Published in final edited form as:

Am J Drug Alcohol Abuse. 2011 September ; 37(5): 446–452. doi:10.3109/00952990.2011.596980.

Baseline matters: The importance of covariation for baseline severity in the analysis of clinical trials

Edward V. Nunes, M.D.^{(a),*}, Martina Pavlicova, Ph.D.^(d), Mei-Chen Hu, Ph.D.^(a), Aimee Campbell, Ph.D.^(a), Gloria Miele, Ph.D.^(a), Denise Hien, Ph.D.^{(a),(b)}, and Donald F. Klein, M.D.^{(a),(c)}

^(a)Columbia University Department of Psychiatry, New York State Psychiatric Institute

^(b)City University of New York, Department of Psychology

^(c)New York University, Department of Child and Adolescent Psychiatry

^(d)Columbia University, Department of Biostatistics

INTRODUCTION

Clinical trials testing the efficacy or effectiveness of interventions for behavioral health problems are important to the goal of advancing evidence-based treatment. Such trials are expensive and time consuming to conduct, but all too often the findings are disappointing, failing to show evidence of treatment effects. In one of the early NIH sponsored multisite trials comparing the effectiveness of behavioral and medication treatments for depression (Cognitive Behavioral Therapy, Interpersonal Therapy, imipramine plus medical management and placebo plus medical management), the main outcome paper reported no differences in outcome between treatments (1). Subsequently, Klein and Ross (2) reanalyzed the data using more robust methods of analysis of covariance, covarying for the baseline scores of the outcome variables and testing baseline-by-treatment interactions. A number of different treatment effects emerged, including that imipramine was superior to placebo among patients with greater severity of depression at baseline.

Covariation for baseline may enhance detection of treatment effects in randomized clinical trials in one of two ways. To the extent that the score of the outcome measure at baseline correlates with its score at outcome (often the case in treatment studies), the variance explained by the baseline score is removed from the error variance in the estimate of the difference in outcome between treatments, improving power to detect treatment effects (3). Or the treatment effect may manifest as a baseline-by-treatment interaction, suggesting the treatment is effective in a subgroup defined by baseline severity, as observed in the NIMH depression trial (2). A baseline factor that interacts with treatment is commonly called a moderator.

In addiction treatment research, strong associations between baseline and outcome scores of the main dependent measures are frequently observed. For example, in clinical trials of treatments for cocaine dependence, abstinence at the time of randomization, as measured by a cocaine negative urine toxicology, has been consistently and strongly associated with the outcome of abstinence during the subsequent trial (4,5,6,7). Wide variation in baseline levels

*Corresponding Author: Edward V. Nunes, M.D. New York State Psychiatric Institute, 1051 Riverside Drive, Unit 51, Room 3717, New York, NY 10032, nunesed@pi.cpmc.columbia.edu, phone: 212-543-5581.

Declaration of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

of severity may be particularly likely in community-based effectiveness research and other 'practical clinical trials,' since the goal is to test the usefulness of treatments in real-world clinical settings, and the effort is to minimize exclusionary criteria and recruit a sample that is as broadly representative as possible. Nonetheless, it has been our experience that the role of covariates in statistical models testing treatment effects is often ignored or misunderstood by clinician-researchers and that covariation for baseline severity is often omitted from primary outcome analyses.

Therefore, we review basic principles of covariation for baseline severity in clinical trials and baseline-by-treatment interactions and present an example from the National Institute on Drug Abuse Clinical Trials Network (CTN). We discuss methodological issues and present the argument that covariation for baseline severity of the outcome measure and testing of the baseline-by-treatment interaction should be considered routinely for inclusion in the primary outcome analyses of treatment effectiveness trials.

