Cumulative Proportion of Responders Analysis (CPRA) as a Tool to Assess Treatment Outcome in Alcohol Clinical Trials

DANIEL E. FALK, PH.D.,^{a,*} RAYE Z. LITTEN, PH.D.,^a RAYMOND F. ANTON, M.D.,^b HENRY R. KRANZLER, M.D.,^c BANKOLE A. JOHNSON, M.D., PH.D.,^d AND THE ACTIVE WORKGROUP^e

^aDivision of Treatment and Recovery Research, National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland

^bDepartment of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, South Carolina

^cDepartment of Psychiatry, Perelman School of Medicine at the University of Pennsylvania and Veterans Integrated Service Network 4, Mental Illness Research, Education and Clinical Center, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania

^dDepartment of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland

^eSee Acknowledgments

ABSTRACT. Objective: Several definitions of treatment response have been proposed for alcohol clinical trials (e.g., abstinence and no heavy drinking). However, each of these outcomes allows only one definition of successful response. In contrast, the cumulative proportion of responders analysis (CPRA) includes all of the possible drinking response cutoff points, providing a more complete picture of the therapeutic effects of a treatment. CPRA has been used to examine the efficacy of analgesics but not alcohol pharmacotherapy. To demonstrate its potential utility, we conducted CPRA in two large alcohol treatment trials: the COMBINE (Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence) trial (naltrexone) and a multisite topiramate trial. CPRA was used to demonstrate the efficacy of naltrexone and topiramate on continuous measures of in-treatment drinking—heavy drinking days

and drinks per day—and their reductions from pretreatment. **Method:** All possible cutoff points were portrayed for each measure. We provide graphs to illustrate the effects of the active medications compared with placebo and examined them statistically over a number of salient drinking outcomes to evaluate their efficacy. **Results:** Treatment group responder curves were not parallel across the entire range of cutoff points; rather, they separated only at lower levels of drinking. In general, effect sizes increased by 0.10–0.15 when going from the lowest drinking level cutoff (i.e., abstinence and no heavy drinking) to the cutoff associated with the maximal treatment effect. **Conclusions:** CPRA may be useful in designing subsequent trials and helping to illustrate for treatment providers the likelihood of treatment success given various definitions of a positive response. (*J. Stud. Alcohol Drugs*, 75, 335–346, 2014)

THE OPTIMAL PRIMARY TREATMENT ENDPOINT in alcohol treatment trials has been debated over the past two decades. Endpoints used to evaluate efficacy have included continuous measures of drinking (e.g., percentage days abstinent, drinks per day [DPD], and percentage of heavy drinking days [PHDDs]) and dichotomous measures of drinking (e.g., percentage of subjects abstinent and percentage of subjects with no heavy drinking days [HDDs]) (Falk et al., 2010). Nondrinking endpoints have included alcohol-related consequences, alcohol craving, and a variety of biomarkers (Litten et al., 2010; Lowman et al., 2000; Miller et al., 1995). In general, nondrinking endpoints are less sensitive than drinking endpoints in detecting medication effects.

A workshop sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (Allen, 2003) concluded that the continuous outcome measure, PHDDs,

should be the primary endpoint in trials, with an HDD defined as being five or more drinks in a day for men and four or more for women. The rationale for this decision was that PHDDs reflect both the quantity and frequency of alcohol consumption. The 4+/5+ cutoff used to define an HDD has been associated with alcohol-related consequences (Breslow and Graubard, 2008; Dawson et al., 2008; Greenfield et al., 2005; Jackson, 2008) and has helped to define the NIAAA guidelines for heavy and high-risk drinking (NIAAA, 2007).

Recently, use of a dichotomous endpoint has been promoted to increase acceptance of medication by the treatment community (O'Malley et al., 2003, 2007). A dichotomous outcome is more readily interpreted than a continuous outcome, allowing for the establishment of concrete guidelines for medications development (Falk et al., 2010). The U.S. Food and Drug Administration (FDA), for instance, suggests the dichotomous outcome “percent subjects with no HDDs” (PSNHDDs) as a primary endpoint for Phase 3 trials (FDA, 2006, p. 4). This endpoint appears to be as sensitive as many of the traditional continuous outcomes in detecting differences between medication and placebo groups (Falk et al., 2010). The European Medicines Agency (EMA) recommends total abstinence, as well as a 50%, 70%, and 90% reduction in

Received: April 19, 2013. Revision: July 22, 2013.

*Correspondence may be sent to Daniel E. Falk at the National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5635 Fishers Lane, Room 2040, Rockville, MD 20852, or via email at: falkde@mail.nih.gov.

alcohol consumption as indicators of successful response to treatment (EMA, 2010). Despite their greater clinical utility, these measures have limitations. For example, PSNHDDs allow for only one definition of a successful response to treatment, and the suggested percentage reduction levels in alcohol consumption may be considered arbitrary and of limited clinical importance, depending on the perspective of the clinician, insurer, or researcher.

