

# Why Research Design and Methods Is So Crucial to Understanding Drug Use/Abuse: Introduction to the Special Issue

Evaluation & the Health Professions

2018, Vol. 41(2) 135-154

© The Author(s) 2018

Reprints and permission:

[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)

DOI: 10.1177/0163278718756161

[journals.sagepub.com/home/ehp](http://journals.sagepub.com/home/ehp)



Lawrence M. Scheier<sup>1</sup>

## Abstract

The collection of articles in this special issue both raise the bar and inspire new thinking with regard to both design and methodology concerns that influence drug use/abuse research. Thematically speaking, the articles focus on issues related to missing data, response formats, strategies for data harmonization, propensity scoring methods as an alternative to randomized control trials, integrative data analysis, statistical corrections to reduce bias from attrition, challenges faced from conducting large-scale evaluations, and employing abductive theory of method as an alternative to the more traditional hypothetico-deductive reasoning. Collectively, these issues are of paramount importance as they provide specific means to improve our investigative tools and refine the logical framework we employ to examine

---

<sup>1</sup>LARS Research Institute, Inc., Scottsdale, AZ 85260, USA

## Corresponding Author:

Lawrence M. Scheier, LARS Research Institute, Inc., 15029 N, Thompson Peak Parkway, Suite B111-443, Scottsdale, AZ 85260, USA.

Email: [scheier@larsri.org](mailto:scheier@larsri.org)

the problem of drug use/abuse. Each of the authors addresses a specific challenge outlining how it affects our current research efforts and then outlines remedies that can advance the field. To their credit, they have included issues that affect both etiology and prevention, thus broadening our horizons as we learn more about developmental processes causally related to drug use/abuse and intervention strategies that can mitigate developmental vulnerability. This is the essential dialogue required to advance our intellectual tool kit and improve the research skills we bring to bear on the important questions facing the field of drug use/abuse. Ultimately, the goal is to increase our ability to identify the causes and consequences of drug use/abuse and find ways to ameliorate these problems as we engage the public health agenda.

### **Keywords**

missing data, abductive reasoning, propensity scoring, weighting, evaluation

All life is problem solving

Karl R. Popper (1999).

### ***Buddhist Leanings***

I went through a certain “stage” in my life, very different from a Piagetian cognitive developmental stage, but still a stage that was very informative and helped to crystallize my identity. For a while, I was deeply immersed in Buddhist readings (Suzuki, 1949), Vedic stuff, and I coupled this effort with an existential search for meaning (Frankl, 1959). To be quite frank, I credit this brief hiatus from scientific reasoning for a spiritual awakening that drove me to pursue a career in psychology. During the course of this experience, I read a book by the Indian philosopher and spiritual teacher Jiddu Krishnamurti.<sup>1</sup> The book was titled *Think on These Things* (1964) and it contained a collage of personal narratives addressing several fundamental concerns regarding the basic composition of the “good” life. In his own contemplative nontheosophical style, Krishnamurti discussed education, love, freedom, ambition, beauty, conformity, and self-discipline to name a few topics. The book is truly a reflection outlining his own psychological awakening, cast in terms of life’s fundamental challenges, who are we, what is our purpose, and what should we be doing? Remembering back from when I read this influential book, I believe that his authorial goal was to

capture the reader's imagination and make him or her realize that the way we think about things provides a pretext for how we live and make sense of our world.

Recounting the influence of this book serves merely as backdrop to my present role as a guest editor of this *special issue*, but it gives some credence to the issue's focus on *design and methodology*. Charged with the task of putting together a special issue, I took it upon myself to accomplish two things in particular. First, I wanted the contributors to tackle some important topics that allude to who we are as scientists (i.e., our identity) and especially how we conceptualize problems. Second, and harking back to my overtures to carefully *think about these things*, I wanted that the contributing authors examine some of the more trenchant issues we face studying drug use/abuse but from a different perspective. Finding ways to advance our knowledge base with different tools is what prompted me to ask contributors to "free wheel" a bit and think outside the proverbial box (i.e., recall I asked them to *think on these things*). I then cautioned them to direct their energies and focus toward design and methodological issues that are of paramount importance to research in drug use/abuse.<sup>2</sup>

### *The Future Is Here Now*

Design and methodology are both critical and fundamental to how we address psychological issues, they are no doubt the foundation upon which rests our experimental rigor (Cook & Campbell, 1979). The design and methodological tools we possess in our scientific tool kit allow us to exact rigorous control, statistical, and otherwise and identify the causal processes responsible for behavior change. In the case of alcohol and drug etiology, we need methods that enable us to partition complex relations reflecting personality, peer, family, and community influences, all of which are eventual targets of intervention. Prevention scientists benefit from the careful scrutiny of causal models adapting them to create logic models containing specific intervention postulates. Regardless of whether our focus rests with etiology or prevention, we test and refine these models with statistical tools based on rigorous methods and study designs that enable us to infer causality. Along the way, we test confounders, manipulate variables through transformation, prune irrelevant variables, refine our understanding of the causal processes that set vulnerability into motion, and eventually achieve parsimonious explanations of the causes, prevention, and treatment of drug use/abuse.