COVARIATION FOR BASELINE IN TESTING TREATMENT EFFECTS: BASIC PRINCIPLES

Covariation for Baseline with no Baseline-by-Treatment Interaction

Figure 1 displays in graphical terms a generic analysis of covariance for a two-arm clinical trial--a linear model where the outcome score is modeled as function of randomly assigned treatment group and the score on the outcome variable measured at baseline prior to randomization. The linear equation for the model takes the following form:

$$\text{Outcome Score} = \text{Intercept} + C_b \text{Baseline} + C_{Tx} \text{Treatment} + \text{error}$$

This example assumes that the outcome variable is a single summary score that is continuous and that higher scores indicate worse outcome--e.g. days using substances, episodes of unprotected sex, or a scale score, such as a craving scale or a mood scale. For a two-arm trial, the treatment variable is dichotomous (e.g. control condition = 0; active treatment condition = 1). The vertical axis represents the outcome score (with increasing values indicating greater severity), and the horizontal axis represents the score of the outcome measure at baseline prior to randomization (increasingly values represent increasing severity). Each individual patient is represented by a point on this outcome-by-baseline graph (solid circles represent control patients; open circle represent active treatment patients). The solid line displays the regression line representing the association between baseline and outcome scores within the control group. The dashed line represents the analogous regression within the active treatment group.

The coefficient of the baseline term (C_b) represents the slope of the regression lines and the strength of association between baseline and outcome scores. The stronger the association between baseline and outcome (i.e., the greater C_b becomes), the less scatter there is in the observed values around the regression lines. Note the graph depicts very little scatter around the regression lines, indicating a very strong association between baseline and outcome scores.

In graphical terms, the main effect of treatment is the vertical distance between the regression lines, visualized by the vertical arrow. This is represented in the equation by the coefficient of the treatment term, C_{Tx} . The standard error of the difference between treatments (and thus the confidence limits of the difference) is determined by the variation around the regression lines, which is smaller than the variation around the means for each treatment group, in proportion to the strength of the baseline-outcome association. Note that

in the graph, there is virtually no overlap between the scatter of data points around the regression line for the treatment group versus the regression line for the control group (given the strong association depicted), whereas there is considerable overlap between the outcome scores for the two treatment groups, ignoring baseline. The variation attributable to baseline score is removed from the estimate of the treatment effect. Thus, a given effect size can be detected with greater power or smaller sample size in proportion to the strength of the baseline-outcome association.

Baseline-by-Treatment Interaction

The model described above and displayed in Figure 1 assumes that the regression lines for the two treatment groups are parallel—i.e. that there is no interaction between baseline score and treatment. Equivalently, this means a constant treatment effect across all initial severities. However, prior to fitting that model, the baseline-by-treatment interaction should be tested, if that is a reasonable hypothesis given prior understanding of the treatment. The linear model covarying for baseline and including the baseline-by-treatment interaction can be represented as follows:

$$\text{Outcome Score} = \text{Intercept} + C_b \text{Baseline} + C_{Tx} \text{Treatment} + C_{int}(\text{Baseline} \times \text{Treatment}) + \text{error}$$

If the estimated coefficient of the interaction term (C_{int}) is significant, the null hypothesis that the regression lines for the treatment and control groups are parallel is rejected. This indicates that the treatment effect differs according to initial severity. The shape of the interaction depends on the slopes and intercepts of the respective regression lines for each treatment group. Further statistical procedures can be applied to determine the region of baseline scores where treatment differs from control. Figure 2 displays three prototypical baseline-by-treatment interactions that have been observed in behavioral health treatment research. Figure 2, Panel A shows an interaction of the form detected by Klein and Ross (2) in their reanalysis of the NIMH treatment of depression trial. The shape of this interaction indicates the difference in outcome between treatment groups increases with increasing baseline severity. The treatment appears to reduce symptoms particularly among those with more severe symptoms to begin with.

Figure 2, Panel B shows the interaction with opposite shape, namely that treatment produces superior outcome among those with low severity at baseline. Examples of this pattern include the multi-site efficacy trial of injectable naltrexone for alcohol dependence, in which the beneficial effect on drinking outcome was found to be restricted to those patients who were abstinent at the outset of the trial (8) and a multisite trial in the CTN in which contingency management with prize-based incentives was found to improve the likelihood of abstinence from cocaine and other stimulants, more so among patients who were abstinent at the time of randomization (9). Anticonvulsants in treatment of cocaine dependence may show similar trends (6,7,10). When a treatment seems to work specifically among patients who have achieved abstinence at baseline, this suggests a relapse prevention effect.