Cumulative proportion of responders analysis (CPRA) provides a more informative evaluation of effectiveness, presenting the treatment effect at every possible responder definition (i.e., drinking cutoff; Farrar et al., 2006). CPRA is a comparison of the cumulative distribution functions (CDFs) of response variables between two groups (Good, 2008). The term *CPRA* appears to have gained traction in the field (Cappelleri et al., 2009; Moore et al., 2008). CPRA presents the proportion of responders at all possible cutoff points, thus allowing a more complete picture of the therapeutic effects of an experimental medication. CPRA has been used to show the efficacy of analgesics (Cappelleri et al., 2009; Farrar et al., 2006; Kim, 2011; Moore et al., 2008, 2010, 2011; Pritchett et al., 2007; Straube et al., 2010) and is illustrated graphically in the FDA-approved labeling for several medications, including Lyrica (pregabalin) and Cymbalta (duloxetine hydrochloride) (Farrar, 2008; *Physicians' Desk Reference*, 2011a, 2011b). Although CPRA has not been used to evaluate medications for alcoholism treatment and is not part of the FDA labeling for these medications, it allows a more comprehensive understanding of the clinical benefit of a treatment and may assist in the selection of appropriate endpoints for clinical trials.

In this article, we present findings using CPRA in a secondary analysis of data from two large, multisite alcohol treatment trials of varying design: the COMBINE (Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence) trial (Anton et al., 2006) and a multisite topiramate trial (Johnson et al., 2007). CPRA was used to represent the full range of possible cutoff points for a number of commonly used continuous drinking outcomes, including frequency of HDDs, total alcohol consumption (DPD), and their reductions from pretreatment. For each cutoff point, the proportions of responders to active medication and placebo were compared to describe how the treatment effect varies by cutoff and to identify the cutoff that maximizes the treatment effect. In addition, we analyzed the data statistically over a range of potentially clinically acceptable endpoints.

Method

Data

Data were obtained from the COMBINE trial (Anton et al., 2006) and a multisite topiramate trial (Johnson et al., 2007)—two randomized, double-blind, placebo-controlled

alcohol dependence treatment trials that yielded evidence of medication efficacy. The studies differed in experimental design, type of medication examined, and the alcohol-dependent population studied (Falk et al., 2010). The COMBINE trial, conducted between 2001 and 2004, compared the efficacy of oral naltrexone (100 mg/day) and acamprosate (3 g/day), administered alone or in combination, and two behavioral interventions: medical management (MM) alone or MM with a combined behavioral intervention (CBI). Data from a CBI-only (no pills) group were not analyzed in the current study. We present CPRA graphs for subjects who received naltrexone or placebo (with or without acamprosate) and MM only, as described below; however, secondary analyses included only subjects who received both MM and CBI. Patients were required to be abstinent for 4–21 days before randomization. The trial randomized 1,383 patients and had a medication adherence rate of 86% during the 4-month treatment phase. The topiramate trial, conducted between 2004 and 2006, examined the efficacy of oral topiramate (300 mg/day) in conjunction with a behavioral intervention (brief behavioral compliance enhancement therapy). This trial randomized 371 subjects and had a medication adherence rate of 91% during the 3-month treatment phase. Unlike the COMBINE trial, patients in the topiramate trial were included only if they continued to drink up to the time of randomization. The rationale, background, design, methodology, sample characteristics, and main findings of both studies have been documented extensively elsewhere (Anton et al., 2006; Johnson et al., 2007).

Measures

For both the COMBINE and topiramate trials, the following continuous drinking outcomes were computed in an identical manner: (a) HDDs during treatment, (b) relative reduction (RR) in PHDDs from pretreatment, (c) DPD during treatment, and (d) RR in DPD. RR from pretreatment is a percentage calculated as: $\{[(\text{pretreatment value}) - (\text{treatment value})] / (\text{pretreatment value})\} \times 100$. For the COMBINE data, drinking outcomes were computed during the last 3 months of the treatment period (Months 2–4) based on our previous work, which showed a significant effect of naltrexone on the PSNHDDs outcome given a 1-month grace period (Falk et al., 2010). Pretreatment drinking was computed for the period of 31–90 days before randomization. This period was chosen because drinking decreased, on average, during the 30 days before randomization (due, in part, to the required 4–21 days of abstinence before randomization). For the topiramate data, drinking outcomes were computed during the last 8 weeks of treatment (Weeks 7–14), reflecting a 6-week grace period during which topiramate was titrated to full dosage. Pretreatment drinking was computed during days 1–28 before randomization because the study protocol required that subjects not be abstinent before randomization.

An HDD was defined as five or more drinks in a day for men and four or more for women. A standard drink was 0.5 oz. of absolute alcohol, which is equivalent to 10 oz. of beer, 4 oz. of wine, or 1 oz. of 100-proof distilled spirits (Miller et al., 1991). In the COMBINE trial, data used to compute pretreatment drinking measures were obtained using the Form-90 assessment of daily alcohol consumption (Miller, 1996)—an instrument based on the Timeline Followback method (Sobell and Sobell, 1992), which was used to collect within-treatment daily drinking data. In the topiramate study, drinking measures were obtained using the Timeline Followback method and a diary card as a memory aid.

Statistical methods

The overarching goal of the CPRA approach is to demonstrate a broad range of possible treatment outcomes; therefore, we chose to use the strongest reported results of the two studies first for illustrative purposes. According to results reported in the main COMBINE article (Anton et al., 2006), naltrexone was efficacious, but only when used in conjunction with MM and not with CBI. Thus, in the present study, the primary analyses included only subjects who received MM and not CBI. The naltrexone group (with MM only) ($n = 302$) included subjects taking naltrexone and acamprosate placebo ($n = 154$) plus subjects taking naltrexone and acamprosate ($n = 148$). The placebo group ($n = 305$) included subjects taking naltrexone placebo and acamprosate placebo ($n = 153$) and naltrexone placebo and acamprosate ($n = 152$). As a validity check against the main COMBINE article, we also conducted secondary CPRAs and included only subjects who received both MM and CBI (i.e., the other four pill-taking cells). In this sample, the naltrexone group ($n = 312$) included subjects taking the naltrexone and acamprosate placebo ($n = 157$) plus subjects taking naltrexone and acamprosate ($n = 155$). The placebo group ($n = 307$) included subjects taking naltrexone placebo and acamprosate placebo ($n = 156$) and naltrexone placebo and acamprosate ($n = 151$). In contrast to the COMBINE trial, the topiramate trial had a simple two-group design (medication and placebo groups). Because a significant treatment effect was found in the main topiramate analyses using all available data (Johnson et al., 2007), we included data from both groups: topiramate ($n = 179$) and placebo ($n = 185$).

