Published in final edited form as:

Behav Ther. 2012 June ; 43(2): 381–392. doi:10.1016/j.beth.2011.08.004.

Predicting Psychotherapy Benefit: A Probabilistic and Individualized Approach

Oliver Lindhiem¹, David J. Kolko¹, and Yu Cheng²

¹University of Pittsburgh School of Medicine, Department of Psychiatry, Pittsburgh, PA

²University of Pittsburgh, Department of Statistics, Pittsburgh, PA

Abstract

We describe the development of the Probability of Treatment Benefit (PTB) chart which incorporates, integrates, and extends more recent approaches to describing treatment effects, such as the Reliable Change Index (Jacobson & Truax, 1991) and normative comparisons (Kendall et al., 1999), by including parameters that are simultaneously *probabilistic* and *individualized*. To illustrate the PTB chart, data are taken from an effectiveness trial ($N = 139$) of a modular treatment for disruptive behavior disorders. The results highlight both individual variability and the probabilistic nature of psychotherapy benefit. Finally, we discuss the utility of the PTB chart in terms of research, policy, and practice implications. Probability tables, such as the PTB chart, have the potential to be used as simple clinical tools to supplement traditional effect sizes and help patients make truly informed decisions about treatment participation.

Keywords

treatment; response; outcome; prediction

Traditional effect sizes, such as Cohen's d , are useful summaries of how well particular treatments work *on average* and have great utility for meta-analytic research (e.g., Lipsey & Wilson, 1993; Smith & Glass, 1977; Weisz, Jensen-Doss, & Hawley, 2006; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995) and systematic reviews (e.g., Eyberg, Nelson, & Boggs, 2008). However, effect sizes do not provide information about the *likelihood* or *probability* that a treatment will provide benefit to a particular patient. An effect size characterizes the average amount of change that was observed in a study, but not the likelihood or probability that a given patient will experience that average benefit. An effect size answers the question, "How *big* is the *average* benefit?" It does not answer the question, "What are the *chances* this treatment will benefit *me*?" As we will illustrate in this paper, these are two equally important but distinct questions. The answers to these two questions provide complementary information, both of which are necessary to make an informed decision about beginning a treatment. We will see, for example, that some patients can have a rather low probability (e.g., 21% chance) of

© 2010 Published by Elsevier Ltd on behalf of Association for Behavioral and Cognitive Therapies.

Corresponding Author: Oliver Lindhiem, Ph.D., University of Pittsburgh School of Medicine, Department of Psychiatry, 3811 O'Hara St., Pittsburgh, PA 15213, Office: 548 Bellefield Towers, Phone: 412-683-7693, Fax: 412-246-5341, lindhiemoj@upmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

experiencing reliable gains from psychotherapy even when the treatment has a “large” effect size by traditional standards.

Even though psychotherapy response and outcome can be modeled by variables such as initial treatment severity (e.g., Kazdin & Whitley, 2006; Reyno & McGrath, 2006), comorbid conditions (Beauchaine, Webster-Stratton, & Reid, 2005; Hartmen, Stage, & Webster-Stratton, 2003; Lavigne, et al., 2008), and life stress (Lavigne, et al., 2008; Reyno & McGrath, 2006), these variables only account for a relatively small proportion of the variability in psychotherapy response and outcome. In any psychotherapy outcome study, the overwhelming proportion of variability in response and outcome remains unaccounted for by any single predictor variable or group of predictor variables (e.g., Reyno & McGrath, 2006). This is true regardless of the treatment’s effect size. As a result, there is always uncertainty concerning the likelihood that a treatment will provide benefit to a given patient, regardless of the treatment’s effect size. The absence of probabilistic information about the expected likelihood or probability of benefiting from psychotherapy makes it difficult for patients to make truly informed decisions before beginning a treatment, even when the treatment has a large mean effect size. Without knowing the likelihood of gains, patients are investing time and financial resources based on only the magnitude of the *average gain* (i.e., effect size). Predicting benefit from psychotherapy, we suggest, is complicated by the fact that individuals (and groups of individuals) do not all have the same likelihood or probability of receiving benefit.

The Probability of Treatment Benefit (PTB) Chart

Overview

In this paper we introduce the Probability of Treatment Benefit (PTB) chart which was developed to quantify the likelihood or probability of benefiting from psychotherapy for a given individual. It is intended as a supplement to traditional effect sizes. Specifically, the PTB chart summarizes the likelihood or probability of benefiting from psychotherapy in percentage points. This provides a straightforward and intuitive interpretation that can easily be understood by patients. We illustrate the PTB chart and its utility with examples from the field of child and adolescent psychotherapy for externalizing disorders.

Prediction must be probabilistic

The PTB chart is based on the premise that prediction must always be probabilistic. For any two variables with a less than perfect correlation (i.e., less than 1.0), the value of one variable can only be used to predict the other with a degree of uncertainty. For example, even though baseline symptom severity is a robust predictor of treatment outcome, a baseline symptom severity score cannot tell us with certainty whether or not a patient will have a favorable outcome. Furthermore, the degree of certainty with which one can predict treatment outcome depends on the level of baseline symptom severity itself. This example is illustrated in Figure 1. Here we have the most certainty about the outcome of patient C and the least certainty about the outcome of patient B.

