Longitudinal Analysis when the Experimenter does not Determine when Treatment Ends: What is Dose–Response?

Daniel J. Feaster[1,*], **Frederick L. Newman**[2], and **Christopher Rice**[2]

1 *Center for Family Studies, Department of Psychiatry and Behavioral Sciences, University of Miami School of Medicine, FL, USA*

2 *College of Health & Urban Affairs, Florida International University, USA*

Abstract

The most appropriate amount of psychotherapy to address a particular problem is of interest to clinicians, consumers and those responsible for funding of care. The dose–response relationship has been examined within the context of randomized clinical trials, meta-analysis as well as naturalistic studies; however, each of these approaches has limits. Many of these approaches have conceptually blurred two distinct concepts: do participants with different characteristics need different amounts of therapy and do otherwise equal participants show different outcomes when given different levels of (a particular type of) therapy? For any study design, if the experimenter does not determine the duration of therapy, then the length of therapy is said to be endogenous. This endogeneity can bias any attempt to untangle the answer to these two questions. An extension of the biasing effect of this endogeneity involves the choice of times to assess outcome; if outcome assessment depends on when therapy is terminated (rather than exogenously chosen) then estimates of the trajectory of outcome may be biased. Design considerations to minimize this effect are discussed.

INTRODUCTION

Traditional treatment research is designed to estimate the efficacy of a clinical intervention by comparing the outcomes of two or more interventions with a control condition (either a placebo, waiting list, or a treatment as usual control) with the dosage of each condition held constant. We now have a large number of studies and meta-analyses using this design (e.g. Baucom, Mueser, Shoham, Daiuto, & Stickle, 1998; McCrady, 2000; McDermut, Miller, & Brown, 2001; Miller & Wilbourne, 2002; Moyer, Finney, Swearingen, & Vergun 2002; Prendergast, Podus, & Chang, 2000; Seligman, 1995; Shadish et al., 1993, 1997; Stanton & Shadish, 1997; Wilson, 2000) that inform policy makers on whether psychotherapy of a particular dose is efficacious (e.g. Barkham et al., 1996; Howard, Kopta, Krause, & Orlinsky, 1986; Kopta, Howard, Lowry, & Beutler, 1994; Maling, Gurtman, & Howard, 1995). However, there is considerably less information about how much psychotherapy is sufficient (Baucom et al., 1998), and even less agreement on this in managed care. Consider two examples of what might happen if many health care insurers' standards for the provision of mental health services were applied to the treatment of physical problems.

> Ms Jones—Your child's head and shoulders are now out of the womb. Unfortunately, you have used up your 5h of delivery time for this week. Please come back next week.

---

* Correspondence to: Dr D. J. Feaster, Center for Family Studies, Department of Psychiatry and Behavioral Studies, University of Miami School of Medicine, 1425 N. W. 10th Ave. #307E, Miami, FL 33136, USA. Tel: +305 243 4592. Fax: +305 243 5577. E-mail:dfeaster@miami.edu.

> Mr Smith—90% of the tumour is now benign. Only 10% remains malignant. But you have used up your 20 radiation treatment sessions for this year. Please come back next year for further treatment (cf. Newman & Tejeda, 1996).

In this paper, we point out the lack of randomized research on duration of treatment and some of the problems with the analytic approaches to the dose–response relationship. Similarly, we note that naturalistic studies have great potential to yield knowledge about the dose–response nature of treatment and point out that the choice or determination of assessment times is important to ensure statistical validity in naturalistic studies of psychotherapy. We discuss two possible alternative approaches to this design issue, including the strengths and limitations of these approaches on inferences regarding cause and effects.

CLINICAL TRIAL APPROACHES

In traditional medical research designs, competing treatments (or a treatment and a control condition) are compared using a randomized clinical trial. In mental health clinical trials, frequently competing treatments are compared, holding dosage constant across the two treatments. Randomization of participants to condition, controls for unobserved characteristics of participants by balancing these across conditions. The success of the experimental treatment relative to the control treatment is determined by performing a statistical test for between-group and repeated measures differences among behavioural trajectories over time (Morris & DeShon, 2002).