We calculated CPRA as described by Farrar et al. (2006). For HDDs and DPD during treatment, for each level of drinking outcome, the proportion of responders was the percentage of subjects whose drinking was less than or equal to the cutoff point. In contrast, for the RR outcomes, the proportion of responders was the percentage of subjects whose drinking was greater than or equal to the cutoff point. For each cutoff point, the proportion of responders in the active medication group was compared with that of the placebo group, and the medication effect size, Cohen's h for inde-

pendent proportions (Cohen, 1992), was computed. Cohen's $h = \varphi_{\text{medication}} - \varphi_{\text{placebo}}$, where $\varphi = 2\arcsin(\text{proportion})^{1/2}$ with the following effect size interpretations: small = .20, medium = .50, and large = .80 (Cohen, 1992). Results are presented graphically for each drinking outcome. Graphical outputs of cumulative distribution functions can be readily obtained using the CDFPLOT statement in Proc Univariate in SAS Version 9.2 (SAS, 2010). We conducted statistical inferential tests on 11 cutoff points for each outcome, using Pearson's chi-square and a Bonferroni correction (family-wise $p = .05$) to control type 1 error (a conservative approach in CPRA analysis because adjacent drinking cutoffs are highly correlated), resulting in a $p < .0045$ threshold necessary to achieve statistical significance. The 11 cutoff points were chosen a priori and focused on the lower thresholds of drinking (i.e., the right half on the CPRA graphs) because both FDA and EMA guidance recommend cutoffs within this range. For example, EMA guidelines suggest RR cutoffs of 50%, 70%, 90%, and 100% (total abstinence) (EMA, 2010). Thus, we chose to test 11 cutoffs to enable significance testing of the RR cutoffs from 50% to 100% by increments of 5%. For HDDs in the COMBINE trial, the 11 cutoffs ranged from 0 to 50 HDDs by increments of 5 (covering approximately 60% of the possible 84 days); in the topiramate study, the 11 cutoffs tested ranged from 0 to 30 HDDs by increments of 3 (covering a comparable percentage of possible days—54% of the possible 56 days). For DPD in both studies, the 11 cutoffs ranged from 0 to 10 DPD by increments of 1.

Drinking data were missing for approximately 6% of the subjects in the COMBINE trial (distributed evenly across the medication and placebo groups), largely because of study dropout. The topiramate trial had 37% missing data, compared with 22% for placebo, a significant difference. In both trials, for heavy drinking, days with missing data were recoded to reflect heavy drinking. For the other drinking variables, days with missing data were assumed to be drinking days and were imputed using the subject's mean drinks per drinking day during the 30 days before the individual's last drink before randomization (COMBINE trial) and the last week before randomization (topiramate trial). Although this conservative imputation method helps to determine the sensitivity of the data given strict assumptions, it may result in biased errors for the outcomes. The CPRA graphs presented in this study (Figures 1–4) used imputed data, whereas both imputed and nonimputed results are summarized in Table 1 for completeness. All analyses were conducted using SAS Version 9.2 (SAS Institute Inc., Cary, NC).

Results

Across each of the four outcomes and two treatment trials, the CPRA graphs (Figures 1–4) showed a separation between the medication and placebo groups, consistent with

TABLE 1. Summary of cumulative proportion of responder analyses

Number of HDD during treatment				
Study	Imputed	Cutoff where peak h occurs	h at peak	h at 0 HDD
COMBINE	Yes	10–30 HDD	.29	.17
	No	17–30 HDD	.33	.17
Topiramate	Yes	6–12 HDD	.43	.34
	No	15 HDD	.68	.55
RR in % HDD from pretreatment				
Study	Imputed	Cutoff where peak h occurs	h at peak	h at 100% RR
COMBINE	Yes	70%–92% RR	.28	.17
	No	70%–97% RR	.27	.17
Topiramate	Yes	75%–80% RR	.44	.34
	No	61% RR	.69	.55
Number of DPD during treatment				
Study	Imputed	Cutoff where peak h occurs	h at peak	h at 0 DPD
COMBINE	Yes	1–3 DPD	.25	.13
	No	3 DPD	.26	.13
Topiramate	Yes	0 DPD	.44	.44
	No	2 DPD	.62	.55
RR in % DPD from pretreatment				
Study	Imputed	Cutoff where peak h occurs	h at peak	h at 100% RR
COMBINE	Yes	84% RR	.33	.13
	No	85% RR	.34	.13
Topiramate	Yes	95% RR	.50	.44
	No	60%–70%, 95% RR	.65	.55

Notes: HDD = heavy drinking days; COMBINE = Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (trial); RR = relative reduction; DPD = drinks per day; h = Cohen's h , a measure of effect size for the difference between two proportions where $h = \varphi_{\text{Medication}} - \varphi_{\text{Placebo}}$, where $\varphi = 2\arcsin(\text{proportion})^{1/2}$.

the previously reported results. Consistently, the responder curves were not parallel across the entire range of cutoff points but, rather, separated only toward the right side of the spectrum, with cutoffs that represent less drinking. We detail below the outcome for each trial, with the greatest focus on HDDs.