The field of prevention science is well aware that it faces its own trenchant issues (e.g., Greenberg, 2008). Take, for instance, a program evaluation that requires parsing variance in the dependent measure into the portion attributed to the school or classroom and the portion ascribed to the individual. Statistical innovations like hierarchical linear modeling (Raudenbush, 1997) allow us to deal with the phenomenon of clustering, where students are nested within classrooms or schools. Although clustering estimates are somewhat small as evidenced in smoking (e.g., Murray et al., 1994) and drug prevention trials (Scheier, Griffin, Doyle, & Botvin, 2002), failure to address them increases the risk of making a Type I error and biasing parameter estimates. Consider a second example, where we examine program effects on “trends” in drug use using growth modeling. One can recall that, up until recently, program efficacy was examined either by contrasting group means between experimental conditions using analysis of covariance (ANCOVA) or by hypothesizing “static” influences using fixed-effect linear regression or some type of structural path model. Advances in methodology now provide the tools to contrast slope trajectories between treated and control students with a myriad of ways to model baseline intercepts (e.g., Mason, Kosterman, Hawkins, Haggerty, & Spoth, 2003; Taylor, Graham, Cumsille, & Hansen, 2000).

This shift in our emphasis (and hypotheses) resulted from methodological improvements and statistical innovations that have made it possible to examine the timing of intervention effects and do this in the context of dynamic features of development.<sup>3</sup> Indeed, as a field, we have witnessed greater emphasis on *developmental epidemiology* (Kellam, Koretz, & Mościcki, 1999), and this has provided impetus for investigators to model program effects on developmental trajectories of target risk mechanisms and behavior across time. Investigators employing these techniques are encouraged to possess a priori hypotheses regarding differential rates of growth in target risk mechanisms or outcome behaviors following intervention exposure. Furthermore, these methods would require at a minimum to have measures of preintervention (baseline), postintervention, and follow-up data. Even more data collection is plentiful, generally boosts power (Petras, 2016), and provides a means to assess curvature and perturbation or decay of effects (e.g., Park et al., 2000). Worth noting, however, is the considerable gap in time between the introduction of drug prevention efficacy trials (1970s) and the application of latent growth modeling for program evaluation (1990s). In these intervening years, the tool kit we use to assess program efficacy expanded considerably. Given another decade or more, we will most assuredly witness introduction of methodological approaches

with increasing complexity that will influence the way we go about our business in years to come.

### *All Things Being Equal*

My overtures to the contributing authors were not by any stretch of the imagination a “blind” or unsupported request. Searches of the general behavioral science literature indicate that numerous fields have copiously addressed design and methodological issues (e.g., Clarke, 1995; Ledgerwood, 2014). These efforts have been extended to include studies of sexual risk (Schroder, Carey, & Vanable, 2003) and, closer to home, issues related to evaluating prevention programs (Masseti, Simon, & Smith, 2016). Although the conceptual foundations of these studies may vary, they provide a template upon which to understand how different fields address their methodological challenges and the ways in which they formulate their concerted responses.

As part of meeting this challenge, we revisit the sanctity of the randomized control trial (RCT), which has been called the “gold standard” in experimental research (e.g., Cartwright, 2007; Sanson-Fisher, Bonevski, Green, & D’Este, 2007). We use RCTs for the level of control they offer by equating extraneous factors (systematic sources of bias are made random) between experimental conditions that could influence treatment outcomes (e.g., what Cook and Campbell called *probabilistic equivalence*). Confounding is, to say the least, messy, and RCTs provide a strong research design that achieves balance of pretreatment covariates and elevates our ability to infer causation by producing unbiased estimates of treatment effects (e.g., Rubin, 1974).

There will, however, be situations where an RCT is not feasible, for ethical or design considerations, or because they are cost prohibitive. In some cases, particularly when working with marginalized or racial/ethnic populations, RCTs may not be deemed as culturally acceptable (Henry, Tolan, Gorman-Smith, & Schoeny, 2017). Moreover, under real-world conditions, and given choices regarding adopting an evidence-based treatment, many communities or schools will summarily reject the concept of random assignment. Holding back effective treatments is not in the best interests of public policy, particularly with regard to schools facing mounting pressures to eradicate behavioral problems and vying for effective “health promotion” programs. It is not uncommon then, even with RCT designs that contamination (i.e., diffusion of treatments) and compensatory rivalry will come into play, biasing findings in ways that were beyond the researcher’s

control.<sup>4</sup> According to Rubin (2005), contamination is a violation of the *stable unit treatment value assumption*, which posits that for RCTs to work, the actions of one set of units (treatment) cannot affect the actions of the other (control group). In light of the potential for this to happen in the “real world,” we need to shift our mind-set to alternative strategies that can yield the same level of statistical control and experimental rigor.<sup>5</sup>