The Phase I clinical trial (Friedman, Furberg, & DeMets, 1998), normally a relatively small safety and dose-finding trial to find the maximum sustainable dose, is a mainstay of drug development, but less used in psychotherapy. This type of trial frequently does not have a control group, but randomizes participants to varying doses of treatment. For behavioural therapies, a slightly different nomenclature has been proposed. The Stage I study (Onken, Blaine, & Battjes, 1997; Rounsaville, Carroll, & Onken, 2001) involves model and manual development and clinical experience is frequently the guide in determining treatment length for a larger efficacy or Stage II study. To achieve definitive evidence on the dose–response of therapy, we would need to extend a larger Stage II design to include randomization to varying doses of psychotherapy (and perhaps varying doses of the 'control' therapy). This type of design is rare but not untried (for a good example, see Shapiro et al. (1994) and Barkham et al. (1996)). Note that this type of design would, in some cases, subject participants to more therapy than required to meet the therapist/participant goals for therapy, which is not the way therapy is generally conducted.

META-ANALYTIC APPROACH TO DOSE–RESPONSE

The preponderance of fixed duration clinical trials has resulted in attempts to identify the relationship between dose and outcome by comparing results across trials using meta-analysis (Steenbarger, 1994). Results of these trials have tended to show only weak (Smith, Glass, & Miller, 1980) or no (Miller & Berman, 1983; Robinson, Berman, & Neimeyer, 1990; Shapiro & Shapiro, 1982) evidence for a dose–response relationship. The variability in the types of therapies compared and, indeed the variability in populations studied may explain the lack of a dose–response relationship in meta-analytical studies (Steenbarger, 1994). It appears that the meta-analytic approach, as applied, has mixed two conceptually different effects. The first is how different types of people may need different amounts of therapy. Thus clients with higher severity do show more results over longer periods of time, because they had more room for improvement and needed more therapy. The second is more truly the dose–response relationship: do otherwise equal clients show different results when given different levels of a particular type of therapy?

Another suggested reason for this lack of findings has been the differing time frames of the component studies and differing timing in assessment of outcomes (Salzer, Bickman, & Lambert, 1999; Steenbarger, 1994). Most studies in the meta-analyses were pre–post designs with wide-variability in the time between pre and post (which actually is used to identify the dose–response relationship). The lack of follow-up data was particularly problematic for assessing lasting success of therapy. This combined with the variability of therapy and heterogeneity of samples contributes to the interpretation that the dosage–outcome link might be stronger within a treatment rather than across (Steenbarger, 1994).

NATURALISTIC REPEATED MEASURES CLINICAL STUDIES

Another popular approach in psychotherapy is the single group (naturalistic) repeated measures design, which is used to investigate the change in the targeted characteristics from pre-treatment through the follow-up period as a trajectory over time. With repeated measures studies it is also possible to examine the within-subjects similarities or differences among the trajectories. The truly naturalistic study design allows participants and therapists to determine the length of therapy. This decision might include consideration of multiple factors, e.g. costs, achievement of behavioural targets, demoralization, futility, other contextual factors, and so forth.

Howard and his colleagues have strongly argued for performing naturalistic studies on the trajectory of change for individuals in treatment, and from these data, identify the type and amount of therapeutic effort that was needed to achieve a clinically significant level of behavioural change (Howard, Moras, Brill, Martinovich, & Lutz, 1996; Newman, Saunders, & Feaster, 2003). These types of investigations have shown that, in fact, there appears to be a negatively accelerating relationship between dose and outcome.

This approach has evolved into a method of benchmarking an individual's progress in therapy relative to his peers in therapy (Lambert, Hansen, & Finch, 2001; Leon, Kopta, Howard, & Lutz, 1999; Lueger et al., 2001; Lutz, Martinovich, & Howard, 1999). Here the emphasis of analysis is no longer 'is more therapy better', but is rather 'is the amount of therapy that you have had doing as much as it should be?' This individualized approach to research has potential to really bolster clinical care.

A POTENTIAL PROBLEM RELATED TO TIMES OF MEASUREMENT

The variability in times of assessment has been suggested as one of the problems associated with the application of meta-analysis to the question of a dose–response relationship (Salzer et al., 1999; Steenbarger, 1994). Many early studies had just pre–post assessments and the post assessment was conducted immediately after the participant finished therapy, which might have varied by participant. This clearly does not include information on the longer-term well-being of the client, because of the lack of follow-up. It also may have been overstating the effect of treatment, if termination was determined by reference to the behavioural (or outcome) status of the client, as good clinical practice would dictate. In this section we explain why randomized clinical trials generally avoid this problem and why naturalistic studies should pay more attention to the problem.

