rockville, MD: Agency for Health Care Policy and Research, U.S. Public Health Service, 1990, pp 145–149.
10. TD Cook, DT Campbell. Quasi-experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton-Mifflin, 1979, pp 137–147, 202–205.
11. WMK Trochim. Research Design for Program Evaluation: The Regression-Discontinuity Approach. Beverly Hills, CA: Sage Publications, 1984.
12. SL Coyle, RF Boruch, CF Turner, eds. Evaluating AIDS Prevention Programs. Expanded edition. Washington, DC: National Academy Press, 1991, pp 144–159.
13. LB Mohr. Impact Analysis for Program Evaluation. Newbury Park, CA: Sage Publications, 1995, pp 133–155.
14. JC Cappelleri, WMK Trochim, TD Stanley, CS Reichardt. Eval Rev 18:141–152, 1991.
15. AS Goldberger. Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations. Madison, WI: Institute for Research on Poverty, 1972, Discussion paper #123.
16. DB Rubin. J Educ Stat 2:1–26, 1977.
17. CS Reichardt, WMK Trochim, JC Cappelleri. Eval Rev 19:39–63, 1995.
18. JC Cappelleri, WMK Trochim. J Clin Epidemiol 47: 261–270, 1994.
19. MO Finkelstein, B Levin, H Robbins. Am J Public Health 86:691–695, 1996.
20. MO Finkelstein, B Levin, H Robbins. Am J Public Health 86:696–705, 1996.

21. WMK Trochim, JC Cappelleri. Controlled Clin Trials 13:190–212, 1992.
22. JC Cappelleri, WMK Trochim. Med Decis Making 15: 387–394, 1995.
23. RF Boruch. Socio Meth Res 4:31–53, 1975.
24. BE Havassy, BR Wesson, JM Tshann, SM Hall, CJ Henke. Efficacy of cocaine treatment: A Collaborative Study. Grant proposal funded by the National Institute on Drug Abuse (NIDA #DA05582), 1989.
25. JC Cappelleri. Cutoff-based designs in comparison and combination with randomized clinical trials. PhD dissertation, Cornell University, Ithaca, NY, 1991.
26. SJ Senn. Statistical Issues in Drug Development. Chichester, UK: Wiley, 1997, pp 89–92.
27. SJ Senn. Statist Med 15:114–116, 1996.

Joseph C. Cappelleri
William M. K. Trochim