Heavy drinking days

In the COMBINE trial, the naltrexone and placebo responder curves began to separate at a cutoff of ~66 HDDs (i.e., up to and including 66 of 84 possible HDDs as a positive response) and continued to separate gradually until reaching maximal separation at cutoffs between 10 and 30 HDDs, followed by a slight diminution in separation at cutoffs between 0 and 9 HDDs (Figure 1a). Using a cutoff of no HDDs (PSN-

HDDs) to define a treatment responder, the proportion of responders was 7.9% greater in the naltrexone group than the placebo group (37.1% vs. 29.2%, respectively), a nonsignificant treatment effect ($h = .17$, $p = .039$; Bonferroni-corrected significance requires $p < .0045$) (see cutoff on right side of graph). The maximal treatment effect was achieved for higher cutoffs, ranging from 10 to 30 HDDs and was statistically significant for all cutoffs tested within this range (i.e., 10, 15, 20, 25, and 30 HDDs; $ps < .0029$). For instance, the proportion of subjects having 10 or fewer HDDs was 13.8% greater in the naltrexone group than the placebo group (65.6% vs. 51.8%, respectively), reflecting a larger, statistically significant treatment effect ($h = .28$, $p = .0006$) than for 0 HDDs.

Although the overall pattern of results for the topiramate trial and COMBINE trial was similar, the topiramate effects were larger (Figure 1b). Between the cutoffs of 45–55 HDDs, the placebo group did slightly better than the topiramate group, a trend that reversed starting at a cutoff of less than 44 of 56 HDDs. The curves continued to separate gradually, favoring topiramate, until reaching maximal separation between 6 and 12 HDDs, followed by a slight diminution at cutoffs of less than 5 HDDs. The topiramate group performed significantly better than the placebo group at all cutoffs tested from 0 to 15 HDDs ($ps < .0014$). The proportion of subjects with no HDDs was 10.2% greater in the topiramate group than the placebo group (15.6% vs. 5.4%, respectively), a significant treatment effect ($h = .34$, $p = .0014$). This treatment effect increased to a maximum of $h = .43$ ($p = .00005$) at a cutoff of less than 6 HDDs. The proportion of subjects with this outcome (irrespective of treatment group) was much lower in the topiramate trial than in the COMBINE trial, which likely reflects the latter study's required abstinence at study entry and its greater emphasis on a goal of complete abstinence.

Relative reduction in percentage of heavy drinking days from pretreatment

Results for the RR in PHDDs from pretreatment are shown in Figure 2. These reflect the number of HDDs during treatment, adjusted for each participant's pretreatment number of HDDs. In the COMBINE trial, naltrexone was only slightly more efficacious than placebo, given cutoffs reflecting a 0%–21% RR in PHDDs (all $hs < 0.10$). However, a gradual increase in the effect of medication over placebo began to appear at 22% or more RR in PHDDs, peaking from 70% to 92% (and statistically significant at all cutoffs tested in this range, $ps < .0024$), followed by a noticeable diminution and loss of statistical significance thereafter (Figure 2a). Topiramate displayed gradually increasing treatment effects starting at 13% or more RR in PHDDs, peaking at 75%–80% (Figure 2b) but with statistical significance evident for every tested cutoff from 50% to 100% RR ($ps < .0035$). In neither study was the maximal treatment effect observed at the cutoff that reflected a maximal reduction in HDDs (100%

FIGURE 1. Cumulative proportion of responders analysis: Heavy drinking days. COMBINE = Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (trial).

*significant, $p < .0045$; -non-significant, $p \geq .0045$.

RR in PHDDs—equivalent to 0 HDDs during treatment). In the COMBINE trial, the maximal treatment effect occurred within the range of 70%–92% RR in PHDDs. For instance, the proportion of subjects with 88% or more RR in PHDDs was 13.7% higher in the naltrexone group than in the placebo group (58.9% vs. 45.2%, respectively; $h = .28$). In the topiramate trial, the maximal treatment effect occurred in the range of 75%–80% RR in PHDDs (e.g., $h = .44$ at 76% RR in PHDDs).

Drinks per day and relative reduction in drinks per day from pretreatment

Results for DPD during treatment and the RR in DPD from pretreatment are illustrated in Figures 3 and 4, respectively. In the COMBINE trial, for DPD, the naltrexone and placebo curves essentially overlapped (i.e., most h s < 0.10) with responder cutoffs greater than 3 DPD (Figure 3). The treatment effect reached a relative maximum when

FIGURE 2. Cumulative proportion of responders analysis: Relative reduction in percentage of heavy drinking days. COMBINE = Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (trial).

*significant, $p < .0045$; $\neg$ nonsignificant, $p \geq .0045$.

the responder definition ranged from 1 to 3 DPD (h s = 0.23–0.25) yet was only statistically significant for 2 and 3 DPD ($p = .0033$ and $p = .0022$, respectively) before dropping substantially for total abstinence, which was not statistically significant (i.e., 0 DPD; $h = .13$, $p = .105$). In the topiramate trial, however, the treatment effect was greatest and statistically significant at total abstinence (i.e., 0 DPD; $h = .44$, $p = .0003$) and remained elevated at 1–3 DPD ($ps < .0013$) be-

fore steadily decreasing to zero and becoming nonsignificant thereafter.