There are other examples of challenges that have produced thinking *outside the box*. Null hypothesis testing has been scrutinized by several authors (Cohen, 1994; Krantz, 1999; Loftus, 1996) and yielded debate about practical versus statistical significance (Jacobson & Truax, 1991). The conceptual foundation behind utilizing discrete diagnostic “classification” as the foundation of psychiatric nosology has also been questioned (Meehl, 1995) as has the issue of dimensional versus categorical classification (Muthén, 2006). There have been criticisms of dichotomization to truncate meaningful quantitative measures, the end result losing important information, and biasing correlations (MacCallum, Zhang, Preacher, & Rucker, 2002). The latter approach is frequently used to transform highly skewed frequency of drug use measures encountered with relatively young samples prior to drug use debut.<sup>6</sup> Other possible methodological challenges include the effects of reporting bias, sample selection bias, participation bias, recall effects, survey methodology (ACASI vs. paper and pencil), and measurement concerns (i.e., reliability and temporal measurement invariance). In each case, these illustrations remind us that the way we “think” or perceive the problem can be a shackle to real discovery.

### *How We Got Here*

The focus on design and methodology as it influences drug use/abuse research stems from my own long-standing professional interests in studies of drug etiology (Scheier, 2010) and prevention (Scheier, 2015). However, there is more to that story as well. The history of studies of drug use/abuse and the broader issues that surround studies of deviant behavior in general is rife with examples of design or methodological challenges. For one thing, it would be ethically reprehensible to experimentally assign people to become drug users; thus, we have to implement alternative designs that allow us to examine the natural developmental course of drug use. This has been a strong suit for etiology over the years, evidenced by numerous naturalistic longitudinal studies of youth at different ages, taking place in many different settings using samples with heterogeneous racial composition (Scheier, 2001). Statistically speaking, low base rates for self-reported drug use in

youthful populations have also caused some consternation. Solutions have included negative binomial or Poisson regression, zero-inflated count models (Buu, Li, Tan, & Zucker, 2012), and two-part semicontinuous models (Olsen & Schafer, 2001); all techniques that are robust to highly skewed distributions (e.g., Atkins & Gallop, 2007).

In dealing with the prevention side of the equation, we face similar challenges. A handful of examples include dealing with attrition and selection effects (Hill, Rosenman, Tennekoon, & Mandal, 2013), noncompliance in treatment studies (Jo, Ginexi, & Ialongo, 2010) as well as prevention trials (Jo, 2002; Little & Yau, 1998; Stuart, Perry, Le, & Ialongo, 2008), assessing fidelity (e.g., Hill, Maucione, & Hood, 2007; Lee et al., 2008), and implementation concerns (e.g., Crowley, Coffman, Feinberg, Greenberg, & Spoth, 2014; Payne & Eckert, 2010); all factors that influence going to scale (e.g., Botvin & Griffin, 2010).

Despite these scientific challenges, there are numerous examples of how our colleagues' design alternatives and methodological enhancements have become a beacon for other behavioral scientists. This includes work examining the validity and accuracy of self-report (e.g., Del Boca & Darkes, 2003), application of missing data estimation methods using the three-form planned missing design (e.g., Graham, Taylor, Olchowski, & Cumsille, 2006), advanced statistical treatment of nonresponse (e.g., Enders, 2010), addressing clustering that occurs in group-randomized trials (e.g., Murray & Hannan, 1990), alternative randomized designs (e.g., Brown & Liao, 1999), person-centered classificatory strategies to test program effects with subgroups (e.g., Lanza & Rhoades, 2013), and the decomposition of program effects using mediation analyses (e.g., MacKinnon, Taborga, & Morgan-Lopez, 2002; Scheier, Botvin, & Griffin, 2001) to name a few.<sup>7</sup> To me, as important as these challenges are to both etiology and prevention, they only scratch the surface. It should be clear that there are even greater challenges that await us, and this special issue is devoted to sketching out a few of these challenges in a rudimentary, conceptually driven manner.

### *Article Anthology*

With this in mind, the articles in this special issue address different design issues and methodological challenges that influence drug use/abuse research. In the first article, Mason and colleagues bring into question the hegemony of hypothetico-deductive reasoning and offer a completely different view of the scientific reasoning that we apply to support logical inferences. Their counterpart is the *abductive theory of method*, which has

strong roots in philosophical thinking regarding theories of evidence. Unlike hypothetico-deductive reasoning, which deduces from cause to effect, the abductive theory of method works “backward” from the problem to find the best fitting explanation(s). Included in their wonderful article is a challenge to the supremacy of the RCT. For a variety of reasons, the RCT does not offer the best experimental approach or offer the certitude required to resolve several issues at hand (i.e., studies of interventions seeking policy shifts or environmental interventions involving whole communities). After they remind us of the various caveats that come along with design and methods, they offer several recommendations to push the event horizon, with each suggestion building a composite of reasoning and understanding regarding why a prevention program or intervention works.