In a well-designed randomized clinical trial, treatment assignment, assessment, termination, and dosage are *exogenous variables* (i.e. treatment arm, times of assessment, dose, and termination are controlled by the experimenter). This feature is a necessary element if statistical tests of between-group differences in the behavioural changes over time are to have validity in addressing the question: which treatment of comparable dosage produces the best outcome or outcome trajectory over time? Which dosage of a particular therapy is likely to achieve results for what type of client? Likewise, for a single-group repeated measures design, there

can be a valid statistical test of outcome testing the question: do the subjects within the group exhibit statistically significant changes over time in a desired direction or meet desired criteria? Do patient characteristics, therapist characteristics, their interaction, or events within therapy (aspects of the therapeutic process) relate to the trajectory of the outcome and the time course of therapy?

In randomized clinical trial designs, we can achieve valid statistical inference regarding outcome or the trajectory of outcome over time, in part, because the times of measurement and treatment termination are exogenous. These are exogenous because they are determined by the researcher, and are uninfluenced by the therapist's or the patient's behaviours. Under these conditions, fixed treatment and random error effects should be independent of each other. Because of this independence it is possible to construct a valid statistical test for fixed (treatment) effects when contrasting behavioural trajectories of individuals over time, both within and between groups.

Whereas fixing the type and amount of treatment exogenously is desirable for measuring behavioural changes over time, the resulting data may not give sufficient information about how much is needed and for whom it is needed. The two examples of applying the logic of fixed dosage to the delivery of a child or to the treatment of a cancerous tumour make abundantly clear what is already understood by most practicing clinicians—dosing decisions are not nor should they always be fixed. But what are the research alternatives? How can we appropriately supplement the information found from controlled clinical trials?

Suppose a researcher is interested in testing the effectiveness of a treatment in an actual (naturalistic) mental health service setting. In naturalistic settings treatment dosage or the reason for termination are not under the control of the researcher and that is problematic when attempting to find a valid test statistic regarding between-or within-group differences in the trajectory of behavioural change over time. In such naturalistic settings treatment dosage is variable and termination occurs for any of at least seven different reasons, none of which is under the control of the experimenter: (1) the patient or the patient's guardian decides to terminate treatment; (2) a third party payer makes the decision to terminate; (3) the therapist decides that a goal has been achieved and it is reasonable to terminate; (4) the therapist decides that further treatment will not benefit the patient; (5) the patient and therapist jointly decide that a goal has been achieved; (6) the patient and therapist jointly decide that further treatment will not benefit the patient; or (7) the patient terminates without regard to any of the above (e.g. 'against medical advice'). In each of these situations, termination is an *endogenous* variable, i.e. the point at which termination of treatment occurs is influenced by those directly or indirectly involved in the therapy.

At issue here is what happens to the analysis of change (e.g. growth curve analysis) when termination is endogenous; that is, when termination is not determined by the researcher. If the outcome variable is measured at the termination point, then a growth curve that includes this observation time will be biased. The derivation that is given in Appendix A clearly indicates that one cannot validly test for the goodness of fit of a growth curve nor can one validly contrast growth curves to detect fixed within-group or between-group differences if this endogenous measurement time is included in the growth curve data.

The problem can be stated slightly more intuitively than the derivation. Let us assume that the client's well-being is the factor behind both the clinical behavioural criteria (determining termination) and our outcome measure. We know that the client's well-being is affected by multiple processes not least of which we hope is the therapy that he or she is undergoing. Despite our hopes of the importance of therapy in the client's well-being, there are multiple other determinates of well-being which because we do not observe them can be thought of (and

modelled) as random. Thus clients will have different 'random' factors or shocks at different times that will either increase or decrease their well-being. These shocks might be short or longer lasting. If the therapist is clinically gauging the client's well-being to determine when it is appropriate to terminate therapy, it is likely that the decision to terminate will be made when a positive random shock has raised the client's well-being. Note that this is not to say that there is not real progress. The true direction of change for well-being may be up, but there are also more transient factors that will also affect the direction and magnitude of change over any finite period. The point is, termination is likely to occur when the client is doing better (has higher well-being) than is his or her long-run course of well-being. Since our outcome measure is also measuring well-being, it will reflect this fact as well, and our measurement of pre–post change is probably overstated. If the time of measurement is predetermined and not affected by the clinical course of the client, this type of bias is minimized.