The responder profiles for the RR in DPD for both the COMBINE trial and the topiramate trial (Figure 4) were similar to their responder profiles for the RR in PHDDs (Figure 2). In the COMBINE trial, the treatment effect gradually increased from near zero at a cutoff of 0% RR in DPD, reached a maximum with a responder definition of

FIGURE 3. Cumulative proportion of responders analysis: Drinks per day. COMBINE = Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (trial).

*significant, $p < .0045$; -nonsignificant, $p \geq .0045$.

at least an 84% RR in DPD ($h = .33$), followed by a fairly sharp diminution thereafter (Figure 4a). The treatment effect was statistically significant at the 80%, 85%, and 90% cutoffs ($ps < .0015$). The topiramate and placebo curves essentially overlapped until a cutoff of at least 33% RR in DPD, at which point the treatment effect increased steadily (with the exception of a small dip between 82% and 92%) before reaching a maximum at a 95% RR in DPD ($h =$

.50). The treatment effect was significant at all points tested from 50% to 100% ($ps < .0035$).

Effect of missing drinking data imputation on responder profiles

The above analyses were repeated without imputing missing data. As evident from the CPRA summary (Table 1), in the COMBINE trial, where there were few missing data,

FIGURE 4. Cumulative proportion of responders analysis: Relative reduction in drinks per day. COMBINE = Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence (trial).

*significant, $p < .0045$; -nonsignificant, $p \geq .0045$.

imputation had a small effect on the key summary indices. In contrast, in the topiramate trial, where there was greater attrition in the topiramate group, treatment effects generally were smaller when data were imputed.

Secondary results: CPRA for subjects receiving both combined behavioral intervention and medical management

CPRA curves were also calculated, including only those subjects who received both MM and CBI (i.e., the other four pill-taking cells). Consistent with the main COMBINE article (Anton et al., 2006), across all cutoffs for each of the four outcomes, our results reveal nearly overlapping CPRA curves—with zero to very small treatment effects (figures not shown). For example, for HDDs, reduction in PHDDs, and DPD, Cohen's h ranged nonsignificantly from -0.09 to 0.11 across the cutoffs. For the RR in DPD, the range of Cohen's h widened slightly at the negative end, from -0.18 to 0.11, indicating that naltrexone actually did somewhat worse than the placebo for certain cutoffs, particularly those between a 65% and 80% RR in DPD ($h = -.18$ to $-.12$).

Discussion

We used CPRA to evaluate a variety of drinking outcomes from two large alcohol clinical trials: the COMBINE trial and the topiramate trial. CPRA allowed us to examine the effects of naltrexone and topiramate at all drinking outcome cutoffs, providing a more complete profile of their efficacy based on different definitions of a “successful treatment response.” Although CPRA is empirically driven and describes the full range of clinical trial data, like any post hoc analysis, replication with other data sets is crucial to understand how the CPRA curves differ or maintain similarities across trials. In a sense, this is no different than normal replication of clinical trials containing fewer outcomes, except that CPRA compares more outcomes across trials. With this caveat, results from our analyses generally confirm those of the primary reports (Anton et al., 2006; Johnson et al., 2007), demonstrating small and medium treatment effects for naltrexone and topiramate, respectively. However, the non-parallel nature of the responder curve trajectories from the two trials indicates that the magnitude of the observed treatment effect depends on the cutoff threshold for the drinking outcome, with greater treatment effects generally occurring at cutoffs associated with lower drinking levels (and greater reductions in drinking). We cannot determine from these data the effects of differences in study design, mechanism of action of the medications, or the robustness of the medication response. Thus, these data are only illustrative and are not meant to provide a direct comparison between medications.

CPRA is useful to determine the cutoff point or points where the treatment effect is maximized. This is relevant to

one of the major goals of the Alcohol Clinical Trials Initiative (ACTIVE), that is, to explore the points at which the maximal (and potentially statistically significant) effects on drinking reduction occur in alcoholism pharmacotherapy trials (Anton et al., 2012). Identifying the most sensitive endpoints to detect treatment effects is a crucial first step. Evaluating the clinical relevance of these endpoints is another key consideration for both treatment providers and regulatory agencies. Historically, the FDA has recommended responder endpoints with the lowest possible threshold of drinking, which previously was considered to be total abstinence and, more recently, “no heavy-drinking days” (FDA, 2006). Falk and colleagues (2010) reported modest differences in the treatment effect when responder definitions of no heavy drinking were compared with those that allowed up to a few days of heavy drinking. The present study extended this analysis by evaluating how treatment effects change when responder definitions allow for more drinking and by identifying the cutoff points for the maximal treatment effects.