Figure 2, Panel C, shows an interaction in which treatment is beneficial compared to control at one end of the baseline severity spectrum and harmful (outcome better on control than treatment) at the opposite end of the baseline severity spectrum. While this might seem farfetched, a relationship of this type appears to hold for treatment of alcohol dependence with serotonin reuptake inhibitor (SRI) antidepressant medications; three trials (11,12,13) have shown SRI to produce better drinking outcome than placebo among patients with the late onset type of alcohol dependence (associated with lower overall severity), while among patients with the early onset type (associated with greater overall severity) the trend is

toward worse drinking outcome on SRI than on placebo. Such a pattern would seem to suggest two biologically distinct subtypes of the disorder with different treatment implications.

Example: HIV Risk Behavior Outcome in the Women and Trauma Study (CTN-0015)

The Women and Trauma Study (CTN-0015) of the National Institute on Drug Abuse Clinical Trials Network (CTN) was a randomized trial testing the impact of a cognitive behavioral intervention, Seeking Safety (14), compared to an attention control condition of equal credibility, Women's Health Education (15), on outcomes of PTSD and substance use (16) and HIV risk behavior (17) among 346 women with substance dependence and post-traumatic stress disorder. This example was chosen because the finding of a treatment effect (17) depended upon including baseline and baseline-by-treatment in the model. We illustrate what happens when the baseline terms are not included in the analysis.

Frequency of vaginal or anal sex without the use of a condom in the prior 30 days was assessed with the Risk Behavior Survey (18) at baseline (prior to randomization) and at 3, 6 and 12 months after treatment completion. The distribution of HIV risk behavior at baseline showed wide variation with 54% ($n = 185$) of the 343 randomized patients reporting zero unprotected sexual occasions, 26% of patients ($n = 90$) reporting 1 to 14 unprotected sexual occasions, 14% of patients ($n = 47$) reporting 15 to 29 unprotected sexual occasions, and 6% of patients ($n = 21$) reporting 30 or more unprotected sexual occasions.

A zero-inflated negative binomial regression was used to model HIV risk behavior across the 3, 6 and 12 month follow-up points, as a function of time (follow-up occasion), treatment assignment, and baseline HIV risk behavior. This is a type of linear model that is most appropriate for the distribution of the outcome data observed here, which has an excess of zeros and relatively uniform distribution of the non-zero values (19). As with other types of linear model, it ultimately yields regression coefficients estimating the effect of each of the terms in the model. Interactions with time were not significant and were dropped from the model, but the final model revealed a significant baseline-by-treatment interaction (17).

To illustrate the effects of covariation for baseline, Table 1 presents the estimated adjusted odds ratios, derived from the respective regression coefficients, first for the effect of treatment when the baseline and the baseline-by-treatment interaction terms are omitted from the model (Model A), then when the baseline term is added to the model (Model B), and when both baseline and the baseline-by-treatment are included (Model C). Odds ratios greater than 1 indicate that Seeking Safety reduces the probability of high risk sex compared to control.

As can be seen in Table 1, in Model A without covariation for baseline, the odds ratio for the treatment term (1.23) is not significant. In Model B, which includes baseline unprotected sexual occasions as a covariate, the odds ratio for the baseline term is highly significant, indicating the strong positive association between unprotected sex at baseline and unprotected sex at outcome, while the odds ratio for the treatment term tends to be greater (1.41) with confidence limits that approach significance ($p < .10$). This is analogous to the situation illustrated in Figure 1: baseline and outcome are associated, and detection of the treatment effect tends to be enhanced. In Model C, which includes both the terms for baseline unprotected sex and the baseline-by-treatment interaction, the interaction is significant, indicating the slopes relating baseline to outcome differ between the treatment groups and hence that the effect of treatment differs according to baseline severity. The pattern of the interaction resembles that illustrated in Figure 2, Panel A--that Seeking Safety was superior to control at reducing HIV risk behavior among those patients with higher levels of HIV risk behavior at baseline.