Treatment response versus treatment outcome

The PTB chart also takes into consideration the important distinction between treatment *response* and treatment *outcome*. Figure 2 illustrates this distinction. We define *response* as the magnitude of change from pre-assessment to post-assessment and *outcome* as post-treatment status. These definitions can be applied to symptom counts, standardized scores, clinical cut-off scores, or diagnostic status. Even though both patients in Figure 2 are participating in the same treatment, patient A has a larger treatment response than patient B. At the same time, patient B has a better treatment outcome than patient A. Unfortunately,

the terms *response* and *outcome* have often been used interchangeably even though the nonequivalence of these terms is evident from the psychotherapy literature (e.g., Lavigne, et al., 2008; Lindhiem & Kolko, in press). Specifically, terms such as “response to treatment” have been used to refer both to post-treatment status (i.e., *outcome*) and the change from pre-treatment to post-treatment (i.e., *response*). In one study, for example, children were classified as treatment responders if they were functioning in the normal range at post-treatment per parent and teacher report (Reid, Webster-Stratton, & Hammond, 2003). In another study, response to treatment was defined in terms of the rate-of-change (slope) of externalizing symptoms (Beauchaine, Webster-Stratton, & Reid, 2005). We suggest that the lack of consensus on definitions for terms such as *response* and *outcome* has contributed to confusing and contradictory reports. This is especially true of the literature on moderators and predictors of treatment effects, which has been observed to yield mixed and even contradictory conclusions (Lochman & Salekin, 2003). We argue that terms such as *response* and *outcome* should be defined precisely and used consistently across studies, and that both are important to evaluate and predict the outcomes of a given treatment. The distinction between these terms is further illustrated in the PTB chart. Throughout this manuscript and in the title we use the more general term “benefit” to include both treatment response and treatment outcome.

Current Study

The aim of this study was to develop a supplement to traditional effect sizes that accounts for both systematic and nonsystematic variability in treatment effects, in order to quantify likely benefit at the individual level. First, we describe the development of the Probability of Treatment Benefit (PTB) chart using real patient data from an existing data set of clinically referred children and adolescents. Our method incorporates, integrates, and extends more recent approaches to describing treatment effects, including the Reliable Change Index (RCI; Jacobson, Follette, & Revenstorf, 1984; Jacobson & Traux, 1991) and the use of normative comparisons (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999), by adding parameters that are simultaneously *probabilistic* and *individualized*. Next, we illustrate how a PTB chart is interpreted. Finally, we discuss several advantages of the PTB chart and emphasize its utility for researchers, clinicians, and patients.

Method

Participants

We illustrate the PTB chart with data from an effectiveness trial of a modular treatment for child and adolescent disruptive behavior disorders (Kolko, Dorn, Bukstein, Pardini, Holden, & Hart, 2009; Lindhiem & Kolko, 2010). In summary, the sample consisted of 139 clinically referred children (118 boys; 21 girls) aged 6–11 ($M = 8.8$; $SD = 1.6$) years from ethnically and socioeconomically diverse backgrounds. One-hundred-fifteen children (83%) met diagnostic criteria for Oppositional Defiant Disorder (ODD) and 29 (21%) met diagnostic criteria for Conduct Disorder (CD). Five children (4%) met criteria for both ODD and CD. Sixty-four children (46%) were identified by their parent as African American--not of Hispanic origin, 66 (47%) were identified as White--not of Hispanic origin, 8 (6%) were identified as biracial, and 1 (1%) was identified as Hispanic. Participants were recruited from clinics affiliated with the University of Pittsburgh Medical Center (UPMC). Approximately half of the participants represented ethnic minorities. All children had diagnoses of Oppositional Defiant Disorder (ODD) or Conduct Disorder (CD) at pre-treatment. Roughly three quarters of the children met diagnostic criteria for comorbid ADHD. Additional participant characteristics are described in the original outcome study (Kolko et al., 2009).

Procedure

General procedure—All study data were collected in accordance with the Institutional Review Board (IRB) at the University of Pittsburgh. Informed consent and assent were obtained from each parent and child respectively. Families participated in a treatment consisting of primarily cognitive-behavioral therapy (CBT) and parent-management training (PMT), but which included other modules designed to address related clinical targets (e.g., parent-child skills training, medication for ADHD, school consultation, etc.). Over 82% of the sample received 12 or more treatment sessions ($M = 17.7$; $SD = 8.2$). Child Behavior Checklist (CBCL; Achenbach, 1991) externalizing scale T scores decreased an average of 7.3 ($SD = 9.6$) points from 71.3 ($SD = 7.6$) to 63.8 ($SD = 10.0$) between pre and post-treatment. This represents a pre-post effect size of 0.84. Additional details on the treatment content and study procedures can be found in the original outcome study (Kolko et al., 2009).

Development of the PTB chart—In order to provide a more individualized description of the treatment's effectiveness, we began by selecting a baseline variable that is strongly associated with treatment outcome with which to stratify the sample. A review of the literature suggested that baseline symptom severity is a reliable and robust predictor of treatment outcome (e.g., Kazdin & Whitley, 2006; Lindhiem & Kolko, in press; Reyno & McGrath, 2006). Specifically, greater symptom severity at baseline tends to be associated with greater treatment *response* but poorer treatment *outcome*. A recent meta-analysis estimated the association between severity of child behavior and treatment outcome at $r = .40$ based on eight efficacy studies (fail safe $N = 168$) of psychosocial treatments for child externalizing behavior problems (Reyno & McGrath, 2006). In the current sample, the association between pre and post-treatment scores on the CBCL externalizing scale was $r = .42$, consistent with the meta-analytic findings.

