Methodological Issues

Application of Regression-Discontinuity Analysis in Pharmaceutical Health Services Research

Ilene H. Zuckerman, Euni Lee, Anthony K. Wutoh, Zhenyi Xue, and Bruce Stuart

Objective. To demonstrate how a relatively underused design, regression-discontinuity (RD), can provide robust estimates of intervention effects when stronger designs are impossible to implement.

Data Sources/Study Setting. Administrative claims from a Mid-Atlantic state Medicaid program were used to evaluate the effectiveness of an educational drug utilization review intervention.

Study Design. Quasi-experimental design.

Data Collection/Extraction Methods. A drug utilization review study was conducted to evaluate a letter intervention to physicians treating Medicaid children with potentially excessive use of short-acting β_2 -agonist inhalers (SAB). The outcome measure is change in seasonally-adjusted SAB use 5 months pre- and postintervention. To determine if the intervention reduced monthly SAB utilization, results from an RD analysis are compared to findings from a pretest-posttest design using repeated-measure ANOVA.

Principal Findings. Both analyses indicated that the intervention significantly reduced SAB use among the high users. Average monthly SAB use declined by 0.9 canisters per month ($p < .001$) according to the repeated-measure ANOVA and by 0.2 canisters per month ($p < .001$) from RD analysis.

Conclusions. Regression-discontinuity design is a useful quasi-experimental methodology that has significant advantages in internal validity compared to other pre-post designs when assessing interventions in which subjects' assignment is based on cutoff scores for a critical variable.

Key Words. Regression discontinuity, drug utilization review, quasi-experimental design

Health services research often involves evaluation of medical interventions or service programs outside a controlled environment. Such interventions or programs may be the result of public policy or may result from diffusion of

newly introduced services or standards. The health services researcher is challenged to evaluate the effectiveness of such programs. Even though the randomized-controlled trial may be the ideal study design, it is frequently infeasible because of ethical, financial, logistical, or political constraints. Researchers have devised many alternative designs to assess the impact of interventions when randomized trials are not possible. These include pretest–posttest comparisons (both with and without control groups), interrupted time-series, and multiperiod panel models (Shadish et al. 1996; Wooldridge 2003). The pretest–posttest design is the easiest to implement and is a common method for evaluating drug utilization review (DUR) interventions (Gurwitz, Noonan, and Soumerai 1992; Ahluwalia et al. 1996; Farris et al. 1996; Collins et al. 1997; Monane et al. 1998). In this paper, we compare results from a pretest–posttest analysis of a DUR intervention to findings from a relatively underused technique known as the regression-discontinuity (RD) design. We show that the latter is robust to the major threats to internal validity associated with simple pretest–posttest designs. We then discuss the applicability of the RD methodology to other aspects of pharmaceutical health services research.

BACKGROUND

The defining characteristic of the pretest–posttest design is that members of the study group are compared with themselves instead of to a control or nonequivalent comparison group. Observations on the variable of interest are collected from the study group prior to the intervention (X_{pre}), and after the intervention (X_{post}). The difference, $X_{\text{pre}} - X_{\text{post}}$, is interpreted as the change resulting from the intervention. This is a reasonable way to achieve the goal of an experiment in the sense that all possible time-invariant factors associated with the study subjects are controlled. However, this design does not control for time-varying factors that may be coincident with the study time frame (Kerlinger and Lee 2000).

Address correspondence to Euni Lee, Pharm.D., Ph.D., Center for Minority Health Services Research, Department of Clinical and Administrative Pharmacy Sciences, School of Pharmacy, Howard University, 2300 Fourth Street, NW, Washington, DC, 20059-0001. Ilene H. Zuckerman, Pharm.D., Ph.D., and Bruce Stuart, Ph.D., are with the Department of Pharmaceutical Health Services Research, School of Pharmacy, University of Maryland, Baltimore, MD. Anthony K. Wutoh, Ph.D., and Zhenyi Xue, M.S., are with the Center for Minority Health Services Research, Department of Clinical and Administrative Pharmacy Sciences, School of Pharmacy, Howard University, Washington, DC.