Across the drinking measures assessed in the present study, maximal treatment effects were generally in the “small” range for naltrexone ($h = .25-.34$) and in the “medium” range for topiramate ($h = .43-.69$). These maximal treatment effect sizes are comparable to, if not slightly larger than, the effect sizes of traditional continuous outcomes analyzed as mean differences between treatment groups (Falk et al., 2010). Of interest, the CPRA graphs in the present study indicate that the maximal treatment effect generally did not occur at the cutoff associated with the least drinking (i.e., total abstinence and no HDDs). For example, in the COMBINE trial, using a cutoff of 0 DPD to define a responder (i.e., total abstinence), the treatment effect was $h = .13$ and not statistically significant, whereas the treatment effect was greater and statistically significant using a cutoff point of 3 or more DPD, reaching a maximum of $h = .25$ ($p = .0022$). Likewise, going from a cutoff of no HDDs (PSNHDDs) to one that allows up to 10 HDDs resulted in an increase in treatment effect from $h = .17$ (nonsignificant, $p = .039$) to $h = .28$ ($p = .0006$). In fact, in the COMBINE trial, an increase in treatment effect of .10–.21 was observed for all four of the outcomes when the definition of a responder allowed more than the lowest level of drinking for that outcome. This trend was also observed in the topiramate trial, except that, for DPD, total abstinence carried the largest treatment effect when the missing data were imputed ($h = .44$).

In addition to using total abstinence as a primary responder endpoint, we examined every cutoff from 0% to 100% reductions in alcohol consumption. The EMA recommends several secondary endpoints, including reductions in alcohol consumption of 50%, 70%, and 90% (EMA, 2010). We found that peak, statistically significant treatment effects occurred at high RRs in DPD (approximately 85% in the COMBINE trial; 60%–95% in the topiramate trial) and in

PHDDs (70%–92% in the COMBINE trial; 75%–80% in the topiramate trial). Thus, the effects of both naltrexone and topiramate appear to be greatest when responders are defined by relatively large reductions in drinking. Interestingly, the opposite pattern was observed in some analgesic clinical trials. For instance, when Farrar and colleagues (2006) used CPRA to assess pregabalin-induced pain reductions, the largest treatment effects were at cutoffs ranging from approximately 10% to 60%, with no difference between the treatment responder curves at cutoffs above 80%. Pregabalin had the greatest separation from placebo when positive responder definitions included subjects with relatively small pain reductions. CPRA analysis should be performed using data from other alcohol pharmacotherapy trials to corroborate these findings.

The effect of imputing missing drinking data was minimal in the COMBINE trial (few missing data and no group differences) and more pronounced in the topiramate trial (more missing data overall and more in the topiramate group than in the placebo group). Thus, like more traditional analytic approaches, the imputation of data in CPRA can produce results that differ from those when data are not imputed.

We propose two applications for CPRA curves: (a) in Phase 2 trials to inform sensitive and clinically meaningful outcome cut-points for dichotomous outcomes in Phase 3 trials, and (b) in Phase 3 trials to help inform treatment providers (e.g., in product labels) of the likelihood of treatment success given various definitions of a positive response. In Phase 2 trials, researchers might propose, *a priori*, that the primary outcome measure for a Phase 3 clinical trial would be based on a cutoff obtained from a Phase 2 trial. Although the cutoff that maximizes the treatment effect in a Phase 2 trial could help to ensure adequate statistical power, the ultimate selection of a Phase 3 outcome should also take into account regulatory requirements, safety, and clinical utility. For example, the responder endpoint PSNHDDs is based on the threshold of 4+/5+ drinks per drinking day. Even relatively infrequent violations of this endpoint (e.g., <1 HDD per month) have been associated with an increased risk of alcohol use disorders, tobacco use, and nicotine dependence (Dawson et al., 2008). With a greater number of days drinking or amount drunk per day, the risk of these adverse outcomes increases monotonically. Precise cutoffs for commonly used drinking outcomes (e.g., HDDs and DPD) have seldom been validated against important outcomes, such as health status, alcohol-related consequences, or treatment utilization and costs. In a previous study, when using the COMBINE trial data to correlate the number of HDDs with alcohol-related consequences during the last 2 months of treatment, the number of consequences increased significantly between 0 and 1 HDD but stayed at a relatively low level up to 6 HDDs before rising thereafter (Falk et al., 2010). Whether RR in HDDs or DPD might afford improved

estimates of consequence change compared with only using drinking levels within the treatment period is not clear. The answer to these questions could have significant clinical and public health consequences.

Not validating the choice of a cutoff produced by CPRA against clinically meaningful measures could have unintended, untoward results. For example, in the present study, the treatment effect in the COMBINE trial was maximized and statistically significant when we allowed up to 10 HDDs in an 84-day period. However, this outcome may not necessarily be a desirable clinical goal because drinking heavily on up to 10 days could cause substantially greater adverse consequences and costs than acceptable lower drinking outcomes. Moreover, this level of heavy drinking may be more detrimental to some individuals than others, such as in those with severe alcohol dependence or comorbidities. On the other hand, compared with relatively high pretreatment drinking, this level of drinking during treatment may represent a substantial decrease that might reduce morbidity and mortality, as well as other social consequences. Thus, although CPRA results are a valuable first step in narrowing the myriad potential responder cutoffs, the best treatment outcome is the one that maximizes the treatment effect while also minimizing potential adverse alcohol-related consequences and costs. These adverse consequences might be valued differently in primary care practices (where a significant reduction in drinking might be acceptable), substance use disorders clinics (where abstinence might be the only goal), or in other settings. CPRA allows each situation to be assessed and informed in a more comprehensive manner. Nevertheless, future studies should address the clinically relevant consequences and costs associated with the use of different responder cutoff points and consider important moderating risk factors, such as disease severity and treatment history.