In the second article, Hansen and colleagues lay the foundation for data harmonization and manipulations that will eventually provide a means to create faux “virtual” controls. Their motivation for integrating data is the tremendous costs associated with collecting data on concurrent control students (or schools) that are part of drug prevention trials. Controls represent the counterfactual or potential outcomes that would occur in the absence of an intervention. Finding ways to reasonably and cost-effectively create *synthetic comparators* (Hansen, Derzon, & Reese, 2014) creates cost economies that would have ripple effects in the prevention industry, to say the least. Synthetic comparator controls are derived from extant data that can be used to evaluate already disseminated drug prevention studies or local programs where funding is not sufficient to collect data on both treated and control youth. However, we currently lack systematized strategies to engage data harmonization procedures leading Hansen and colleagues to produce a workable framework and then testing its heuristic utility. Their article elaborates several crucial considerations including the necessary steps to create “conceptual” concordance between items and scales from different studies. They also apply Random Forest imputation procedures to handle missing data and using receiver operating characteristic curves to evaluate model performance. The end result of their data harmonization efforts is to achieve parsimony and efficiency with an eye cast toward cost-effectiveness.

The past few years has seen increasing discussion, acrimonious at times, regarding the replicability crisis in psychology (Maxwell, Lau, & Howard, 2015), and this has been extended to other disciplines as well (e.g., Goodman, Fanelli, & Ioannidis, 2016; Valetine et al., 2011). A good deal of this discussion revolves around the need for clearer guidelines in reproducing study findings particularly noting the differences between “exact” and

conceptual replications and the manner in which we can discover regularity in human behavior (Stroebe & Strack, 2014). Clearly, this debate applies to intervention effects that may be considered “specious” if not replicated when moving from efficacy to effectiveness trials. Scientists have a much greater chance to detect reliable and consistent behavior patterns as well more judiciously test theory if they “pool” existing data sets, overcoming the limitations that may be ascribed to smaller underpowered studies when examined in a stand-alone manner. In the next article, Curran and colleagues highlight this important discussion over replication and then illustrate the many benefits of integrated data analysis (IDA). They typify its application by examining youth polydrug use with multiple longitudinal samples delineated based on whether a child’s parent was an alcoholic. In this particular case, they ask whether polysubstance use is theoretically and psychometrically consistent when assessed as a latent construct across several studies.

The studies share similar design profiles (i.e., following children of alcoholics longitudinally) and are all concerned about the deleterious effects of an alcoholic parent on child outcomes. However, they don’t contain the exact same observed measures of polydrug use, requiring strategies aimed at achieving data harmonization, weighting, and imputation. The fact that three studies are used provides a rigorous means to test the “idiosyncrasies” of the study, its population, design, and setting versus the staying power of the ideas being tested (i.e., theoretical representations). The authors outline the numerous strengths of IDA (i.e., stretched cohorts increasing the age span studied) and also note its limitations because it is not a panacea in some cases (i.e., measurement differences between studies). Still, it is a very important tool in the arsenal as we combat detractors who suggest prevention science consists of a smattering of small studies, producing small effects sizes and failing to uncover reproducible behavioral patterns that demonstrate the lasting utility of our efforts.

In the fourth article, Chang and Little synthesize three design innovations including a multiform questionnaire protocol using planned missingness, visual analog scaling (VAS), and retrospective pretest–posttest design. From an evaluation point of view, blending these design tactics is quite novel for several reasons, not the least of which is cost-efficiency. Their article includes a wonderful exposition on the fundamentals of missing data and the use of multiform protocols. The cost-efficiencies represent one advantage, reducing negative emotions from participants faced with burdensome surveys is another one. The overall gains include increased power and precision in the parameter estimates. The second offering is visual

analog scaling, perhaps the strongest response to the multi-item Likert-type scales (e.g., Guyatt, Townsend, Berman, & Keller, 1987), offering a surefire alternative to the categorical scale problem. Visual analog scales include extreme anchors with clear interval level demarcations (60% is twice 30%)<sup>8</sup> that is not available with categorical Likert-type scales (4 or *agree* is not twice 2 or *disagree* response). Moreover, the VAS is perhaps more sensitive (i.e., responsive) to detect clinically significant behavior “changes” over time (Hasson & Arnetz, 2005). The final piece in the equation is the retrospective pretest–posttest design, which also provides cost-efficiencies (eliminating the pretests entirely), and avoids response shift bias (Howard, 1980). Asking the respondent at posttest how they feel “at this time” or currently and how they felt “when they started the study” (pretest) levels the playing field with respect to their “frame of reference” or the internal standard respondents use to gauge their attitudes or behavior. This approach also eliminates any bias that may arise from overestimation of one’s capabilities, behavior, or attitudes, producing a more veridical estimation. If there is any change worth noting, the respondent can answer based on their own true assessment, without having to oscillate back and forth searching for a frame of reference (respondents need “anchors” to set their mind on some internal comparison; Jacowitz & Kahneman, 1995), which is lacking in a true pretest. Failure to address response shift bias poses a threat to internal validity. The authors then coalesce their methodological innovations using two excellent empirical illustrations using longitudinal data.