Between the periods of pretest and posttest, many things may occur in addition to the studied intervention. The longer the time interval, the greater the chance that extraneous factors will affect the variable of interest. Regression to the mean is one such factor. It refers to the phenomenon where lower scores on the pretest tend to be higher and higher scores lower on the posttest (i.e., regress to the mean) irrespective of the impact of the intervention. In circumstances where study subjects are selected for an intervention precisely because they fall at the lower or upper end of a distribution in the pretest (as in the case of DUR interventions) then regression presents a significant potential threat to validity. The traditional way to control for regression effects in observational designs is to have a comparison group matched on values of the variable of interest during the pretest. However, in operational DUR programs, *all* subjects exceeding the threshold value of the critical variable are typically assigned to the intervention, thereby negating the possibility of a contemporaneously matched comparison group. Lack of matched comparison groups also makes it difficult to control for other time-varying threats to internal validity including the effects of history and maturation.

The RD design has its roots in educational literature from the 1960s (Thistlewaite and Campbell 1960). In the 1970s, RD was used to evaluate compensatory education programs as well as criminal justice and social welfare programs (Trochim 1984). Carter, Winkler, and Biddle (1987) adopted the RD design to evaluate the effectiveness of NIH research career development awards. Not until the 1990s is there any significant discussion of using RD in health services research. The Agency for Healthcare Policy and Research (AHCPR, now known as the Agency for Healthcare Research and Quality [AHRQ]) sponsored a conference and published proceedings on research methodology and nonexperimental data in 1990; one paper involved RD design (Trochim 1990). Subsequently, Trochim and Cappelleri (1992) published simulation models of randomized clinical trials versus RD design and cutoff-based randomized clinical trials. Finkelstein, Levin, and Robbins (1996a, b) also modeled risk-based allocation using principles of RD design.

RD design is characterized by its method of assigning subjects. Briefly, a cutoff score on an assignment measure, rather than random assignment, is employed. All subjects who score on one side of the cutoff are assigned to the intervention group while those scoring on the other side are assigned to a control group. This method of subject assignment uniquely aligns the method to nonexperimental interventions that employ threshold selection procedures (like our DUR intervention). Figure 1 provides an example where subjects with preintervention scores higher than the cutoff value were assigned to the

Figure 1: Hypothetical Regression Discontinuity Design Showing an Intervention Effect (Decreased Posttest Score after the Intervention)*

*Adapted from Trochim (1990).

intervention group; those with scores below the cutoff value were assigned to the control group and received no treatment. Based on the control group's regression equation, one could predict what the intervention group's values would have been if the program had no effect. In this simple illustration, the difference or "discontinuity" in the two regression lines *at the cutoff* provides an estimate of the intervention effect.

The RD design has the same strength as the pretest–posttest design in controlling for time-invariant individual characteristics. However, the robustness of the RD design to time-varying threats to internal validity represents the method's primary attraction. For example, in the RD design we expect that all participants will mature (e.g., naturally occurring improvement in the subject's asthma condition) and that, on average, maturation may be different for the two groups. However, the RD design is not measuring the intervention effect as the difference in the posttest averages of the two groups, but rather by a change in the pre–post relationship at the cutoff point. Thus for maturation to introduce bias into the RD design, the difference in maturation rates would need to cause a discontinuity in the pre–post relationship that coincides with the cutoff point—a very unlikely event. The same reasoning applies to the threat of bias because of history, regression to the mean, or external environmental influences.