Finally, providing CPRA curves in Phase 3 trials and product labels can also help to inform treatment providers as to the likelihood that a treatment response could reasonably be attributed to the medication. For instance, a clinician could check the CPRA curves to see whether a patient's response to 3 months of naltrexone therapy was consistent with a medication effect. If the patient was found to have reduced the number of DPD by 20%, the treatment provider could see that there was an equivalent chance of achieving the expressed goal with placebo (89.7%) as with naltrexone (87.2%) and might deem treatment unsuccessful. However, in the case of a patient who reduced the number of DPD by at least 85%, the treatment provider could infer that the response was likely attributable to the naltrexone treatment (i.e., 61.6% vs. 45.6% for placebo; $h = .32$, $p = .00008$) and imply a successful outcome, especially if related to other clinical improvements. This evidence-based approach could be used to inform the physician's discussion with the patient of the likelihood of

success with different treatment options and potentially enhance the patient's continued success.

In sum, CPRA appears to be an informative and useful way to present clinical trial results. We recommend that alcohol pharmacotherapy clinical trials consider the utility of CPRA as a supplement to traditional statistical analyses such that a more complete picture of the efficacy of a study medication might be attained.

Acknowledgments

In addition to the authors, the following individuals are or were members of the ACTIVE workgroup and provided intellectual input into this article during attendance at workgroup meetings: Earle Bain, M.D., Abbott Laboratories; Carla Canuso, M.D., Johnson & Johnson Pharmaceutical Research and Development, LLC; Marc de Somer, M.D., Alkermes; Jay Graham, Pharm.D., GlaxoSmithKline; Thomas Kosten, M.D., Baylor College of Medicine; Karl Mann, M.D., Central Institute of Mental Health, Mannheim, Germany; David McCann, Ph.D., National Institute on Drug Abuse; Didier Meulien, M.D., Lundbeck; Roger Meyer, M.D., Best Practice Management; Charles O'Brien, M.D., Ph.D., University of Pennsylvania; Stephanie O'Malley, Ph.D., Yale University School of Medicine; Joseph Palumbo, M.D., formerly of Johnson & Johnson Pharmaceutical Research & Development, LLC; Thomas Permutt, Ph.D., FDA; Beatrice Rendenbach-Mueller, Ph.D., Abbott Laboratories; Rebecca Robinson, M.S., Eli Lilly & Company; Bernard Silverman, M.D., Alkermes; Lars Torup, Ph.D., Lundbeck; Susan VanMeter, M.D., GlaxoSmithKline; Celia Winchell, M.D., FDA; and Conrad Wong, Ph.D., Eli Lilly. Sarah Timm (American College of Neuropsychopharmacology) has provided important administrative support to the ACTIVE workgroup. The ACTIVE workgroup expresses sincere appreciation to Leah R. Zindel, R.Ph., M.A.L.S., for her editing of the article. The ACTIVE workgroup also acknowledges the contribution of Dr. Vivek Kusumakar who was instrumental in establishing the initial ACTIVE effort but whose untimely death prevented his invaluable contribution to be fully realized.