In the fifth article, Rhew and colleagues examine program effects of a relatively large-scale community-based drug prevention intervention using inverse probability weighting (IPW) to adjust for students leaving their school catchment area during the course of the longitudinal study. No RCT can be completely pristine, and students may exit a study for a variety of reasons, some academic (i.e., performance), some based on socioeconomic factors (i.e., parent’s earning potential, job relocation), and extraneous factors (i.e., neighborhood factors compel parents to move). Regardless of the prime reason for relocating, failing to address subject loss (especially if this loss differs by condition) threatens the internal validity of the study and can bias findings. Earlier multilevel analyses using the “intention-to-treat” method provided evidence of favorable intervention effects. The reanalysis using IPW considers “exposure” as a major factor in prevention outcomes. This provides a more sophisticated means of grappling with the basic tenets of prevention science (i.e., “how much dose is enough?”) and also creative ways to handle missing data. The article discusses additional strengths of the IPW approach including maintaining the integrity of the RCT design by

avoiding biases that arise from postrandomization data manipulations. Overall, the authors report there are clear benefits to youth staying in the communities assigned to the experimental condition, which gave them more exposure to prevention activities and was associated with better behavioral outcomes compared to control youth, who also stayed in their community.

In the sixth article, Tein and colleagues tackle the perennial problem of high rates of dropout encountered with randomized community-based interventions. In their case, they offered court mandated parenting classes to divorced parents in an effort to improve parenting skills (e.g., ability to navigate divorce with reduced parent conflict) and improve their children's behavioral outcomes. The problem is that many parents signed up and then showed up for at most one or two classes. In some cases, parents failed to show up at all and basically dropped out of the study. The subject loss undermines power and also makes it hard to discern intervention efficacy. To develop a methodology that can resolve this issue, the authors hypothesized that some exposure to the "active ingredients" is better than nothing. However, the study design called for an active control group receiving some form of intervention. To rectify this situation, the authors applied a propensity scoring method also using IPW to reintroduce dropout parents back into the analytic scheme. Essentially, they demarcated parents that attended any sessions, those administered the "active control" condition, and parents in any condition who attended no sessions at all. With this postrandomization strategy in hand, they used propensity score methods to "balance" covariates between the conditions, essentially preserving the ability to make strong causal inferences regarding program effects. In so doing, they effectively reduced the high dimensionality of the covariates (one scalar weight vs. numerous covariates), balancing the different experimental groups on preexisting conditions that could confound outcomes.<sup>9</sup> By reducing bias that might exist between conditions, even after program implementation, the strategy offers a modicum of precision to ascertain whether the intervention "worked," particularly when contrasting parents with some intervention exposure (i.e., a dose effect) versus those that had nothing.<sup>10</sup>

Shifting gears somewhat, in the final article in this special issue, Derzon recounts the many pitfalls we encounter conducting large-scale evaluations; providing potential remedies that we have at our fingertips. One formidable issue in evaluation science is the inclusion of "macro" or host setting measures that affect how well a program is implemented. Derzon calls these the "activities, contexts, and characteristics of the settings implementing the intervention," but more generally, they are the "conditions of

implementation.” The quandary faced by an evaluator is “what happens when we implement the same program in two different settings and obtain two different results?” In order to formulate an answer, it is essential that we first recognize that macro or contextual factors *can and do* affect implementation and trickle down to the level of individual behavior change. Implementation research repeatedly shows that teachers lacking “buyin,” feeling that their training was insufficient, or believing the school organizational climate failed to support the intervention don’t deliver the same precise curriculum as teachers who possess greater passion and zeal (e.g., Pas, Waasdorp, & Bradshaw, 2015; Wanless, Patton, Rimm-Kaufman, & Deutsch, 2013). Recognizing the sheer impact of these setting-level factors can help to improve our understanding of how to maximize positive program outcomes (Durlak & DuPre, 2008). Requirements in the process of evaluating setting influences (and implementation quality) include obtaining reliable and valid measures that can be used to monitor organizational and training factors that may inhibit program delivery (i.e., moderate program effects). Also, finding the right balance between qualitative and quantitative methods, use of existing archival data (i.e., putting a finger on the pulse of organizational factors), and implementing a sophisticated design (e.g., rapid cycle evaluations that are used by hospitals) to tease apart the myriad of different setting effects are all factors that can improve precision in estimating program effects (and intersite differences). Truth be told, if a program evaluation entails 30 schools or 30 communities, each site is really an intervention of its own; an independent “trial” with its own set of crisis revolving around delivering the intervention with fidelity.