DISCUSSION

The example of the Women and Trauma Study illustrates wide variation in the distribution of baseline, pre-randomization severity of the outcome variable, and the strong association between baseline and outcome scores, features typical in clinical trials testing treatments for the addictions and other behavioral health targets. The example also illustrates the frequently occurring variation in treatment effect across levels of baseline severity, which needs detection.

The ubiquity of these effects of baseline severity reinforces the recommendation that baseline severity be included as a covariate in primary outcome analyses, especially if it is known from prior data that baseline and outcome are likely to be strongly associated (20). Further, treatment effects may not be constant across initial levels of severity. Thus, inclusion of the baseline-by-treatment interaction in the model may also enhance the detection of treatment effects, as well as help to specify those patients more likely or unlikely to benefit. This is important for shaping treatment guidelines as well as future treatment development efforts.

A baseline-by-treatment interaction may also offer clues regarding the mechanism of action of the treatment or of the targeted disorder (6,7). For example, the finding that injectable naltrexone reduced drinking among alcoholics who entered the trial having already achieved abstinence (8) may indicate clinically meaningful subtypes of the disorder. The lower severity (abstinent) subgroup may have stronger motivation or represent a biologically distinct form of the disorder where the endogenous opioid system mediates alcohol relapse.

It is commonly thought that the purpose of covariation for baseline factors is to adjust for differences between treatment groups at baseline in order to correct for imbalances in the randomization. However, as pointed out in the CONSORT statement (21, 22, 23, 24), possible sample bias in a randomized clinical trial is addressed by randomization. CONSORT therefore does not even recommend reporting statistical tests on differences between treatment groups at baseline. Covariation for a baseline factor in a randomized clinical trial is useful when the baseline factor is expected to correlate with outcome, not to the extent that the factor appears to significantly differ between treatment groups at baseline (20, 25).

An alternative approach to covariation would be exclusion of patients not in the range of severity thought to be responsive to the intervention being tested. This approach may be considered when prior evidence clearly suggests lack of effectiveness at a particular level of severity. However, stringent inclusion/exclusion criteria are generally discouraged in the design of effectiveness trials, because the resultant sample will be less representative, reducing external validity. In addition, a relationship between baseline severity and treatment effect, observed previously, may not replicate, and restriction of the range of severity of the sample eliminates the opportunity to observe this.

The role of covariation for baseline features and the analysis of subgroup effects has been debated in a variety of clinical fields, particularly the concern that subgroup analyses can be misleading (26,27,28). When a subgroup is derived from an inherently continuous or ordered variable (such as severity) by using an arbitrary cutoff point, this may either capitalize on chance (with respect to detecting a treatment effect within a subgroup) or miss a subgroup effect. Testing a number of subgroups or testing many baseline covariates for interactions with treatment can obviously capitalize on chance, and such analyses must be considered exploratory (20,29). However, a covariate such as baseline severity can be specified a priori as part the primary outcome analysis (20,25).

Recommendations

In a randomized clinical trial, in which baseline severity is expected to be associated with outcome, it is recommended that the plan for the primary outcome analysis consider covariation for baseline and testing of the baseline-by-treatment interaction (20). For studies in which covariation for baseline severity was not part of the primary outcome analysis, re-analysis with covariation for baseline severity should be considered on an exploratory basis. Adding the interaction does incur further statistical issues. Adding baseline terms to the model costs some degrees of freedom each, which is of little consequence in an effectiveness trial with a large sample size. However, determination of power and sample size for a model that includes an interaction is less straightforward, and adding the baseline-by-treatment interaction term will result in a larger sample size being needed. The benefit is the potential to detect a treatment effect, dependent on baseline severity that would be missed otherwise. Yet planned primary outcome analyses in the field rarely include the baseline-by-treatment interaction. Given the expense, time, and effort needed to mount large-scale multi-site effectiveness trials, it seems potentially disappointing and wasteful to limit firm conclusions from a trial to main effects and relegate baseline-by-treatment interactions to exploratory status. Guidelines should be developed for the field regarding the inclusion of the interaction of baseline severity by treatment in the primary outcome analyses of effectiveness trials.