The logic of this statement also holds for any interim measurement strategy that in any way depends on the participant, therapist, or therapy. Thus, the potential problem of endogenous measurement times is not limited to termination assessments, but also holds for assessments at any time during or after therapy. Thus, in order to meet the requirements of the analysis, for both naturalistic designs and controlled clinical trials, the times of measurement should be determined prior to the initiation of therapy. Further, procedures should be established that ensure that the participant's functioning (or therapist's judgment) does not affect the probability of assessment. Clearly, if you only assess a depressed participant when he or she 'feels like' being assessed, you are likely to understate the severity of their trajectory of depression.

ARE THERE ALTERNATIVE APPROACHES?

Does endogeniety paint us into a corner such that we cannot perform valid naturalistic studies? We do not feel that this is the case. There are reasonable alternatives. The most pervasive reason for pursuing such alternatives is that current research findings based upon traditional research designs are not sufficiently pervasive to provide a sound scientific basis for setting policies about the delivery of mental health services (Kazdin, 2001). We are not arguing that performing naturalistic longitudinal studies that map and contrast the trajectories of change among patients will provide all solutions, but we are arguing that significant information exists in the collective experience of clinicians and that developing a scientific base for presenting this information could aid in policy development. Likewise, accurate patient profiling and potential utilization of feedback to clinical practice have the potential to really improve evidence-based practice. It is in this sense that we now offer a discussion of some likely alternative approaches to the analysis of longitudinal data collected in naturalistic studies. Specifically, we explore two alternative approaches that could be used to address research questions about the nature of change in behaviours over time, given that treatment termination is endogenous.

The first and perhaps simplest solution to this endogeniety problem is to set fixed points in time that are independent of when treatment terminates (and the therapeutic process) to assess behaviours and to assess the behavioural trajectories within and/or between groups. The intervals between behavioural assessments do not necessarily have to be of equal interval, just so long as the time when an assessment is carried out is independent of the patient's or the therapist's behaviour and preferably specified prior to the initiation of therapy. This means that outcome should not be assessed at the point of termination, because this is dependent on the therapeutic process. If outcome is assessed at termination, the outcome at this time should not be included in the growth curve estimate of the trajectory of change. This may seem counter-intuitive (that more information is not better). However, remember that information that is obtained when the patient or therapist have decided to terminate therapy is not a random sampling of the trajectory of the outcome. It is likely that participants are doing better (on

average) at this point, so inclusion of this data would bias the slope or measurement of true change.

Of course the researcher should select intervals that at least cover the expected duration of treatment. Moreover, the researcher should select intervals such that the expected changes in behaviours can be detected. For example, one interpretation of the phase model of psychotherapy (Howard, Lueger, Maling, & Martinovich, 1993) would predict very rapid changes in behaviour in the first few sessions (i.e. during the remoralization phase of changes in feelings of well-being), followed by slightly slower changes during the remediation (changes in symptoms) and this would be followed by even slower changes during the rehabilitation phase (where long-term habits are modified). Because there are at least three different rates of change anticipated, one may want to have at least five well-placed assessment points.

A second approach would be to estimate jointly both time to termination and the growth curve describing the outcome variable (Hogan & Laird, 1996; Touloumi, Pocock, Babiker, & Darbyshire, 1999). Structural equation methods (Bentler & Weeks, 1980; Muthén, 1991) or other maximum likelihood approaches (Duncan et al., 1997; Khoo & Muthén, 2000; Schluchter, 1992) would need to be used to estimate such a model. This joint modelling strategy can allow time to termination to affect the growth curve and the growth curve to affect time to termination. This approximates and describes the decision making involved by the relevant parties. If we further model the decision as specific to one of the involved parties (patient, therapist or third party payer), a separate equation could be added for the behaviours of each of the parties. In fact, the modelling of reasons for treatment withdrawal can be an important research question in its own right.

Lambert et al. (2001) use a derivative of this approach to examine the effect of feedback to therapist on treatment effects. In their Study 1, rather than examine the growth curve of outcome, they reframe the question to be how long does it take to achieve a clinically significant change? Thus they are conceptually looking at what we are calling time to termination. Their data came from multiple provider organizations. Only in one subpart of their data is the actual length of therapy open-ended; in large part their data comes from patients among whom insurance rules would have given an exogenous termination to therapy. However, because they did not control the data collection process, they cannot be sure that their data collection times might not have been endogenous (i.e. affected by therapy, managed care supervisors, etc.). Their decision to look only at the time to clinically significant change avoids the problem (in Study 1) of which we are advising herein, but also does not give any information on the growth curve of outcome. In their Study 2, they use recovery curves that are based on growth curves, which are then potentially subject to the bias we describe. If they had jointly estimated the growth curves with the model of time to clinically significant improvement, they would have negated any bias introduced by having an endogenous treatment termination (although not any bias of the therapy process affecting the other, nontermination, assessment times). Because their eventual goal (their Study 3)—to show an effect of feedback of participant progress to therapists—was successful, the bias (if it existed) was not so large as to entirely obviate the effect of feedback. In fact, this bias might be much less important for making comparisons between individuals and a group (all of whom are affected by the bias, if it exists).