Figure 3 shows the stratification of the sample by CBCL externalizing scale T scores at pre-treatment. The sample was stratified by half standard deviation intervals (i.e., 5 points) for ease of interpretation. As expected, the figure indicates the nonequivalence of treatment response and treatment outcome. Patients with greater baseline symptom severity evidenced a greater response to treatment ($r = .36$, $p < .001$) but a poorer outcome ($r = .42$, $p < .001$). The correlations are both positive because higher scores on the CBCL indicate greater symptom severity. Although these associations are not surprising, and perhaps due partly to a floor effect or regression-to-the-mean, the important point is that response and outcome are nonequivalent constructs. Statistical regression could plausibly explain why baseline severity is associated with response, but not why baseline severity predicts outcome. In other words, it does not account for why the rank order of the stratified groups is the same at pre and post treatment (i.e. why none of the lines in Figure 3 intersect). This is due, rather, to the stability of symptom severity. Traditionally, baseline differences have been regarded as unwanted variance or “noise” and controlled for in statistical analyses. However, the point here is that baseline symptom severity is not simply unwanted variance that needs to be statistically controlled. Rather, it is an important participant characteristic that has predictive value in terms of treatment outcome and can be used in our models to predict treatment outcome. The figure suggests differential effectiveness of the treatment for patients who begin treatment with differing levels of symptom severity. This variation illustrates why a single effect size is a poor characterization of a treatment's overall efficacy or effectiveness and the rationale behind stratification.

Next, we created several variables to reflect both treatment response and treatment outcome. Specifically, we created four dichotomous variables (Yes = 1; No = 0) to indicate whether or not each participant met specific post-treatment criteria. We chose the variables to reflect

concepts of clinical significance (e.g., Jacobson et al., 1984; Jacobson & Traux, 1991; Kendall et al., 1999). The first variable, *Normal Range*, indicates whether or not a patient is within the normal range on the CBCL externalizing scale ($T \leq 60$) at post-treatment. The second variable, *Clinical Range*, indicates whether or not a patient is within the clinical range on the CBCL externalizing scale ($T \leq 64$) at post-treatment. The third variable, *Improvement*, indicates whether or not a patient made a reliable improvement on the CBCL externalizing scale between pre and post-treatment. We defined a reliable change in terms of the Reliable Change Index (RCI) which we estimated as 8 points based on the formula provided by Jacobson and Truax (1991). The fourth variable, *Deterioration*, indicates whether or not a patient's score on the CBCL externalizing scale became reliably worse between pre and post-treatment, based on an RCI of 8 points. This definition of deterioration is based on the extant literature (e.g., Lambert, Hansen, & Finch, 2001; Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010). Finally, we calculated the probabilities that patients within each stratified band met each of these four criteria at post treatment. These probabilities, converted to percentages, are displayed in the center of the PTB chart. Logistic models were used to model probabilities of the outcome or response of interest (i.e., within normal range, within clinical range, improvement, and deterioration) across the five baseline severity groups. The probabilities were estimated from the logistic models by inverting the intercept and logit coefficients back to the probability scales. For example, the probability of having an outcome or response of interest for a subject with reference level of severity is

$p_0 = \frac{\exp(\alpha)}{1 + \exp(\alpha)}$, where α is the intercept of a logistic model, and the probability of having the same outcome or response for a subject with severity other than the reference level is,

$p_1 = \frac{\exp(\alpha + \beta)}{1 + \exp(\alpha + \beta)}$, where β is the estimated log odds ratio for that level. The 80% and 95% confidence intervals of the probabilities were also estimated. To supplement these probabilities, we also report the expected (median) treatment response and outcome for each stratified group. These values are displayed in the right two columns of the PTB chart. As there were no deterioration cases among the three highest severity groups, we compared any of the two groups using Fisher's exact test and computed the confidence intervals using the exact method based on binomial distributions.

Cross-validation—In order to conduct a cross-validation without significantly reducing the sample size, we used the “leave-one-out” method (see Chernick, 1999). The leave-one-out cross-validation method is conceptually similar to the more familiar split-half cross-validation method. With the leave-one-out method, however, a single observation is left out of the model and used as the validation sample. This is repeated once for each set of data points (N) in the sample (in this case $N = 139$) until each set of data points has served as the validation sample for N different (albeit very similar) models (in this case PTB charts), each based on $N - 1$ set of data points. Although more labor intensive than the split-half cross-validation, the leave-one-out method makes the most of each set of data points and is therefore useful for performing cross-validation with small data sets.

Replication with a related construct—We also replicated the results of the PTB method described above using global impairment instead of symptom severity. As with symptom severity, global impairment at baseline was associated with a greater response to treatment ($r = 0.41$, $p < .001$) but a poorer outcome ($r = 0.56$, $p < .001$). The same four dichotomous variables (Normal Range, Clinical Range, Improvement, and Deterioration) were created to examine response and outcome in terms of global impairment. The sample was stratified by five point intervals into the following bands: “30+” ($n = 15$), “25–29” ($n = 31$), “20–24” ($n = 35$), “15–19” ($n = 35$), and “14” ($n = 22$). Cross-validation was also performed using the “leave-one-out” method described above.

Measures

Externalizing symptoms—The externalizing behavior problems scale of the Child Behavior Checklist (CBCL; Achenbach, 1991) was used to measure treatment response and outcome. Parents completed the CBCL at the pre and post-treatment assessments (see Kolko et al., 2009 for additional details on the assessment procedures). The CBCL has well established reliability and validity. We used *T* scores because of the variability in child age and repeated assessments. *T* scores also have the advantage of being easily interpretable as they are standardized with a mean of 50 and standard deviation of 10.