References

- Allen, J. P. (2003). Measuring outcome in interventions for alcohol dependence and problem drinking: Executive summary of a conference sponsored by the National Institute on Alcohol Abuse and Alcoholism. *Alcoholism: Clinical and Experimental Research*, 27, 1657–1660.
- Anton, R. F., Litten, R. Z., Falk, D. E., Palumbo, J. M., Bartus, R. T., Robinson, R. L., . . . the ACTIVE Workgroup (2012). The Alcohol Clinical Trials Initiative (ACTIVE): Purpose and goals for assessing important and salient issues for medications development in alcohol use disorders. *Neuropsychopharmacology*, 37, 402–411.
- Anton, R. F., O'Malley, S. S., Ciraulo, D. A., Cisler, R. A., Couper, D., Donovan, D. M., . . . Zweben, A., & the COMBINE Study Research Group. (2006). Combined pharmacotherapies and behavioral interventions for alcohol dependence: The COMBINE study: A randomized controlled trial. *Journal of the American Medical Association*, 295, 2003–2017.
- Breslow, R. A., & Graubard, B. I. (2008). Prospective study of alcohol consumption in the United States: Quantity, frequency, and cause-specific mortality. *Alcoholism: Clinical and Experimental Research*, 32, 513–521.
- Cappelleri, J. C., Bushmakina, A. G., Zlateva, G., & Sadosky, A. (2009). Pain responder analysis: Use of area under the curve to enhance interpretation of clinical trial results. *Pain Practice*, 9, 348–353.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Dawson, D. A., Li, T. K., & Grant, B. F. (2008). A prospective study of risk drinking: At risk for what? *Drug and Alcohol Dependence*, 95, 62–72.
- European Medicines Agency. (2010). *Guideline on the development of medicinal products for the treatment of alcohol dependence*. London, England: Author. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/03/WC500074898.pdf
- Falk, D. F., Wang, X. Q., Liu, L., Fertig, J., Mattson, M., Ryan, M., . . . Litten, R. Z. (2010). Percentage of subjects with no heavy drinking days: Evaluation as an efficacy endpoint for alcohol clinical trials. *Alcoholism: Clinical and Experimental Research*, 34, 2022–2034.
- Farrar, J. T. (2008). Author's response [Letter to the editor]. *Journal of Pain Symptom Management*, 36(2), e7.
- Farrar, J. T., Dworkin, R. H., & Max, M. B. (2006). Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: Making clinical trial data more understandable. *Journal of Pain and Symptom Management*, 31, 369–377.
- Food and Drug Administration. (2006). *Medical Review of Vivitrol*, 21–897. Rockville, MD: Author.
- Good, P. I. (2008). Re: Cumulative proportion of responders [Letter to the editor]. *Journal of Pain Symptom Management*, 36(2), e7.
- Greenfield, T. M., Bond, J., Kerr, W. C., Nayak, M. B., Anton, R. F., Kranzler, H. R., & Kaskutas, L. A. (2005, May). *Risks of alcohol use disorders for treated and concerned individuals are higher than for others at given patterns of drinking*. Presented at the Annual Epidemiological Symposium of the Kettil Bruun Society for the Epidemiological and Social Research on Alcohol, Riverside, CA.
- Jackson, K. M. (2008). Heavy episodic drinking: Determining the predictive utility of five or more drinks. *Psychology of Addictive Behaviors*, 22, 68–77.
- Johnson, B. A., Rosenthal, N., Capece, J. A., Wiegand, F., Mao, L., Beyers, K., . . . Swift, R. M., the Topiramate for Alcoholism Advisory Board, & the Topiramate for Alcoholism Study Group. (2007). Topiramate for treating alcohol dependence: A randomized controlled trial. *Journal of the American Medical Association*, 298, 1641–1651.
- Kim, Y. (2011). Missing data handling in chronic pain trials. *Journal of Biopharmaceutical Statistics*, 21, 311–325.
- Litten, R. Z., Bradley, A. M., & Moss, H. B. (2010). Alcohol biomarkers in applied settings: Recent advances and future research opportunities. *Alcoholism: Clinical and Experimental Research*, 34, 955–967.
- Lowman, C., Hunt, W. A., Litten, R. Z., & Drummond, D. C. (2000). Research perspectives on alcohol craving: An overview. *Addiction*, 95, Supplement 2, S45–S54.
- Miller, W. R. (1996). *Form 90: A Structured Assessment Interview for Drinking and Related Behaviors* (Project MATCH Monograph Series, Vol. 5, DHHS Pub. No. [NIH] 96–4044). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Miller, W. R., Heather, N., & Hall, W. (1991). Calculating standard drink units: International comparisons. *British Journal of Addiction*, 86, 43–47.
- Miller, W. R., Tonigan, J. S., & Longabaugh, R. (1995). *The Drinker Inventory of Consequences (DrInC): An instrument for assessing adverse consequences of alcohol abuse* (Test Manual). (Project MATCH Monograph Series, Vol. 4, DHSSS Pub. No. 95–3911), Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Moore, R. A., Moore, O. A., Derry, S., & McQuay, H. J. (2008). Numbers needed to treat calculated from responder rates give a better indication of efficacy in osteoarthritis trials than mean pain scores. *Arthritis Research & Therapy*, 10, R39. Retrieved from <http://arthritis-research.com/content/10/2/R39>
- Moore, R. A., Moore, O. A., Derry, S., Peloso, P. M., Gammaitoni, A. R., & Wang, H. (2010). Responder analysis for pain relief and numbers needed to treat in a meta-analysis of etoricoxib osteoarthritis trials: Bridging a gap between clinical trials and clinical practice. *Annals of the Rheumatic Diseases*, 69, 374–379.
- Moore, R. A., Straube, S., Paine, J., Derry, S., & McQuay, H. J. (2011). Minimum efficacy criteria for comparisons between treatments using individual patient meta-analysis of acute pain trials: Examples of etori-

- coxib, paracetamol, ibuprofen, and ibuprofen/paracetamol combinations after third molar extraction. *Pain*, 152, 982–989.
- National Institute on Alcohol Abuse and Alcoholism. (2007). *Helping patients who drink too much: A clinician's guide, updated 2005 edition*. Bethesda, MD: Author.
- O'Malley, S. S., Garbutt, J. C., Gastfriend, D. R., Dong, Q., & Kranzler, H. R. (2007). Efficacy of extended-release naltrexone in alcohol-dependent patients who are abstinent before treatment. *Journal of Clinical Psychopharmacology*, 27, 507–512.
- O'Malley, S. S., Rounsaville, B. J., Farren, C., Namkoong, K., Wu, R., Robinson, J., & O'Connor, P. G. (2003). Initial and maintenance naltrexone for alcohol dependence using primary care vs. specialty care: A nested sequence of 3 randomized trials. *Archives of Internal Medicine*, 163, 1695–1704.
- Physicians' Desk Reference*. (2011a). Cymbalta (duloxetine hydrochloride). Retrieved from <http://www.pdr.net/drug-information/cymbalta?druglabelid=288>
- Physicians' Desk Reference*. (2011b). Lyrica (pregabalin). Retrieved from <http://www.pdr.net/drug-information/lyrica?druglabelid=467>
- Pritchett, Y.L., McCarberg, B.H., Watkin, J.G., & Robinson, M.J. (2007). Duloxetine for the management of diabetic peripheral neuropathic pain: Response profile. *Pain Medicine*, 8, 397–409.
- SAS. (2010). The UNIVARIATE procedure (Chapter 4). In *Base SAS 9.2 procedures guide: Statistical procedures, third edition*. Cary, NC: SAS Publishing.
- Sobell, L.C., & Sobell, M.B. (1992). Timeline followback: A technique for assessing self-reported ethanol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biological methods* (pp. 41–72). Totowa, NJ: Humana Press.
- Straube, S., Derry, S., Moore, R.A., Paine, J., & McQuay, H.J. (2010). Pregabalin in fibromyalgia—Responder analysis from individual patient data. *BMC Musculoskeletal Disorders* 11, 150. Retrieved from <http://www.biomedcentral.com/1471-2474/11/150>