Endogenous measurement times are not the only threat to statistical inference involved in naturalistic studies. For example, we have focused here on single-group naturalistic studies. If the study being contemplated involves multiple types of therapy (a multiple group) naturalistic study, then there is an additional source of endogeniety—the choice of therapy. In a naturalistic study, participants are not randomized to therapy—they either choose a particular type of therapy, or the therapist decides that a particular type of therapy would be best for them. In this case, assignment to therapy is endogenous. There has been significant attention to this

situation in the policy evaluation field (Heckman, 1997; Heckman & Hotz, 1989), with attention focused on matching individuals in the different types of treatment using, for example, the propensity score (Hahn, 1998; Heckman, Ichimura, & Todd, 1998; Rosenbaum & Rubin, 1984; Rubin & Thomas, 1996). The propensity score uses the participant's observed characteristics to predict what treatment they actual choose or receive. Many solutions to this endogeniety problem such as the propensity score are similar to what we have proposed here because they involve jointly modelling the decision to go into a particular therapy (rather than when to terminate) and the time path of outcome if a particular therapy is chosen.

Randomized clinical trials are not immune from endogeneity problems. Our discussion of these trials has focused on their planned predetermined length or duration of treatment. In fact, most trials do not have all participants completely complying with the planned course of treatment. Thus, even within the randomized clinical trial, actual treatment duration varies and is endogenous. To look at the relationship of treatment duration and outcome within this framework requires modeling of the nature discussed above. Indeed, much recent research has examined how to incorporate compliance data into tests of the randomized clinical trial (Angrist, Imbens, & Rubin, 1996; Imbens, & Rubin, 1997; Jo, 2002; Korhonen, Laird, & Palmgren, 1999; Rochon, 1995) not without some controversy (Pocock & Abdalla, 1997).

SUMMARY AND WHAT IS REALLY MEANT BY DOSE–RESPONSE

We have pointed out that measurement times must be independent (exogenous) from the path of therapy for statistical inference on the trends in outcome to be valid in both randomized clinical trials and in naturalistic studies. The best way to ensure this is to have a predetermined plan for the timing of outcome assessment. This would ensure that what actually is happening in therapy is in no way related to the timing of measurement. A particular implication of this is that in a naturalistic study that does not have a fixed termination time, the outcome assessments made at termination should not be included when calculating individuals' growth curves. If it is desired to include the outcome assessment at termination, then time to termination should be an explicit part of the statistical model and a structural (or simultaneous) equation model estimated.

Naturalistic studies of particular types of psychotherapy have the potential to supplement the efficacy data obtained from randomized clinical trials. Because of the fixed dosage involved in most randomized clinical trials, there is little information from randomized trials that is available about the minimally sufficient or optimal dose. Naturalistic studies of actual therapy practice have and should continue to add information important to making dosing decisions.

We would like to assert that dosing decisions are not exactly the same as a dose–response relationship. Much of the research on dose–response relationships in psychotherapy research has mixed two conceptually distinct questions. First, what is the appropriate dose for someone with a particular history, set of presenting problems (including severity) and contextual factors? This is really trying to establish optimal clinical assessment, diagnosis and targeting, not a dose–response relationship. The second is the true dose–response question; given persons that are otherwise equal, will the person with more therapy show significantly better outcomes? These are two very different questions. In general, randomized clinical trials can have an advantage in answering the second question (if they randomly assign to varying lengths of therapy). And at present, the naturalistic, patient-focused, patient-profiling methodologies appear to have an advantage in answering the first question. There is no reason why answers from both types of well-run trials should not inform clinician-researchers.

References

Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American Statistical Association 1996;91:444–455.