Global impairment—Child impairment in daily functioning was also assessed at pre and post treatment using the Columbia Impairment Scale – Parent Version (CIS-P; Bird, Shaffer, Fisher, & Gould, 1993). The CIS-P consists of 13 items each rated on 5- point Likert scales (“no problem” to “very big problem”) with a total score ranging from 0 to 52. The measure has good internal consistency (.85–.89) and two-week test-retest reliability (.89). The CIS also corresponds well to other measures of impairment and functioning and has an established clinical cutoff of 15.

Results

Interpreting the PTB Chart

The PTB chart is presented in Figure 4. Patients with baseline symptom severity of 80 (CBCL externalizing scale *T* score) or higher only had a 9% chance of being in the normal range at post-treatment. The probability of being in the normal range at post treatment increased systematically with decreasing baseline symptom severity. Patients with baseline symptom severity of 64 or lower had the highest probability (68%) of being in the normal range at post treatment. They also had the lowest probability of being in the clinical range (16%) at post treatment.

Although patients with the greatest baseline symptom severity had the greatest likelihood of being in the clinical range and the least likelihood of being in the normal range at post treatment, they had the best chance (64%) of making a significant improvement during the course of treatment. The likelihood of making a significant improvement during the course of treatment decreased systematically with decreasing baseline symptom severity. Patients with baseline symptom severity of 64 or lower had the lowest probability of making a significant improvement during the course of treatment. Although the probability of deteriorating during the course of treatment was very low overall (4%), the probability was highest (16%) for patients with the lowest baseline symptom severity and lowest (0%) for patients with the highest baseline symptom severity. This is likely due, at least in part, to a ceiling effect as patients with the highest baseline symptom severity have less room to deteriorate.

The expected response (median) was 9 points for patients with the highest baseline symptom severity and 3 points for patients with the lowest baseline symptom severity. However, patients with the highest baseline symptom severity were still more likely to have a worse outcome (*T* score: 73) compared to patients with the lowest baseline symptom severity (*T* score: 56). In summary, the probabilities of being in the normal range and outside of the clinical range at post treatment decreased as baseline symptom severity increased. At time same time, patients with the highest baseline symptom severity were the most likely to show improvements and least likely to deteriorate during the course of treatment.

Cross-Validation

Results from the cross-validation using the leave-one-out method yielded probability values that were all within four percentage points ($M = 1.5\%$; $SD = 1.1\%$) of those reported in the PTB chart (Figure 4). The two sets of probabilities were therefore close to each other and the decreasing or increasing trends across severity groups were observed for both sets of probabilities. The cross-validation therefore indicated that the proposed method will likely provide reliable predictions with a new sample taken from the same population.

Confidence Intervals

Figure 5 presents the PTB chart from Figure 4 with 80% and 95% confidence intervals. This format is perhaps more useful for researchers and clinicians. For example, we can be 80% confident that patients with baseline symptom severity scores between 70 and 74 have a 29–33% chance of being in the normal range at post treatment. As another example, we can be 95% confident that patients with baseline symptom severity scores between 65 and 69 have a 2–24% chance of deteriorating during the course of treatment.

Significant Differences

Table 1 summarizes which probability values in the PTB chart are significantly different from one another. All comparisons were based on the chi-square test for odds ratios or the Fisher exact test when numbers of deterioration cases were small. Again, these data will be most relevant to researchers who are using a PTB chart to summarize the efficacy or effectiveness of a treatment. We see, for example, that the probability of ending treatment in the normal range for patients with baseline symptom scores of 64 or below (68%) was significantly higher than the other four groups. However, the probability of ending treatment in the normal range for patients with baseline symptom scores between 70 and 74 (31%) was not significantly higher than the probability of ending treatment in the normal range for patients with baseline symptom scores between 75 and 79 (28%). Similar relative comparisons could be made to highlight the likelihood of treatment response or outcome based on the data in the chart.

Global Impairment

The same pattern of results was replicated using global impairment (see Figures 6 and 7). Patients with lower baseline global impairment were more likely to end treatment outside of the clinical range compared to patients with higher baseline global impairment. At the same time, patients with higher baseline global impairment evidenced greater response to treatment and were more likely to show reliable improvements. The probability of deteriorating during the course of treatment was very low for all groups (3% or less). The pattern of results was consistent with the results based on symptom severity.

Discussion

The PTB chart was designed to provide clinical information that is complementary to that provided by traditional effect sizes. An effect size answers the question, “How big is the average benefit?”, whereas the PTB chart answers the question, “What are the chances this treatment will benefit me?” In the current data set, for example, the overall effect size (pre-post $ES = 0.84$) of the treatment tells us that the average patient can expect to end treatment 0.84 standard deviations below his or her pre-treatment level. This effect size does not convey the likelihood or probability that any particular individual will experience this 0.84 SD gain. As Figure 3 illustrates, there is substantial variability in terms of both response and outcome *within* the sample. The PTB chart characterizes the level of nonsystematic and systematic variability in treatment effects by summarizing expected benefits (in percentage

values) for groups of patients with particular baseline characteristics. As such, the PTB chart can be used to make individualized predictions about the expected benefit of a particular treatment.