We believe that the RD design is flexible with respect to situations that are more complex than the simple linear regression example shown in Figure 1. The addition of quadratic and higher order polynomials can accommodate curvilinear relationships between pretest and posttest values. Adding

Figure 2: Hypothetical Regression Discontinuity Design Showing an Intervention Effect with an Interaction Term*

*Adapted from Trochim (1990).

interaction terms between the treatment indicator and pretest values (and polynomials of the same) tests whether there are significant differences in the slopes of the regression lines for the intervention and control groups (Figure 2). In the typical RD application these higher-order terms and interactions are included in the initial test, and then discarded if the coefficients are insignificant (Trochim 1990). The result is a parsimonious model.

It is also possible in the RD framework to test the sensitivity of discontinuities at the cutoff point by estimating models in which the cutoff point is artificially lowered. In the event of a true intervention effect, the measured discontinuity at the artificial cutoff should be lower because the “nonintervention” posttests are contaminated with intervention effects. If the estimated discontinuities are insensitive to changes in the cutoff point, then the analyst has reason to suspect a spurious relationship between the intervention and the true outcome. We present an example of this type of sensitivity test in the DUR application described in the next section. More detailed discussion of RD design can be found elsewhere (Trochim 1984; Shadish et al. 1996).

METHODS

To illustrate and compare the results of an RD analysis with a pretest–posttest design, we analyzed data from a Mid-Atlantic state Medicaid drug utilization intervention intended to improve the pharmacologic management of

pediatric asthma. A description of the intervention is reported elsewhere (Lee, McNally, and Zuckerman 2004). Briefly, we identified all Medicaid children with asthma who exceeded the national guideline of no more than an average of one prescription per month for short-acting β_2 -agonist inhalers (SAB) (National Heart, Lung and Blood Institute 1997). Physicians treating these children were subsequently identified and sent educational letters explaining the problem and giving the children's names. We hypothesized that the letter campaign would lead to lowered rates of SAB use among the identified children after the intervention.

For our simple pretest–posttest analysis, we used repeated measure ANOVA to compare the differences in monthly SAB prescription fills for the 333 “high-user” children who were continuously enrolled in Medicaid for a seasonally equivalent period 5 months preintervention and 5 months post-intervention.

To conduct the RD analysis, we compared the pre–post experience of the intervention group of 333 to that of 3,306 continuously enrolled children with asthma who had at least one prescription for a SAB and were below the cutoff value of an average of one SAB canister per month during the seasonally matched preintervention period. We used ordinary least squares (OLS) regression for all versions of the RD model. Our initial model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_3 X_i T_i + \beta_4 X_i^2 + \beta_5 X_i^2 T_i + \varepsilon_i$$

where Y_i is the posttest measure for individual i , X_i is the pretest measure for individual i minus the cutoff value, T_i is the treatment indicator (1 for treatment, 0 for control), β_0 is the intercept, which can be interpreted as the post-intervention estimate for the control group at the cutoff, β_1 is the linear slope parameter, β_2 is the intervention effect estimate, β_3 is the linear interaction term, β_4 is the quadratic pretest coefficient, β_5 is the quadratic interaction term, and ε_i is the random error term.

We examined pre–post relationship using a plot and looked for (1) discernable discontinuity at the cutoff and (2) flexion points. As we did not observe any flexion points from the visual examination, i.e., the bivariate relationship appeared linear, we created a second-order polynomial because the rule of thumb was to go two orders of polynomial higher than was indicated (Trochim 1990). We believe our initial model is an over-specified model including all the necessary covariates. To avoid collinearity, we centered the pretest measurement by subtracting the cutoff value from the pre-intervention measure. This also allows estimation of the intervention effect at the cutoff rather than at 0. This model has low power if a simplified model

happens to be the true model in the reality. In order to improve the efficiency of our statistical test, subsequent revisions of this model were made based on results from the initial model. We sequentially removed insignificant terms until all remaining coefficients were significant. To test whether our findings may have been caused by some other event besides the intervention, we conducted a sensitivity analysis by simulating a cutoff point of 0.5 canisters per month. As explained above, by misclassifying those who did not have the intervention as having the intervention, we expect to see an attenuation of the (true) intervention effect.