Acknowledgments

Supported in part by the National Institute on Drug Abuse (NIDA) Clinical Trials Network grant U10 DA13035 (Dr. Nunes) and National Institute on Drug Abuse grant K24 DA022412 (Dr. Nunes).

References

1. Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB. National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatment. *Arch Gen Psychiatry*. 1989; 46:971–982. [PubMed: 2684085]
2. Klein DF, Ross DC. Reanalysis of the National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. *Neuropsychopharmacology*. 1993; 8(3):241–251. [PubMed: 8507351]
3. Fleiss, JL. Design and Analysis of Clinical Experiments. New York: John Wiley and Sons, Inc; 1986. Appendix A: Sample-size determination; p. 369–417.
4. Alterman AI, Kampman K, Boardman CR, Cacciola JS, Rutherford MJ, McKay JR, Maany I. A cocaine-positive baseline urine predicts outpatient treatment attrition and failure to attain initial abstinence. *Drug Alcohol Depend*. 1997; 46:79–85. [PubMed: 9246555]
5. Kampman KM, Volpicelli JR, Mulvaney F, Rukstalis M, Alterman AI, Pettinati H, Weinrieb RM, O'Brien CP. Cocaine withdrawal severity and urine toxicology results from treatment entry predict outcome in medication trials for cocaine dependence. *Addict Behav*. 2002; 27(2):251–260. [PubMed: 11817766]
6. Bisaga A, Aharonovich E, Garawi F, Levin FF, Rubin E, Raby WN, Vosburg SK, Nunes EV. Utility of lead-in period in cocaine dependence pharmacotherapy trials. *Drug Alcohol Depend*. 2005; 77(1):7–11. [PubMed: 15607836]
7. Bisaga A, Aharonovich E, Garawi F, Levin FR, Rubin E, Raby WN, Nunes EV. A randomized placebo-controlled trial of gabapentin for cocaine dependence. *Drug Alcohol Depend*. 2006; 81(3): 267–274. [PubMed: 16169160]
8. Garbutt JC, Kranzler HT, O'Malley SS, Gastfriend DR, Pettinati HM, Silverman BL, Loewy JW, Ehrich EW. Vivitrex Study Group. Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. *JAMA*. 2005; 293:1617–1625. [PubMed: 15811981]