Barkham M, Rees A, Stiles WB, Shapiro DA, Hardy GE, Reynolds S. Dose–effect relations in time-unlimited psychotherapy. Journal of Consulting and Clinical Psychology 1996;64:927–935. [PubMed: 8916621]

Baucom DH, Mueser KT, Shoham V, Daiuto AD, Stickle TR. Emprically supported couple and family interventions for marital distress and adult mental health problems. Journal of Consulting and Clinical Psychology 1998;66:53–88. [PubMed: 9489262]

Bentler PM, Weeks DG. Linear structural equations with latent variables. Psychometrika 1980;45:289–308.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Duncan TE, Duncan SC, Alpert A, Hops H, Stoolmiller M, Muthen B. Latent variable modeling of longitudinal and multilevel substance use data. Multivariate Behavioral Research 1997;32:275–318.

Friedman, L.M., Furberg, C.D., & DeMets, D.L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York: Springer-Verlag.

Hahn J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 1998;66:315–331.

Heckman JJ. Instrumental variables: A study of the implicit assumptions underlying one widely used estimator for program evaluations. Journal of Human Resources 1997;32:441–462.

Heckman JJ, Hotz VJ. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. Journal of the American Statistical Association 1989;84:862–874.

Heckman JJ, Ichamura H, Todd P. Matching as an econometric evaluation estimator. Review of Economic Studies 1998;65:261–294.

Hogan JW, Laird NM. Intention-to-treat analyses for incomplete repeated measures data. Biometrics 1996;52:1002–1017. [PubMed: 8805765]

Howard KL, Kopta SM, Krause MS, Orlinsky DE. The dose–effect relationship in psychotherapy. American Psychologist 1986;54:159–164. [PubMed: 3516036]

Howard KI, Lueger RJ, Maling MS, Martinovich Z. A phase model of psychotherapy outcome: Causal mediation of change. Journal of Consulting & Clinical Psychology 1993;61:678–685. [PubMed: 8370864]

Howard KI, Moras K, Brill PL, Maretinovich Z, Lutz W. Evaluation of psychotherapy: efficacy, effectiveness, and patient progress. American Psychologist 1996;51:1059–1064. [PubMed: 8870542]

Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with non-compliance. Annals of Statistics 1997;25:305–327.

Jo B. Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. Statistics in Medicine 2002;21:3161–3181. [PubMed: 12375297]

Kazdin AE. Almost clinically significant ($p < 0.10$): Current measures may only approach clinical significance. Clinical Psychology: Science and Practice 2001;8:455–467.

Khoo, S.T., & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. In J.S.E. & Rose, L.E. Chassin et al. (Eds), *Multivariate applications in substance use research: New methods for new questions* (pp. 113–140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kopta SM, Howard KI, Lowry JL, Beutler LE. Patterns of symptomatic recovery in psychotherapy. Journal of Consulting and Clinical Psychology 1994;62:1009–1016. [PubMed: 7806709]

Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: An application to the ATBC study. Statistics in Medicine 1999;18:2879–2897. [PubMed: 10523748]

Lambert MJ, Hansen NB, Finch AE. Patient-focused research: using patient outcome data to enhance treatment effects. Journal of Consulting and Clinical Psychology 2001;69:159–172. [PubMed: 11393594]

Leon SC, Kopta SM, Howard KI, Lutz W. Predicting patients' responses to psychotherapy: Are some more predictable than others? Journal of Consulting and Clinical Psychology 1999;5:698–704. [PubMed: 10535236]

Lueger RJ, Howard KI, Martinovich Z, Lutz W, Anderson EE, Grissom G. Assessing treatment progress of individual patients using expected treatment response models. Journal of Consulting and Clinical Psychology 2001;69:150–158. [PubMed: 11393593]

Lutz W, Martinovich Z, Howard KI. Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. Journal of Consulting & Clinical Psychology 1999;67:571–577. [PubMed: 10450628]

Maling MS, Gurtman MB, Howard KI. The response of interpersonal problems to varying doses of psychotherapy. Psychotherapy Research 1995;5:63–75.

McCrady BS. Alcohol use disorders and the Division 12 Task Force of the American Psychological Association. Psychology of Addictive Behaviors 2000;14:267–276. [PubMed: 10998952]

McDermut W, Miller IW, Brown RA. The efficacy of group psychotherapy for depression: A meta-analysis and review of the empirical research. Clinical Psychology: Science and Practice 2001;8:98–120.

Miller RC, Berman JS. The efficacy of cognitive abhaviour therapies: a quantitative review of the research evidence. Psychological Bulletin 1983;94:39–53. [PubMed: 6353465]

Miller WR, Wilbourne PL. Mesa grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction 2002;97:265–277. [PubMed: 11964100]

Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods 2002;7:105–125. [PubMed: 11928886]

Moyer A, Finney JW, Swearingen CE, Vergun P. Brief interventions for alcohol problems: A meta-analytic review of controlled investigations in treatment-seeking and non-treatment seeking populations. Addiction 2002;97:279–292. [PubMed: 11964101]

Muthén, B.O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L.M.E. Collins, & J.L.E. Horn (Eds), *Best methods for the analysis of change: Recent advances*, *unanswered questions*, *future directions* (pp. 1–17). Washington, DC: American Psychological Association.