RESULTS

The repeated-measure ANOVA analysis from the pretest–posttest design showed a significant decline in average monthly SAB use of 0.9 SAB inhaler canisters per month from 1.6 in the preperiod to 0.7 in the postperiod ($p < .001$). The RD results are summarized in Table 1. A plot of the pre–post data points with the fitted regression lines is shown in Figure 3. The initial OLS model (upper panel of Table 1) showed insignificant quadratic and interaction terms. In the final model, we removed the insignificant terms (lower panel of Table 1) and the intervention effect remained significant (-0.21 canister per month; $p < .001$). Results of the sensitivity analysis shown in Table 2 supported

Table 1: Regression Estimates for the Initial and Final Regression-Discontinuity Models

<i>Variable</i>	<i>Regression Coefficient</i>	<i>Standard Error</i>	<i>p</i>
Initial model			
Constant	0.511	0.051	<.001
Linear (X_i)	0.351	0.199	.078
Intervention effect (T_i)	− 0.186	0.065	.004
Linear interaction (X_iT_i)	0.269	0.219	.220
Quadratic (X_i^2)	− 0.158	0.179	.378
Quadratic interaction ($X_i^2T_i$)	0.138	0.183	.449
	$R^2 = 0.228$		
Final model			
Constant	0.568	0.018	<.001
Linear (X_i)	0.549	0.025	<.001
Intervention effect (T_i)	-0.213	0.038	<.001
	$R^2 = 0.228$		

X_i is the pretest measure minus the cutoff for individual i ; T_i is the treatment indicator (1 for treatment; 0 for control).

Figure 3: Bivariate Distribution of Short-Acting β_2 -Agonist Inhaler Use in the Pre- and Postintervention Period with Regression Lines From the Final Regression Discontinuity

our main study finding. With a simulated cutoff of 0.5, the model was linear with no interaction effect, and the artificial intervention effect was not significant.

DISCUSSION

This paper has presented two approaches to measuring the impact of a DUR intervention designed to reduce overuse of SA β_2 -agonist inhalers among

Table 2: Regression Estimates for the Regression-Discontinuity Model with Simulated Intervention Cutoff of 0.5 Short-Acting β_2 -Agonist Canister per Month

Variable	Regression Coefficient	Standard Error	<i>p</i>
Constant	0.264	0.008	<.001
Linear (X_i)	0.424	0.021	<.001
Intervention effect (T_i)	0.016	0.022	.469
$R^2 = 0.221$			

X_i is the pretest measure minus the cutoff for individual i ; T_i is the treatment indicator (1 for treatment; 0 for control).

children with asthma enrolled in a Mid-Atlantic state Medicaid program. The first test used repeated-measure ANOVA and showed a significant reduction of almost a full canister between the pre- and postintervention periods. However, because the one-group pretest–posttest design suffers from several potential threats to validity arising from both internal (maturation, regression to the mean) and external (history) sources, we could not conclude that the intervention alone was responsible for the observed decline in SAB use. A randomized experiment is often considered the preferred method for obtaining unbiased estimates of an intervention effect because randomization reduces the threat of selection bias. In our case randomization was not feasible, so our discussion is focused on alternative quasi-experimental designs, which have been extensively reviewed elsewhere (Shadish, Cook, and Campbell 2002; Trochim 2003). We believe the RD approach is ideally suited for circumstances such as the asthma example, for the following reasons.