The PTB chart also incorporates several definitions of clinically significant change, such as ending treating within the normal range, falling outside of the clinical range, or resulting in a reliable change (e.g., Jacobson et al., 1984; Jacobson & Truax, 1991; Kendall et al., 1999). These inclusions ground the interpretation of the PTB chart with clinically meaning “anchors” or reference points. Perhaps the most useful property of the PTB chart, however, is its simplicity. Its interpretation is straightforward and intuitive. To understand that one has a 70% chance of being within a normal range at post treatment, for example, is no more complicated than comprehending the likelihood of rain from the evening weather report. This makes the PTB chart accessible to patients who have not had formal training in statistics or research design.

It should be noted that there are other examples of measures in the extant literature that use percentage values to summarize treatment effects, albeit infrequently. For example, Weston and Morrison (2001) advocate the use of several indices including “percent improved,” “percent recovered,” and “percent seeking additional treatment” to supplement traditional effect sizes in reporting the results of clinical trials and meta-analyses. Unfortunately, these sound recommendations have not been incorporated into the requirements for reporting clinical significance by most research journals.

Treatment Response Versus Treatment Outcome

The PTB chart also highlights the important distinction between treatment response and treatment outcome. Specifically, patients with greater symptom severity tend to show a greater response to treatment but still have a worse prognosis in terms of treatment outcome. This is because their response (although greater) is not enough to “catch up” to their counterparts who begin treatment with fewer symptoms. The critical point is that *response* and *outcome* are distinct constructs that should not be used interchangeably. It is unfortunate that these terms have sometimes been used casually and imprecisely in the extant literature, occasionally leading to apparently inconsistent conclusions. We suggest a more general term such as “benefit” be used when referring broadly to response, outcome, or both.

This distinction between response and outcome has important clinical implications, as can be illustrated further by the PTB chart. For example, we can see that a patient might have a reasonably good chance of making a reliable (perhaps even clinically meaningful) improvement from treatment (64%) even though his or her chance of being in the “Normal Range” at post-treatment is dismal (9%). In other words, the treatment might benefit the patient even though he or she is not completely “cured” of all his or her symptoms. A more dramatic example has been given by Kazdin (1999) who suggests that change might be clinically significant even when there is no change in symptoms at all. In this case it might be that the goal of treatment is not to reduce symptoms at all, but rather help the patient to cope with his or her situation.

Limitations and Future Directions

The PTB chart used as an illustration in this paper is based on a relatively small sample size ($N = 139$) with even smaller stratified sub-samples (N s ranging from 11 to 39). As such, the confidence intervals displayed in Figures 5 and 7 are relatively large. However, PTB charts summarizing the results of larger clinical trials will have narrower confidence intervals. Similarly, the number of significant differences summarized in Table 1 are partly a function

of the sample size and resulting statistical power. A larger sample (and sub-samples) would almost certainly result in more contrasts reaching the threshold of statistical significance.

In terms of future directions, it must be noted that this paper provides just two examples of a PTB chart. There are countless improvements and modifications that could be made without sacrificing the essential characteristics or simplicity of the PTB chart. There are only two essential characteristics of the PTB chart, namely 1) reporting of treatment effects in probabilistic terms, and 2) differentiation of treatment effects based on meaningful stratification by baseline patient characteristics. Nonessential characteristics include outcome criteria and stratification method. For example, treatment response and outcome could be evaluated in terms of quality of life (Gladis, Gosch, Dishuk, & Crits-Christoph, 1999) rather than symptom severity or global impairment. For some disorders, reducing symptoms might not be a reasonable goal. Instead, an effective treatment might prevent further deterioration or improve overall quality of life.

Baseline stratification could also be based on variables other than pre-treatment symptom severity or global impairment. Meaningful variables should be selected based on the extant literature (i.e., variables known to be associated with treatment outcome). We selected baseline symptom severity because it is known to be one of the most robust predictors of treatment outcome (e.g., Reyno & McGrath, 2006). For example, a PTB chart could be stratified based on number of comorbid conditions, a variable which has been associated with treatment response and outcome (e.g., Beauchaine et al., 2005; Hartmen, Stage, & Webster-Stratton, 2003; Lavigne et al., 2008). Across studies, as with baseline symptom severity, comorbid conditions have been associated with greater *response* but poorer *outcome*. Other variables that might meaningfully differentiate treatment effects include social support (e.g., Dadds & McHugh, 1992), life stress (e.g., Lavigne et al., 2008), or referral source (e.g., Reyno & McGrath, 2006).

It should also be noted that stratification could be based on more than one baseline variable. For example, baseline profiles could be created by crossing pre-treatment symptom severity and number of comorbid conditions (e.g., High Severity/No Comorbid Conditions; High Severity/One Comorbid Condition; etc.). The only limitation is that this would require a much larger sample size in order to have sufficient sample sizes in each sub-group to calculate differential probabilities. Finally, when such data are available, it would be helpful to include probabilities for control group conditions. These values would reflect the probabilities of meeting outcome criteria if one does not receive the treatment. A patient who is weighing the pros and cons of a treatment could then compare the probability of an outcome if he or she receives the treatment with the probability of the same outcome if he or she does not receive the treatment.

Implications for Research, Policy, and Practice

Research implications—The PTB chart has utility for researchers, clinicians, and patients alike. For researchers, the PTB chart affords a transparent and efficient means to summarize the results of an efficacy or effectiveness trial. Specifically, the PTB chart summarizes both the nonsystematic and systematic variability in treatment effects in probabilistic terms. As noted above, the PTB chart provides a summary of data that complements traditional effect sizes. An effect size answers the question, “How big is the average benefit?”, whereas the PTB chart answers the question, “What are the chances (odds) this treatment will benefit a particular patient with particular baseline characteristics?” Furthermore, the PTB chart is a useful tool that can simultaneously summarize the efficacy or effectiveness of a treatment in terms of both *response* and *outcome*.