## Conclusion

The articles in this special issue represent a vanguard of design and methodological approaches that are set to push the event horizon in prevention science for years to come. Importantly, the articles suggest cutting-edge techniques that can provide unique insight into the questions we pose about etiology and prevention. This is the stamp of imprimatur for those emphasizing methodology, and their mission involves providing answers to questions about very deep and profound issues, in our case the causes, consequences, and prevention of drug use/abuse. This is no easy task, as I have reiterated in several places. It should be patently clear to readers that only when we consider the challenges made threadbare in the special issue articles that we will ultimately avail ourselves of the tools required to pursue scientific resolutions to very important human problems. The

articles in this special issue go a long way toward reinforcing that our techniques require refinement, we need new ways of seeing things, and the balance of our work remains in front of us.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **Notes**

1. I actually came across this book while on a lengthy bike ride in Southern California, back behind Saddleback Mountain east of Mission Viejo, located in Orange County amid windy country roads adorned with California poppies, wild anise, and draped in the shade of large oak trees. I came across a dirt road leading to a Zen Buddhist monastery that actually maintained a bookstore open to the public. I purchased a copy of the book, stuffed it into the rear pocket of my bike shirt, and read it sometime afterward.
2. I also asked the contributing authors to provide nontechnical papers, as there are sufficient publications that address the technical aspects, for instance of propensity scoring, interrupted times series, regression discontinuity regression point displacement, stepped wedge, and dynamic waitlist (rollout) designs. The purpose of the current special issue was to unveil the conceptual foundations for alternative design and methodological strategies and show their intended application and ramifications for the way we study drug use/abuse. It is worth mentioning that the collections of articles do, however, neglect the intersection of methodology and qualitative research, a good discussion of which can be found in Creswell (2009).
3. McArdle and Epstein (1987) should perhaps be credited with introducing growth modeling as a formidable tool in developmental research; however, it was not a household term in drug prevention studies until the 1990s. Several noteworthy examples of how etiology can benefit from elucidating growth trajectories include longitudinal studies examining relations between refusal skills and alcohol use (e.g., Scheier, Botvin, Griffin, & Diaz, 1999), self-esteem and alcohol use (Scheier, Botvin, Griffin, & Diaz, 2000), alcohol use and bar patronage (e.g., Curran, Harford, & Muthén, 1996), and drug use in the context of peer selection versus socialization mechanisms (Wills & Cleary, 1999).

4. Derzon (2014) even points out that political, economic, social, and cultural factors play an integral role in the adoption of programs and the use of randomized control trial (RCTs) to test efficacy. Highly motivated school districts being handed a program and given the resources for implementation are not like the remaining “world” of school districts that may flounder if asked to support their own implementation of an evidence-based program. Thus, RCTs may hinder generalizability of the intervention tested because of the considerable “investment of resources” given schools to test the efficacy of a new or novel program.
5. There is also a good reason to consider that RCTs often work with units that can be randomized but may not be the optimal focus of an intervention. Traditionally, we randomize students, classrooms, or schools to experimental condition. We may want to, however, randomize intact social units based on the level of social persuasion. This would entail delivering interventions to intact “social networks” (e.g., Valente, Hoffman, Ritt-Olson, Lichman, & Johnson, 2003) that may be responsible for drug initiation (peers influence each other at the level of close friendship networks). An example of a network-randomized design would entail conducting small group discussions among peers to dissuade them from conforming to deviant behavior. This type of “ring” design is backed by research examining the peer ecology of drug use and peer deviance talk (e.g., Piehler & Dishion, 2007).
6. Three good examples of where this technique was applied in drug prevention studies can be found in Sun, Sussman, Dent, and Rohrbach (2008), Hansen, Johnson, Flay, Graham, and Sobel (1998), and G. J. Botvin, Baker, Dusenbury, and E. M. Botvin (1995). In all three cases, highly skewed distributions for frequency of use measures led to dichotomization, resulting in a use/nonuse measure that was used to demonstrate program efficacy by modeling program effects on “ever use.”
7. Decomposition of intervention effects in multimodal programs can be achieved with additive or dismantling designs, fractional factorial designs (e.g., Chakraborty, Collins, Strecher, & Murphy, 2009), and componential analysis (West & Aiken, 1997); all rigorous design approaches that tease apart which component provides the biggest bang for the buck.
8. Hasson and Arnetz (2005) point out there is some debate over the appropriate use of parametric (i.e., interval data) versus nonparametric (i.e., ordinal data) statistics for VAS. This debate includes reference to the setting and intended uses of the scale, whether there is preparatory training in scale use, and the suitability for younger populations.
9. As they explain in detail, the balancing is done using generalized boosted regression with a tree-like data-adaptive (i.e., recursive) approach very much

like flexible regression tree approaches that rely on successive splitting to achieve the best model fit (this latter component uses a stochastic subsampling approach).

10. At this point in the comparison, assuming a Gaussian distribution, the outcome is independent of the treatment conditional on the covariates, and consistent with the stable unit treatment value assumption, the comparison reduces to a simple ANCOVA estimating the treatment effect contrasting “weighted” means between groups.