When deciding on an analytic approach, the statistical analysis should match the design. Adding a control group that receives no intervention is an improvement to the one-group pretest–posttest design. Ideally the control group should be similar to the intervention group. However, when circumstances do not allow employing such a control group, a nonequivalent control group design is a frequently used quasiexperimental design. Although the design is susceptible to internal validity threats, it is possible to address some of these threats through variations in the design and analysis. For example, adding additional pretest measurements may allow better assessment of regression and maturation. Use of propensity scores (i.e., the predicted probability of being in the intervention versus the control group) or selection bias modeling using econometric techniques can be implemented to reduce selection bias that may be present when using nonequivalent control groups. Shadish, Cook, and Campbell (2002) point out that selection bias models are closely related to the RD design because both achieve an unbiased estimate of the intervention effects through complete knowledge of selection. In our case, the fact that the intervention enrolled all eligible children meeting the inclusion criterion (greater than one SAB canister per month on average) meant that we could not construct an independent, contemporaneously similar control group against which to compare the experience of those subjected to the DUR intervention. We rejected a historical control group approach because of vagaries in the seasonal severity of asthma over time.

The interrupted time-series design is another strong quasi-experimental design alternative that is available to health services researchers interested in evaluating DUR programs and other pharmaceutical interventions. Wagner et al. (2002) argue that the interrupted time-series design is the strongest quasi-experimental approach to evaluating the longitudinal impact of interventions. Intervention impacts are inherently dynamic. Study subjects learn to adapt to changes in their environment and frequently respond in unexpected ways. Moreover, external influences invariably arise that impinge on the “purity” of the measured response. These dynamic elements cannot be adequately captured with two-period, before-and-after designs, no matter how sophisticated they otherwise may be. But time-series analysis per se is generally not an appropriate choice when individual-level data are available to the analyst. The reason is that time-series is an aggregate-level approach with one observation per time period—or in the case of comparative time-series designs, two observations per time period. However, one limitation in this design is that the researchers collapse individual-level datasets to evaluate mean sample-level differences over time. Panel analysis is the appropriate tool if the analyst has multiperiod individual-level data and wishes to evaluate intervention impacts longitudinally. Panel studies maintain individual variance, and if there are enough periods available for analysis, time-specific shocks can be evaluated much the same way as in traditional time-series analysis.

So, given these design alternatives and the limitations in our data, our situation seemed ideal for the regression-discontinuity design, in which the pre–post experience of children exposed to the intervention was compared contemporaneously to that of children who were below the threshold SAB use and thus not exposed to the intervention. In our case, the subjects were assigned to the intervention based on a cutoff score and only on this score, so the selection process was completely known, an ideal situation for the regression discontinuity design approach. The results of the RD analysis including a sensitivity check confirmed that the intervention had the desired effect of reducing SAB overuse, but the measured impact was much smaller than determined through the pretest–posttest analysis (-0.2 versus -0.9 canisters per month). The lower effect size was expected because the RD design is robust to regression to the mean (a likely source of inflated impact in the pretest–posttest design given that only children with very high SAB use were enrolled in the intervention).

The key to all health service research is having adequate statistical control. RCTs achieve it through random assignment of subjects to intervention and comparison groups. Simple pretest–posttest designs achieve control

under the very restrictive assumption that the intervention impact is the only time-sensitive factor that can influence subject behavior. Multiperiod panel designs may or may not achieve control depending on the distribution of characteristics of those subjected to the intervention and those selected as controls. Traditional interrupted time-series designs achieve control through autoregressive integrated moving average (ARIMA) models that essentially eliminate all time-related information other than that associated with the intervention itself. These models typically require at least 25 observations before and an equal number after the intervention in order to produce reliable results (Johnston, Ottenbacher, and Reichardt 1995).

So where does RD fit within this panoply of research designs? First, it must be remembered that the RD design was developed to address the special case of subject assignment based on a pretest score; those above the score are assigned and those below it are not (or vice versa). The essential requirement of this assignment regimen is that there can be no overlap on the critical test variable between the intervention and nonintervention groups. Traditional regression models (including panel models) will not produce reliable estimates of the intervention impact under these circumstances. RD analysis will produce reliable estimates, but *only* where there is no overlap in assignment. This may seem like a very restrictive use for the technique, but there are potentially large numbers of therapeutic applications (e.g., where physicians prescribe medicines based on threshold values of lab tests). By design, assignment is completely known and perfectly measured, as in a randomized-controlled trial. In the RD design, it is possible that recording errors may occur or social pressures may override assignment (as in the RCT), which introduces selection bias. In RD, it is possible that the cutoff score itself has measurement error. In our study perhaps the measure of SAB overuse is subject to error because the claims do not necessarily measure actual medication use. However, in RD design, the cutoff score is used to measure how subjects were assigned to a group and it contains no error for this purpose.