Policy implications—The focus on enhancing the potential for individual benefit is consistent with the priorities of the National Institute of Mental Health (NIMH) Strategic Plan (U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health, 2008), especially Strategic Objective 3. This objective emphasizes both patient preference and personalization of care based on an understanding of each individual's unique set of characteristics, and the adoption of a personalized approach to optimize outcomes. Data-driven, personalized selection of treatments is beginning to gain traction in medical fields including oncology and psychiatry. In oncology, for example, hormone receptor levels have the potential to classify patients who will benefit from endocrine therapy for breast cancer (Bartlett et al., 2011). In psychiatry, researchers are beginning to identify genetic markers (HTR2A and GRIK4 genes) associated with response and nonresponse to antidepressant treatment (Paddock et al., 2007). An approach such as the PTB chart provides a framework to summarize information about the expected benefit of particular treatments for individuals with particular background, genetic, or other personal characteristics. Information reflecting the expected treatment benefit for an individual patient, could promote the use of personalized medicine in which pharmacological, behavioral, and other forms of treatment are selected based on each patient's clinical profile. PTB chart data may ultimately encourage the use of personalized decision-making in the selection of a given treatment and provide a forum for discussing treatment alternatives. Thus, the chart may help mental health providers to select evidence-based interventions whose outcomes have been demonstrated in a population with an initial level of need that compares to a patient in question. "Similarly, patient preference is a powerful indicator of how well someone will adhere to treatment and must also be part of a personalized approach to mental health intervention" (U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health, 2008, p. 26). Patient preference can only be meaningful when patients are presented with easily understandable yet relevant data on which to base informed decisions about their treatment options. Ultimately, the PTB method extends this objective by providing a simple clinical tool that has the potential to allow individual patients to make informed decisions that are based on personalized data.

Practice implications—For clinicians, the PTB chart is a simple and intuitive tool with which to inform patients about the expected benefit of a treatment. This will aid the decision making process regarding whether a particular treatment is a "good fit" for the patient. For example, a clinician might decide that a patient has unacceptably low chances of benefiting from a given treatment and requires a referral to a more intensive treatment program. Patients perhaps have the most to gain from a clinical tool such as the PTB chart. For patients, the PTB chart affords the opportunity to make more informed decisions before beginning a treatment. Psychosocial treatments in particular can involve significant investments of time and emotional resources. Treatments can also involve significant financial investments in terms of co-payments for the insured and even greater out-of-pocket expenses for the uninsured. The PTB chart allows a patient to base his or her decision about beginning a treatment on the probability that the treatment will benefit a patient with his or her baseline profile, and weigh the costs and benefits accordingly. It is important to underscore that the clinical utility of a PTB chart can only come after the PTB method has been applied to existing datasets of particular treatments and patient populations. That is, the PTB method must first be applied to nomothetic datasets before idiographic predictions can be made. Although the distinction we make between "response" and "outcome" may be a potential source of confusion for some patients, a simple diagram such as Figure 2 can be used to illustrate the distinction. A simplified version of a PTB chart without confidence intervals (Figures 4 and 6) may also be used for simplicity.

Conclusions

In conclusion, we have introduced the PTB chart as a supplemental approach to calculating and summarizing the efficacy or effectiveness of a psychosocial treatment. We suggest that this method bears significant practice and research implication, and lends itself to diverse applications. Of course, the ultimate aim of the PTB chart and related methods is to efficiently summarize the potential clinical benefits and limitations of a specific treatment regimen so that patients and their practitioners can make informed decisions about treatment participation.

Acknowledgments

This study was supported by grant funding from the National Institute of Mental Health to the second author (R01MH057727) and by a T32 Fellowship (T32MH018951; PI: David Brent, MD) and K-award (K01MH093508) to the first author. The authors acknowledge the research and clinical staff of the Resources to Enhance the Adjustment of Children (REACH) program.

References

- Achenbach, TM. Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.
- Bartlett JM, Brookes CL, Robson T, van de Velde CJ, Billingham LJ, Campbell FM, Rea D. Estrogen receptor and progesterone receptor as predictive biomarkers of response to endocrine therapy: A prospectively powered pathology study in the Tamoxifen and Exemestane Adjuvant Multinational trial. *Journal of Clinical Oncology*. 2011; 29(12):1531–1538.10.1200/JCO.2010.30.3677 [PubMed: 21422407]
- Beauchaine TP, Webster-Stratton C, Reid MJ. Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology*. 2005; 73(3):371–388.10.1037/0022-006X.73.3.371 [PubMed: 15982136]
- Bird H, Shaffer D, Fisher P, Gould MS. The Columbia impairment scale (CIS): Pilot findings on a measure of global impairment for children and adolescents. *International Journal of Methods in Psychiatric Research*. 1993; 3:167–176.
- Chernick, MR. Bootstrap Methods: A Practitioner's Guide. New York: John Wiley & Sons, Inc; 1999.
- Dadds MR, McHugh TA. Social support and treatment outcome in behavioral family therapy for child conduct problems. *Journal of Consulting and Clinical Psychology*. 1992; 60(2):252–259. [PubMed: 1592955]
- Eyberg SM, Nelson MN, Boggs SR. Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology*. 2008; 37(1):215–237.10.1080/15374410701820117 [PubMed: 18444059]
- Gladis MM, Gosch EA, Dishuk NM, Crits-Christoph P. Quality of life: Expanding the scope of clinical significance. *Journal of Consulting and Clinical Psychology*. 1999; 67(3):320–331. [PubMed: 10369052]
- Hartman RR, Stage SA, Webster-Stratton C. A growth curve analysis of parent training outcomes: Examining the influence of child risk factors (inattention, impulsivity, and hyperactivity problems), parental and family risk factors. *Journal of Child Psychology and Psychiatry*. 2003; 44(3):388–398. [PubMed: 12635968]
- Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*. 1984; 15(4):336–352.
- Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*. 1991; 59(1):12–19. [PubMed: 2002127]
- Kazdin AE. The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*. 1999; 67(3):332–339. [PubMed: 10369053]