Newman FL, Tejeda M. The need for research to support decisions in the delivery of mental health services. American Psychologist 1996;51:1040–1049. [PubMed: 8975583]

Newman FL, Saunders S, Feaster DJ. Equivalence to normal? Journal of Clinical Psychology 2003;59:735–743. [PubMed: 12808580]

Onken, L.S., Blaine, J.D., & Battejes, R. (1997). Behavioral therapy research: A conceptualization of a process. In S.W. Henngler, & R. Amentos (Eds), *Innovative approaches for difficult to treat populations* (pp. 477–485). Washington, DC: American Psychiatric Press.

Pocock SJ, Abdalla M. The hope and hazards of using compliance data in randomized controlled trials. Statistics in Medicine 1997;17:303–317. [PubMed: 9493256]

Prendergast ML, Podus D, Chang E. Program factors and treatment outcomes in drug dependence treatment: An examination using meta-analysis. Substance Use and Misuse 2000;35:1931–1965. [PubMed: 11138713]

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Robinson L, Berman J, Neimeyer R. Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. Psychological Bulletin 1990;108:30–49. [PubMed: 2200072]

Rochon J. Supplementing the intent-to-treat analysis: accounting for covariates observed postrandomization in clinical trials. Journal of the American Statistical Association 1995;90:292–300.

Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Society 1984;79:516–524.

Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioural therapies research: Getting started and moving from stage I. Clinical Psychology: Science and Practice 2001;8:133–142.

Rubin DB, Thomas N. Matching using estimated propensity scores: Relating theory to practice. Biometrics 1996;52:249–264. [PubMed: 8934595]

Salzer MS, Bickman L, Lambert EL. Dose–effect relationship in children's psychotherapy services. Journal of Consulting and Clinical Psychology 1999;67:228–238. [PubMed: 10224733]

Schluchter MD. Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine 1992;11:1861–1870. [PubMed: 1480878]

Seligman MP. The effectiveness of psychotherapy: The Consumer Reports study. American Psychologist 1995;50:965–974. [PubMed: 8561380]

Shadish WR, Montgomery LM, Wilson P, Wilson MR, Bright I, Okwumabua T. Effects of family and marital psychotherapies: A meta-analysis. Journal of Consulting and Clinical Psychology 1993;61:992–1002. [PubMed: 8113500]

Shadish WR, Matt GE, Navarro AM, Siegle G, Crits-Christoph P, Hazelrigg MD, Jorm AF, Lyons LC, Nietzel MT, Prout HT, Robinson L, Smith ML, Svartberg M, Weiss B. Evidence that therapy works in clinically representative conditions. Journal of Consulting and Clinical Psychology 1997;65:355–365. [PubMed: 9170759]

Shapiro DA, Shapiro D. Meta-analysis of comparative therapy outcome studies: A replication and refinement. Psychological Bulletin 1982;92:581–604. [PubMed: 7156259]

Shapiro DA, Barkham M, Rees A, Hardy GA, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. Journal of Consulting and Clinical Psychology 1994;62:522–534. [PubMed: 8063978]

Smith, M.L., Glass, G.V., & Miller, T.I. (1980). *The benefits of psychotherapy.* Baltimore: John Hopkins University Press.

Stanton MD, Shadish WR. Outcome, attrition, and family-couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. Psychological Bulletin 1997;122:170–191. [PubMed: 9283299]

Steenbarger BN. Duration and outcome in psychotherapy: an integrative review. Professional Psychology: Research and Practice 1994;25:111–119.

Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Estimation and comparison of change in longitudinal studies with informative drop-outs. Statistics in Medicine 1999;18:1215–1233. [PubMed: 10363341]

Wilson DB. Meta-analyses in alcohol and other drug abuse treatment research. Addiction 2000;95(Suppl 3):S419–S438. [PubMed: 11132366]

APPENDIX A: DEMONSTRATING THE PROBLEMS IN DERIVING A VALID STATISTICAL TEST IN NATURALISTIC PRE–POST STUDIES WITH VARYING TIME OF TERMINATION

To demonstrate this we start by defining the model for the individual growth curve (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002):

$$y_{it} = \pi_{0i} + \pi_{1i}\, a_{it} + e_{1it} \tag{1}$$

Where: y_{it} is observation on the *ith* subject at time t, π_{0i} = intercept, π_{1i} = slope, e_{1it} = error, and a_{it} = time since study entry, which need not be equal across subjects. Now, suppose that a_{it} is endogenous, which can be represented by the sum of a component that is decided upon by the client, the therapist, or an external agent, γz_{it}, and a random error, e_{1it}. Thus, a_{it} is not under experimental control such that:

$$a_{it} = \gamma z_{it} + e_{2it} \tag{2}$$

At level 2, as described by Raudenbush and Bryk (2002), individual variation is modelled in Equations 3 and 4, below:

$$\pi_{0i} = \beta_{00} + \sum_{q=1}^{Q_0} \beta_{0q} X_{0qi} + r_{0i} \qquad (3)$$

$$\pi_{1i} = \beta_{10} + \sum_{q=1}^{Q_1} \beta_{1q} X_{1qi} + r_{1i} \qquad (4)$$

Where the baseline level of outcome, π_{0i}, is described as the sum of β_{00}, an intercept term, and the series coefficients β_{0q}, or slope terms, multiplied by individual non-time varying (fixed) characteristics X_{0qi}, and the error term, r_{0i}. Similarly, the rate of change over time, π_{1i}, is described analogously with terms, β_{10}, β_{1q}, and r_{1i}. Substituting Equations 3 and 4 into 1 and assuming only one fixed characteristic (e.g. the type of treatment) in each equation, 3 and 4, the result is given as Equation 5 below:

$$y_{it} = \beta_{00} + \beta_{01} X_{01} + r_{0i} + (\beta_{10} + \beta_{11} X_{11} + r_{1i}) a_{it} + e_{1it} \qquad (5)$$

Completing the multiplication and regrouping terms we create Equation 6:

$$\begin{aligned} y_{it} = \beta_{00} + \beta_{01} X_{01} + \beta_{10} a_{it} + \beta_{11} X_{11} \cdot a_{it} \\ + (r_{0i} + r_{1i} a_{it} + e_{1it}) \end{aligned} \qquad (6)$$

where the portion in parenthesis in Equation 6 is the error structure and with the assumption that r_{0i}, r_{1i}, and e_{1it} are normally distributed error terms that are uncorrelated (and thus independent). This can be estimated using maximum likelihood (or more frequently restricted maximum likelihood). However, if a_{it} is endogenous, Equation 2 must be included:

$$\begin{aligned} y_{it} = \beta_{00} + \beta_{01} X_{01} + r_{0i} \\ + (\beta_{10} + \beta_{11} X_{11} + r_{1i}) \cdot (z_{it} \gamma + e_{2it}) + e_{1it} \end{aligned} \qquad (7)$$

Then completing the multiplication and regrouping of terms we have Equation 8:

$$\begin{aligned} y_{it} = \beta_{00} + \beta_{01} X_{01} + \beta_{10} \gamma z_{it} + \beta_{11} X_{11} \gamma z_{it} \\ + (\beta_{10} e_{2it} + \beta_{11} X_{11} e_{2it} + r_{0i} + r_{1i} \gamma z_{it} + r_{1i} e_{2it} + e_{1it}) \end{aligned} \qquad (8)$$

Note that the error structure (the terms within the parentheses of Equation 7) is now much more complicated. This equation is really no longer a growth curve, because the 'growth curve' depends on time-varying covariates. Instead we now have a system of simultaneous equations. There are several potential ways to estimate this model, although a fully general model is not identified. Typically one would need exclusions between z and X. A possible method of estimation is two- (or three-) stage least squares (2SLS), although maximum likelihood may also be used. Note that the assumption that e_{1it} and e_{2it} are independent is probably not tenable. We would think that the therapist's determination that therapy should end is going to be correlated with outcome measures regardless of whether they know these measures or not.

Looking at Equation 7 is perhaps the quickest way to see the difficulty. Since a_{it} is endogenous, there is an error term associated with it. It is noteworthy that the term within the parentheses greatly resembles a measurement error equation. Not accounting for the endogenous character of the term will cause the parameters on all of the variables in Equation 8 to be biased and inconsistent if the observed a_{it} is used as the regressor without accounting for this error term.

Note that here we have shown the case where all observation times, a_{it}, are endogenous, but the problem exists if any of the observation times are endogenous.