A second factor to keep in mind is that when all is said and done RD is simply a before and after test and cannot address complex time sensitivities. This means that RD applications will generally need to be followed up using other evaluative techniques to determine whether interventions have lasting impacts or not. Lastly, one should be aware of the assumptions of an RD analysis. Briefly, an exclusive cutoff criterion should be employed; pre-post distribution can be described as a polynomial function rather than logarithmic, exponential, or some other function; a sufficient number of pretest values should be present; all subjects must come from a single continuous pretest

distribution; and the intervention program was uniformly delivered to all recipients (Trochim 1990).

Perhaps the biggest obstacle to using the RD design is that it has not been widely employed in health services research. A literature search with key words “regression discontinuity” and “cutoff designs” and their variations covering Medline (1980 to July 2005), International Pharmaceutical Abstracts (1970 to July 2005), and EMBASE (1980 to July 2005) revealed only a few empirical studies using RD design (Daniels et al. 1992; Hoglend 1996; Cullen et al. 1999). Although a few applications of the RD approach have been made in economics (Pitt and Khandker 1998; Anguist and Lavy 1999), program evaluation in health services research has lacked a formal application of this method. One possibility for the lack of articles using the RD design is publication. As Shadish, Cook and Campbell point out, there was a 30-year lag in widespread acceptance of the randomized experiment in social and health sciences research, and perhaps there is a similar lag being observed with RD (Shadish, Cook, and Campbell 2002). This is a challenge to health services researchers who need to find publication outlets for their findings. It is a challenge we hope this article will help address.

ACKNOWLEDGMENTS

This study was partially supported by the Agency for Healthcare Research and Quality grant 1R24HS11673. The authors acknowledge the staff of Pharmaceutical Research Computing, School of Pharmacy, University of Maryland, Baltimore, for assistance with data preparation. We also thank two anonymous reviewers of this manuscript for their valuable input. There are no disclosures or disclaimers.

REFERENCES

- Ahluwalia, J. S., M. L. Weisenberger, A. M. Bernard, and S. E. McNagny. 1996. “Changing Physician Prescribing Behavior: A Low-Cost Administrative Policy That Reduced the Use of Brand Name Nonsteroidal Anti-Inflammatory Drugs.” *Preventive Medicine* 25 (6): 668–72.
- Anguist, J. D., and V. Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114: 533–75.