- Kazdin AE, Whitley MK. Comorbidity, case complexity, and effects of evidence-based treatment for children referred for disruptive behavior. *Journal of Consulting and Clinical Psychology*. 2006; 74(3):455–467.10.1037/0022-006X.74.3.455 [PubMed: 16822103]
- Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*. 1999; 67(3):285–299. [PubMed: 10369049]
- Kolko DJ, Dorn LD, Bukstein OG, Pardini D, Holden EA, Hart J. Community vs. clinic-based modular treatment of children with early-onset ODD or CD: A clinical trial with 3-year follow-up. *Journal of Abnormal Child Psychology*. 2009; 37:591–609.10.1007/s10802-009-9303-7 [PubMed: 19221871]
- 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(2):159–172.10.1037//0022-006X.69.2.159 [PubMed: 11393594]
- Lavigne JV, LeBailly SA, Gouze KR, Cicchetti C, Jessup BW, Arend R, Binns HJ. Predictor and moderator effects in the treatment of oppositional defiant disorder in pediatric primary care. *Journal of Pediatric Psychology*. 2008; 33(5):462–472.10.1093/jpepsy/jsmo75 [PubMed: 17956931]
- Lindhiem O, Kolko DJ. Trajectories of symptom reduction and engagement during treatment for childhood behavior disorders: Differences across settings. *Journal of Abnormal Child Psychology*. 2010; 38(7):995–1005.10.1007/s10802-010-9416-z [PubMed: 20414715]
- Lindhiem O, Kolko DJ. Trajectories of symptom reduction during treatment for behavior problems in pediatric primary-care settings. *Administration and Policy in Mental Health and Mental Health Services Research*. (in press).
- Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*. 1993; 48(12):1181–1209.10.1037/0003-066X.48.12.1181 [PubMed: 8297057]
- Lochman JE, Salekin RT. Prevention and intervention with aggressive and disruptive children: Next steps in behavioral intervention research. *Behavior Therapy*. 2003; 34(4):413–419.
- Paddock S, Laje G, Charney D, Rush JA, Wilson AF, Sorant AJM, McMahon FJ. Association of GRIK4 With Outcome of Antidepressant Treatment in the STAR*D Cohort. *American Journal of Psychiatry*. 2007; 164(8):1181–1188.10.1176/appi.ajp.2007.06111790 [PubMed: 17671280]
- Reid MJ, Webster-Stratton C, Hammond M. Follow-up of children who received the Incredible Years intervention for oppositional-defiant disorder: Maintenance and prediction of 2-year outcome. *Behavior Therapy*. 2003; 34(4):471–491.
- Reyno SM, McGrath PJ. Predictors of parent training efficacy for child externalizing behavior problems – a meta-analytic review. *Journal of Child Psychology and Psychiatry*. 2006; 47(1):99–111.10.1111/j.1469-7610.2005.01544.x [PubMed: 16405646]
- Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *American Psychologist*. 1977; 32(9):752–760.10.1037/0003-066X.32.9.752 [PubMed: 921048]
- U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health. The National Institute of Mental Health strategic plan. NIH Publication No. 08-6368. 2008. Retrieved from <http://www.nimh.nih.gov/about/strategic-planning-reports/nimh-strategic-plan-2008.pdf>
- Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology*. 2010; 78(2):144–155.10.1037/a0018544 [PubMed: 20350026]
- Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*. 2006; 61(7):671–689.10.1037/0003-066X.61.7.671 [PubMed: 17032068]
- Weisz JR, Weiss B, Alicki MD, Klotz ML. Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*. 1987; 55(4):542–549. [PubMed: 3624610]

- Weisz JR, Weiss B, Han SS, Granger DA, Morton T. Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*. 1995; 117(3):450–468. [PubMed: 7777649]
- Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*. 2001; 69(6):875–899.10.1037/0022-006X.69.6.875 [PubMed: 11777114]

\$watermark-text

\$watermark-text

\$watermark-text

Highlights

- We describe the development of the Probability of Treatment Benefit (PTB) chart.
- The method extends existing approaches to describing treatment effects.
- The results highlight the individual and probabilistic nature of psychotherapy benefit.
- We discuss the research, policy, and practice implications of the PTB chart.

Figure 1.

Distributions of baseline symptom severity scores for patients with good treatment outcome and poor treatment outcome. Patient A has a better chance of good treatment outcome than patient B but still a reasonable chance of poor treatment outcome. Patient B has a 50/50 chance of good treatment outcome. Patient C has only a very small chance of good treatment outcome. Initial symptom severity scores that fall along the darker gray rectangle (most patients) represent values for which treatment outcome can only be reasonably anticipated in *probabilistic* terms.