9. Stitzer ML, Petry N, Peirce J, Kirby K, Killeen T, Roll J, Hamilton J, Stabile PQ, Sterling R, Brown C, Kolodner K, Li R. Effectiveness of abstinence-based incentives: interaction with intake stimulant test results. *J Consult Clin Psychol*. 2007; 75:805–811. [PubMed: 17907862]
10. Kampman KM, Pettinati H, Lynch KG, Dackis C, Sparkman T, Weigley C, O'Brien CP. A pilot trial of topiramate for the treatment of cocaine dependence. *Drug Alcohol Depend*. 2004; 75(3): 233–240. [PubMed: 15283944]
11. Kranzler HR, Burleson JA, Brown J, Babor TF. Fluoxetine treatment seems to reduce the beneficial effects of cognitive-behavioral therapy in type B alcoholics. *Alcohol Clin Exp Res*. 1996; 20:1534–1541. [PubMed: 8986200]
12. Pettinati HW, Volpicelli JR, Kranzler HR, Luck G, Rukstalis MR, Cnaan A. Sertraline treatment for alcohol dependence: interactive effects of medication and alcohol subtype. *Alcohol Clin Exp Res*. 2000; 24:1041–1049. [PubMed: 10924008]
13. Brady KT, Sonne S, Anton RF, Randall CL, Back SE, Simpson K. Sertraline in the treatment of co-occurring alcohol dependence and posttraumatic stress disorder. *Alcohol Clin Exp Res*. 2005; 29:395–401. [PubMed: 15770115]
14. Najavits, LM. Seeking safety: A treatment manual for PTSD and substance abuse. New York: Guilford; 2002.
15. Miller, S.; Pagan, D.; Tross, S. Peer Activism for Female Partners of Injection Drug Users. Unpublished treatment manual. Columbia University; 1998. Women's health education.
16. Hien DA, Wells EA, Jiang H, Suarez-Morales L, Campbell AN, Cohen LR, Miele GM, Killeen T, Brigham GS, Zhang Y, Hansen C, Hodgkins C, Hatch-Maillette M, Brown C, Kulaga A, Kristman-Valente A, Chu M, Sage R, Robinson JA, Liu D, Nunes EV. Multi-site randomized trial of behavioral interventions for women with co-occurring PTSD and substance use disorders. *J Consult Clin Psychol*. 2009; 77:607–619. [PubMed: 19634955]
17. Hien DA, Campbell AN, Killeen T, Hu MC, Hansen C, Jiang H, Hatch-Maillette M, Miele GM, Cohen LR, Gan W, Resko SM, Dibono M, Wells EA, Nunes EV. The impact of trauma-focused group therapy upon HIV sexual risk behaviors in the NIDA Clinical Trials Network "Women and Trauma" Multi-site Study. *AIDS Behav*. 2010; 14(2):421–430. [PubMed: 19452271]
18. Booth RE, Watters JK, Chitwood DD. HIV risk-related sex behaviors among injection drug users, crack smokers, and injection drug users who smoke crack. *Am J Public Health*. 1993; 83:1144–1148. [PubMed: 8342724]
19. Lambert D. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*. 1992; 34:1–14.
20. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med*. 2002; 21:2917–2930. [PubMed: 12325108]
21. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schultz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996; 276:637–639. [PubMed: 8773637]
22. Moher D, Schulz KF, Altman D. CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001; 285:1987–1991. [PubMed: 11308435]
23. Schulz KF, Altman DG, Moher D. for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010; 340:c332. [PubMed: 20332509]
24. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010; 340:c869. [PubMed: 20332511]
25. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other mis(uses) of baseline data in clinical trials. *Lancet*. 2000; 355(9209):1064–1069. [PubMed: 10744093]
26. Bhandari M, Devereaux PJ, Li P, Mah D, Lim K, Schünemann HJ, Tornetta P 3rd. Misuse of baseline comparison tests and subgroup analyses in surgical trials. *Clin Orthop Relat Res*. 2006; 447:247–251. [PubMed: 16672904]

27. Hernández AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery*. 2005; 57(6):1244–1253. [PubMed: 16331173]
28. Hernández AV, Boersma E, Murray GD, Habbema JD, Steyerberg EW. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J*. 2006; 151(2): 257–264. [PubMed: 16442886]
29. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991; 266:93–98. [PubMed: 2046134]

Figure 1.

Graph of a linear model for the analysis of a generic two-arm, randomized clinical trial, modeling outcome score as a function of treatment assignment and baseline severity of the outcome measure as covariate: $\text{outcome} = \text{intercept} + C_b \text{ baseline} + C_{Tx} \text{ treatment} + \text{error}$. Vertical axis is the outcome score; horizontal axis is the baseline score. Each individual patient is represented by a point on the baseline-by-outcome graph (filled circles for patients in the control group; open circles for patients in the active treatment group). Regression lines relating baseline to outcome scores for each treatment group are shown (solid line: control group; dashed line: active treatment group) with slopes estimated by the coefficient of the baseline term (C_b). The main effect of treatment, estimated by the coefficient of the treatment term (C_{Tx}), is the vertical distance (represented by the vertical arrow) between the regression lines for each treatment group. The standard error of the difference between treatments (and thus the confidence limits of the difference or treatment effect) is determined by the variation around the regression lines, which is smaller than the variation around the means for each treatment group, in proportion to the strength of the baseline-outcome association. Hence, variance in outcome explained by baseline score is removed from the estimate of the treatment effect.