## References

- Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology, 21*, 726–735.
- Botvin, G. J., Baker, E., Dusenbury, L., & Botvin, E. M. (1995). Long-term follow-up of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association, 273*, 1106–1112.
- Botvin, G. J., & Griffin, K. W. (2010). Advances in the science and practice of prevention: Targeting individual-level etiologic factors and the challenge of going to scale. In L. M. Scheier (Ed.), *Handbook of drug use etiology: Theory, methods, and empirical findings* (pp. 631–650). Washington, DC: American Psychological Association.
- Brown, C. H., & Liao, J. (1999). Principles for designing randomized preventive trials in mental health: An emerging developmental epidemiology paradigm. *American Journal of Community Psychology, 27*, 673–710.
- Buu, A., Li, R., Tan, X., & Zucker, R. A. (2012). Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine, 31*, 4074–4086.
- Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*, 11–20.
- Chakraborty, B., Collins, L. M., Strecher, V. J., & Murphy, S. A. (2009). Developing multicomponent interventions using fractional factorial designs. *Statistics in Medicine, 28*, 2687–2708.
- Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63*, 718–725.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist, 49*, 997–1003.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Harcourt.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed-method approaches*. Thousand Oaks, CA: Sage.

- Crowley, D. M., Coffman, D. L., Feinberg, M. E., Greenberg, M. T., & Spoth, R. L. (2014). Evaluating the impact of implementation factors on family-based prevention programming: Methods for strengthening causal inference. *Prevention Science, 15*, 246–255.
- Curran, P. J., Harford, T. C., & Muthén, B. O. (1996). The relation between heavy alcohol use and bar patronage: A latent growth model. *Journal of Studies on Alcohol, 57*, 410–418.
- Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction, 98*, 1–12.
- Derzon, J. H. (2014). Differences-indifferences in approaches. In G. J. N. Bruinsma, & D. L. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 1054–1064). New York, NY: Springer Science+Business Media.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Frankl, V. E. (1959). *Man's search for meaning*. New York, NY: Beacon Press.
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*, 1–6.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343.
- Greenberg, M. T. (2008). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science, 5*, 5–13.
- Guyatt, G. H., Townsend, M., Berman, L. B., & Keller, J. L. (1987). A comparison of Likert and visual analogue scales for measuring change in function. *Journal of Chronic Diseases, 40*, 1129–1133.
- Hansen, W. B., Derzon, J. H., & Reese, E. L. (2014). A synthetic comparator approach to local evaluation of school-based substance use prevention programming. *Evaluation & the Health Professions, 37*, 258–282.
- Hansen, W. B., Johnson, C. A., Flay, B. R., Graham, J. W., & Sobel, J. (1998). Affective and social influences approaches to the prevention of multiple substance abuse among seventh grade students: Results from Project SMART. *Preventive Medicine, 17*, 135–154.
- Hasson, D., & Arnetz, B. B. (2005). Validation and findings comparing VAS vs. Likert scales for psychosocial measurements. *International Electronic Journal of Health Education, 8*, 178–192.
- Henry, D., Tolan, P., Gorman-Smith, D., & Schoeny, M. (2017). Alternatives to randomized control trial designs for community-based prevention evaluation. *Prevention Science, 18*, 671–680.

- Hill, L. G., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science, 8*, 25–34.
- Hill, L. G., Roseman, R., Tennekoon, V., & Mandal, B. (2013). Selection effects and prevention program outcomes. *Prevention Science, 14*, 557–569.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4*, 93–106.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161–1166.
- Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics, 27*, 385–409.
- Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science, 11*, 384–396.
- Kellam, S. G., Koretz, D., & Mościcki, E. K. (1999). Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology, 27*, 463–482.
- Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association, 44*, 1372–1381.
- Krishnamurti, J. (1964). *Think on these things*. Ojai, CA: Krishnamurti Foundation of America.
- Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*, 157–168.
- Ledgerwood, A. (2014). Introduction to the special section on advancing our methods and practices. *Perspectives on Psychological Science, 9*, 275–277.
- Lee, C.-Y., August, G. J., Realimuto, G. M., Horowitz, J. L., Bloomquist, M. L., & Klimes-Dougan, B. (2008). Fidelity at a distance: Assessing implementation fidelity of the Early Risers Prevention Program in a going-to-scale intervention trial. *Prevention Science, 9*, 215–229.
- Little, R. J., & Yau, L. H.-Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods, 3*, 147–159.
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 161*–171.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

- MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence, 68*, S69–S83.
- Mason, W. A., Kosterman, R., Hawkins, J. D., Haggerty, K. P., & Spoth, R. L. (2003). Reducing adolescents' growth in substance use and delinquency: Randomized trial effects of a parent-training prevention intervention. *Prevention Science, 4*, 203–212.
- Masseti, G. M., Simon, T. R., & Smith, D. G. (2016). Methodological and design considerations in evaluating the impact of prevention programs on violence and related health outcomes. *Prevention Science, 17*, 779–784.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist, 70*, 487–498.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275.
- Murray, D. M., & Hannan, P. J. (1990). Planning for the appropriate analysis in school-based drug-prevention studies. *Journal of Consulting and Clinical Psychology, 58*, 458–468.
- Murray, D. M., Rooney, B. L., Hannan, P. J., Peterson, A. V., Ary, D. V., Biglan, A., . . . Schinke, S. P. (1994). Intraclass correlation among common measures of adolescent smoking: Estimates, correlates, and applications in smoking prevention studies. *American Journal of Epidemiology, 140*, 1038–1050.
- Muthén, B. O. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction, 101*, 6–16.
- Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semi-continuous longitudinal data. *Journal of the American Statistical Association, 96*, 730–745.
- Park, J., Kosterman, R., Hawkins, J. D., Haggerty, K. P., Duncan, T. E., Duncan, S. C., & Spoth, R. (2000). Effects of the “Preparing for the Drug Free Years” curriculum on growth in alcohol use and risk for alcohol use in early adolescence. *Prevention Science, 1*, 125–138.
- Pas, E. T., Waasdorp, T. E., & Bradshaw, C. P. (2015). Examining contextual influences on classroom-based implementation of Positive Behavior Support Strategies: Findings from a randomized controlled effectiveness trial. *Prevention Science, 16*, 1096–1106.
- Payne, A. A., & Eckert, R. (2010). The relative importance of provider, program, school, and community predictors of the implementation quality of school-based prevention programs. *Prevention Science, 11*, 126–141.