Figure 2. Differential response to treatment. Patient A has a larger treatment response than patient B. However, patient B has a better treatment outcome than patient A.

Figure 3. Reductions in CBCL externalizing scale *T* scores from pre to post-treatment with stratification based on pre-treatment symptom severity.

PTB		Probability of Treatment Benefit (CBCL externalizing scale <i>T</i> scores)				Baseline Score: _____	
		Normal Range	Clinical Range	Improvement (Reliably Better)	Deterioration (Reliably Worse)	Expected Response	Expected Outcome
Baseline Symptom Severity	80+	9%	91%	64%	0%	9 pts	73
	75-79	28%	69%	44%	0%	7 pts	68
	70-74	31%	64%	41%	0%	7 pts	65
	65-69	41%	48%	38%	7%	4 pts	63
	≤ 64	68%	16%	21%	16%	3 pts	56
	TOTAL	36%	57%	40%	4%	7 pts	65

Figure 4.
PTB chart for symptom severity.

PTB		Probability of Treatment Benefit (CBCL externalizing scale <i>T</i> scores)							
		Normal Range		Clinical Range		Improvement (Reliably Better)		Deterioration (Reliably Worse)	
		80% CI	95%CI	80% CI	95%CI	80% CI	95%CI	80% CI	95%CI
Baseline Symptom Severity	80+	7-12%	1-43%	88-93%	56-99%	60-67%	34-86%	0-19% ¹	0-28% ¹
	75-79	26-30%	16-44%	67-71%	53-82%	42-47%	29-61%	0-6% ¹	0-10% ¹
	70-74	29-33%	18-47%	62-66%	48-77%	39-43%	27-57%	0-6% ¹	0-9% ¹
	65-69	39-44%	25-60%	46-51%	31-66%	36-40%	22-57%	6-8%	2-24%
	≤ 64	66-71%	45-85%	14-18%	5-39%	19-24%	8-45%	14-18%	5-39%
	TOTAL	35-37%	28-44%	56-59%	49-66%	39-41%	32-49%	3-4%	2-9%

Figure 5.
PTB chart for symptom severity with 80% and 95% confidence intervals (CIs).

PTB		Probability of Treatment Benefit (CIS Global Impairment)				Baseline Score: _____	
		Normal Range	Clinical Range	Improvement (Reliably Better)	Deterioration (Reliably Worse)	Expected Response	Expected Outcome
Baseline Global Impairment	30+	20%	80%	73%	0%	13 pts	23
	25-29	39%	61%	61%	3%	10 pts	16
	20-24	50%	50%	47%	3%	7 pts	15
	15-19	71%	29%	37%	3%	6 pts	11
	≤ 14	95%	5%	19%	0%	4 pts	7
	TOTAL	57%	43%	46%	2%	7 pts	12

Figure 6.
PTB chart for global impairment.

PTB		Probability of Treatment Benefit (CIS Global Impairment)							
		Normal Range		Clinical Range		Improvement (Reliably Better)		Deterioration (Reliably Worse)	
		80% CI	95%CI	80% CI	95%CI	80% CI	95%CI	80% CI	95%CI
Baseline Global Impairment	30+	18-23%	7-47%	77-82%	53-93%	70-76%	47-90%	0-14% ¹	0-22% ¹
	25-29	37-41%	23-57%	59-63%	43-77%	59-63%	43-77%	0-12%	0-17%
	20-24	48-52%	34-66%	48-52%	34-66%	45-49%	31-64%	0-11%	0-15%
	15-19	69-73%	55-84%	27-31%	16-45%	35-39%	23-54%	0-11%	0-15%
	≤ 14	94-96%	73-99%	4-6%	1-27%	17-21%	7-41%	0-10% ¹	0-16% ¹
	TOTAL	56-58%	48-65%	42-44%	35-52%	45-47%	38-55%	1-5%	0-6%

Figure 7.
PTB chart for global impairment with 80% and 95% confidence intervals (CIs).

Table 1
Significant Differences Between Probability Values in the PTB Chart (Figure 4)

	Normal Range	Clinical Range	Improvement (Reliably Better)	Deterioration (Reliably Worse)	Expected Response	Expected Outcome
Baseline Symptom Severity	1. 80+ 4 [†] , 5 ^{***}	4 [*] , 5 ^{***}	5 [*]		5 [*]	2 [*] , 3 ^{***} , 4 ^{***} , 5 ^{***}
	2. 75–79 5 ^{***}	4 [†] , 5 ^{***}	5 [†]	5 [*]	4 [†] , 5 ^{***}	1 [*] , 4 [*] , 5 ^{***}
	3. 70–74 5 ^{***}	5 ^{***}		5 [*]	5 ^{***}	1 ^{***} , 5 ^{***}
	4. 65–69 1 [†] , 5 [†]	1 [*] , 2 [†] , 5 [*]			2 [†]	1 ^{***} , 2 [*] , 5 [*]
	5. 64 1 ^{***} , 2 ^{***} , 3 ^{***} , 4 [†]	1 ^{***} , 2 ^{***} , 3 ^{***} , 4 [*]	1 [*] , 2 [†]	2 [*] , 3 [*]	1 [*] , 2 ^{***} , 3 ^{***}	1 ^{***} , 2 ^{***} , 3 ^{***} , 4 [*]

[†] p < .10.
^{*} p < .05.
^{***} p < .01.