Graphical Pattern*	Description	Examples
Panel A: High severity treatment effect		
	<p>--Treatment more effective (outcome severity reduced more compared to control) among patients with greater severity at baseline.</p> <p>--If the mechanism of the treatment is to reduce symptom severity (or induce remission or abstinence), then this relationship may be observed. Patients with greater severity at baseline have more room to improve or may be more in need of specific treatment and less likely to respond to control treatment.</p>	<p>--NIMH Collaborative Treatment of Depression Trial (1,2). Superiority of antidepressant medication or psychotherapy over placebo when baseline severity greater.</p> <p>--HIV risk behavior analysis (above); see Figure 1, Model C; cognitive behavioral treatment (Seeking Safety) reduced HIV risk behavior more than the control condition among patients with higher levels of HIV risk behavior at baseline.</p>
Panel B: Low severity treatment effect		
	<p>--Patients with low severity at baseline get worse (relapse) in the control condition but remain well in the treatment condition. Whereas among patients with higher severity at baseline, both treatment and control conditions are ineffective, and symptom level remains high at outcome in both treatments.</p> <p>--When a treatment works specifically among patients who are abstinent from the target health behavior at baseline, this can be viewed as a relapse prevention effect.</p>	<p>--Long-acting injectable naltrexone for alcohol dependence more effective compared to placebo among patients who were already abstinent at baseline (8).</p> <p>--Contingency management for stimulant dependence more effective among patients abstinent at baseline (9)</p>
Panel C: Opposite treatment effects at low and high baseline severity levels		
	<p>--Treatment condition yields better outcome than the control condition when baseline severity is low.</p> <p>--Opposite pattern, treatment worse than control (toxic), when baseline severity is high.</p> <p>--The converse pattern, treatment superior at high baseline severity and inferior or toxic at low severity, is also possible.</p>	<p>--Specific serotonin re-uptake inhibitor antidepressants (SSRI) for treatment of alcohol dependence: three trials (11,12,13) show a similar trend that drinking outcome is worse on SSRI compared to placebo among patients with early onset-higher severity subtype of alcohol dependence, while drinking outcome is better on SSRI compared to placebo among patients with later onset-lower severity subtype of alcohol dependence.</p>

*Graphs show the association between baseline severity (horizontal axis) and outcome severity (vertical axis) of the target behavior or symptoms. One regression line shows the baseline-outcome association for the control treatment (solid line), the other for active treatment (dashed line).

Figure 2.

Three patterns of interaction between severity of a behavioral health target measured at baseline prior to randomization (horizontal axis in the graphs) and outcome severity of the target behavior after treatment (vertical axis) in prototype randomized clinical trials of a test treatment (dashed line in the graphs) compared to control treatment (solid line)

*Graphs show the association between baseline severity (horizontal axis) and outcome severity (vertical axis) of the target behavior or symptoms. One regression line shows the baseline-outcome association for the control treatment (solid line), the other for active treatment (dashed line).

Effects in a randomized clinical trial (17) of behavioral treatment--Seeking Safety (SS) versus Women's Health Education control condition (WHE)--on unprotected sexual occasions (USO) during post-treatment follow-up according to three regression models^a: Model A includes the treatment term alone without covariation for USO at baseline; Model B includes treatment and the baseline level of USO as covariate; Model C includes treatment, baseline, and the baseline-by-treatment interaction. Model A shows no significant treatment effect in the absence of covariation for baseline. Model B shows the strong association between baseline USO and outcome and a trend toward a main effect of treatment. Model C shows the significant baseline-by-treatment interaction.

Table 1

Covariates	Model A		Model B		Model C	
	AOR	95% CI	AOR	95% CI	AOR	95% CI
Treatment (WHE vs. SS)	1.23	0.87–1.75	1.41 ⁺	0.95–2.09	0.96	0.55–1.67
USO (baseline)			1.29 ^{***}	1.12–1.50	1.14	0.92–1.41
USO (baseline) × Treatment (WHE vs. SS)					1.34 [*]	1.00–1.79

Notes: USO = unprotected vaginal or anal sexual occasions with a male partner, AOR = adjusted odds ratio, CI = confidence interval;

⁺ p < .10.
^{*} p < .05.
^{***} p < .001.

^a Models are zero-inflated negative binomial models; age, racial/ethnic group, educational level, marital status, site, and measurement occasion (3, 6, and 12 months post treatment) were controlled in all models. Effects shown are the adjusted odds ratios (AORs) derived from the regression coefficients; odds ratios greater than 1 indicate increased probability of unprotected sexual occasions (USO) associated with the control condition, Women's Health Education (WHE), compared to SS.