- Petras, H. (2016). Longitudinal assessment design and statistical power for detecting an intervention impact. *Prevention Science, 17*, 819–829.
- Piehler, T. F., & Dishion, T. J. (2007). Interpersonal dynamics within adolescent friendships: Dyadic mutuality, deviant talk, and patterns of antisocial behavior. *Child Development, 78*, 1611–1624.
- Popper, K. R. (1999). *All life is problem solving*. New York, NY: Routledge.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173–185.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688–701.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association, 100*, 322–331.
- Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine, 33*, 155–161.
- Scheier, L. M. (2001). Etiological studies of adolescent drug use: A compendium of data resources and their implication for prevention. *Journal of Primary Prevention, 22*, 125–168.
- Scheier, L. M. (Ed.). (2010). *Handbook of drug use etiology: Theories, methods, and empirical findings*. Washington, DC: American Psychological Association.
- Scheier, L. M. (Ed.). (2015). *Handbook of adolescent drug use prevention: Research, intervention strategies, and practice*. Washington, DC: American Psychological Association.
- Scheier, L. M., Botvin, G. J., & Griffin, K. W. (2001). Preventive intervention effects on developmental progression in drug use: Structural equation modeling analyses using longitudinal data. *Prevention Science, 2*, 91–112.
- Scheier, L. M., Botvin, G. J., Griffin, K. W., & Diaz, T. (1999). Latent growth models of drug refusal skills and adolescent alcohol use. *Journal of Alcohol and Drug Education, 44*, 21–48.
- Scheier, L. M., Botvin, G. J., Griffin, K. W., & Diaz, T. (2000). Dynamic growth models of self-esteem and adolescent alcohol use. *Journal of Early Adolescence, 20*, 178–209.
- Scheier, L. M., Griffin, K. W., Doyle, M. M., & Botvin, G. J. (2002). Estimates of intragroup dependence for drug use and skill measures encountered in school-based drug abuse prevention trials: An empirical study of three independent samples. *Health Education and Behavior, 29*, 83–101.
- Schroder, K. E., Carey, M. P., & Venable, P. A. (2003). Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytical options. *Annals of Behavioral Medicine, 26*, 76–103.

- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*, 59–71.
- Stuart, E. A., Perry, D. F., Le, H.-N., & Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science, 9*, 288–298.
- Sun, P., Sussman, S., Dent, C. W., & Rohrbach, L. A. (2008). One-year follow-up evaluation of Project Towards No Drug Abuse (TND-4). *Preventive Medicine, 47*, 438–442.
- Suzuki, D. T. (1949). *Essays in Zen Buddhism*. New York, NY: Grove Press.
- Taylor, B. J., Graham, J. W., Cumsille, P., & Hansen, W. B. (2000). Modeling prevention program effects on growth in substance use: Analysis of five years of data from the adolescent alcohol prevention trial. *Prevention Science, 1*, 183–197.
- Valente, T. W., Hoffman, B. R., Ritt-Olson, A., Lichman, K., & Johnson, C. A. (2003). Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. *American Journal of Public Health, 93*, 1837–1843.
- Valetine, J. C., Biglan, A., Boruch, R. F., Castro, F. G., Collins, L. M., Flay, B. R., . . . Schinke, S. P. (2011). Replication in prevention science. *Prevention Science, 12*, 103–117.
- Wanless, S. B., Patton, C. L., Rimm-Kaufman, S. E., & Deutsch, N. L. (2013). Setting-level influences on implementation of the Responsive Classroom approach. *Prevention Science, 14*, 40–51.
- West, S. G., & Aiken, L. S. (1997). Towards understanding individual effects in multiple component prevention programs: Design and analysis strategies. In K. Bryant, M. Windle, & S. West (Eds.), *New methodological developments in prevention research: Alcohol and substance abuse research* (pp. 167–209). Washington, DC: American Psychological Association.
- Wills, T. A., & Cleary, S. D. (1999). Peer and adolescent substance use among 6th–9th graders: Latent growth analysis of influence versus selection mechanisms. *Health Psychology, 18*, 453–463.