- Carter, G. M., J. D. Winkler, and A. K. Biddle. 1987. *An Evaluation of the NIH Research Career Development Award. R-3568-NIH*. Santa Monica, CA: T. R. Corporation.
- Collins, T. M., D. A. Mott, W. E. Bigelow, and D. R. Zimmerman. 1997. "A Controlled Letter Intervention to Change Prescribing Behavior: Results of a Dual-Targeted Approach." *Health Services Research* 32 (4): 471-89.
- Cullen, K. W., L. M. Koehly, C. Anderson, T. Baranowski, A. Prokhorov, K. Basen-Engquist, D. Wetter, and A. Hergenroeder. 1999. "Gender Differences in Chronic Disease Risk Behaviors through the Transition out of High School." *American Journal of Preventive Medicine* 17 (1): 1-7.
- Daniels, V., M. Somers, J. Orford, and B. Kirby. 1992. "How Can Risk Drinking amongst Medical Patients Be Modified? The Effects of Computer Screening and Advice and a Self-Help Manual." *Behavioural Psychotherapy* 20 (1): 47-60.
- Farris, K. B., D. M. Kirking, L. A. Shimp, and R. A. Opdycke. 1996. "Design and Results of a Group Counter-Detailing DUR Educational Program." *Pharmaceutical Research* 13 (10): 1445-52.
- Finkelstein, M. O., B. Levin, and H. Robbins. 1996a. "Clinical and Prophylactic Trials with Assured New Treatment for Those at Greater Risk: I. A Design Proposal." *American Journal of Public Health* 86 (5): 691-5.
- . 1996b. "Clinical and Prophylactic Trials with Assured New Treatment for Those at Greater Risk: II. Examples." *American Journal of Public Health* 86 (5): 696-705.
- Gurwitz, J. H., J. P. Noonan, and S. B. Soumerai. 1992. "Reducing the Use of H2-Receptor Antagonists in the Long-Term-Care Setting." *Journal of the American Geriatrics Society* 40 (4): 359-64.
- Hoglund, P. 1996. "Long-Term Effects of Transference Interpretations: Comparing Results from a Quasi-Experimental and a Naturalistic Long-Term Follow-Up Study of Brief Dynamic Psychotherapy." *Acta Psychiatrica Scandinavica* 93 (3): 205-11.
- Johnston, M. V., K. J. Ottenbacher, and C. S. Reichardt. 1995. "Strong Quasi-Experimental Designs for Research on the Effectiveness of Rehabilitation." *American Journal of Physical Medicine and Rehabilitation* 74 (5): 383-92.
- Kerlinger, F., and H. Lee. 2000. *Foundations of Behavioral Research*. Orlando, FL: Harcourt College Publishers.
- Lee, E., D. L. McNally, and I. H. Zuckerman. 2004. "Evaluation of a Physician-Focused Educational Intervention on Medicaid Children with Asthma." *Annals of Pharmacotherapy* 38 (6): 961-6.
- Monane, M., D. M. Matthias, B. A. Nagle, and M. A. Kelly. 1998. "Improving Prescribing Patterns for the Elderly through an Online Drug Utilization Review Intervention: A System Linking the Physician, Pharmacist, and Computer." *Journal of the American Medical Association* 280 (14): 1249-52.
- National Heart, Lung and Blood Institute. 1997. *Expert Panel Report II; Guidelines for the Diagnosis and Management of Asthma. National Asthma Education Program*. Bethesda, MD: National Heart, Lung and Blood Institute.

- Pitt, M. M., and S. R. Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy* 106: 958–96.
- Shadish, W. R., T. D. Cook, and D. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Shadish, W. R. Jr., T. D. Cook, M. B. Church, S. Ajanaku, J. R. Kim, and R. Cohen. 1996. "The Development of Occupational Aspirations and Expectations among Inner-City Boys." *Child Development* 67 (6): 3368–85.
- Thistlewaite, D., and D. Campbell. 1960. "Regression Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment." *Journal of Educational Psychology* (51): 309–17.
- Trochim, W. 2003. "The Research Methods Knowledge Base" [accessed on August 15, 2003, 2003]. Available at <http://trochim.human.cornell.edu/kb/statrd.htm>
- Trochim, W. M., and J. C. Cappelleri. 1992. "Cutoff Assignment Strategies for Enhancing Randomized Clinical Trials." *Controlled Clinical Trials* 13 (3): 190–212.
- Trochim, W. M. K. 1984. *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Beverly Hills, CA: Sage Publications.
- . 1990. "The Regression-Discontinuity Design." In *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*, edited by L. Sechrest, P. Perrin, and J. Bunker, pp 119–39. Washington, DC: U.S. Dept. HHS, Public Health Service, Agency for Health Care Policy and Research, PHS 70-3454.
- Wagner, A. K., S. B. Soumerai, F. Zhang, and D. Ross-Degnan. 2002. "Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research." *Journal of Clinical Pharmacy and Therapeutics* 27 (4): 299–309.
- Wooldridge, J. M. 2003. *Introductory Econometrics: A Modern Approach*. Mason, OH: Thomson South-Western.