

Assessing Onset of Treatment Benefit in Depression and Anxiety: Conceptual Considerations

Eugene M. Laska, PhD; Craig H. Mallinckrodt, PhD;
Jim C. Mundt, PhD; Paul Leber, MD; Anthony L. Vaccarino, PhD;
Amir H. Kalali, MD; and John H. Greist, MD

Objective: Methods for characterizing the onset of treatment benefit in major depressive disorder and generalized anxiety disorder have been studied for some time, yet there is no universal agreement as to the best approaches. Our purpose is to summarize the conceptual framework underlying modern methods for characterizing onset and detailed approaches for which there is consensus from the perspective of a clinician, clinical researcher, and statistician. Possible alternatives to unresolved issues are discussed.

Participants: There were 17 experts from academia, the pharmaceutical industry, and the US Food and Drug Administration who met on April 19, 2007, to consider the issues. Many others from sponsoring firms observed the proceedings.

Evidence: A series of papers was presented at a consensus meeting and, after discussions, a sense of the participants was obtained. A small group subsequently reviewed the material and articles from the literature and prepared this article, which was reviewed by all of the participants.

Conclusions: The elements that form the basis for describing onset of treatment benefit include defining a clinical event or measurable threshold that validly signals that a treatment has begun to provide clinically meaningful and sustained improvement and utilizing methods for estimating the probability of crossing the onset threshold, the distribution of time to onset for those who do cross, and when to alter or change interventions if the treatment is unsuccessful.

J Clin Psychiatry 2009;70(8):1138–1145

© Copyright 2009 Physicians Postgraduate Press, Inc.

Submitted: February 11, 2009; accepted April 21, 2009
(doi:10.4088/JCP.09cs05129).

Corresponding author: Eugene M. Laska, PhD, Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd, Orangeburg, NY 10962 (laska@nki.rfmh.org).

Onset or time to response is defined as the time at which a treatment first begins to provide a *meaningful* therapeutic benefit to the patient to whom it is administered. It is self-evident that among a set of otherwise therapeutically fungible interventions, those with shorter average times of onset are preferable. However, the complexity of characterizing onset makes this simple proposition conceptually and practically difficult to operationalize. This document is meant to help identify the issues and promote greater understanding of current approaches.

Patients with major depressive disorder (MDD) or generalized anxiety disorder (GAD) may wait 2 or more weeks before a meaningful clinical response is first observed after medication is first administered.^{1–3} Moreover, not all patients given a course of treatment respond. At what point should the treatment of a patient who has not yet exhibited a meaningful response be changed? Definitive evidence bearing on this question, of course, is yet to be obtained. Nevertheless, many believe that with currently available treatments a patient who has not responded by the sixth or eighth week is unlikely to obtain subsequent benefit from further continuation of the medication.⁴

Because of the morbidity and risk of suicide associated with MDD and GAD, there is high motivation for developing new treatments that have a more rapid onset of action than currently marketed products.^{2,5–12} Because of the consequential regulatory uncertainty, a considerable impediment to progress is the absence of clarity within the scientific community as to how best to characterize and compare the onset of treatments for these disorders.

To date, the assessment of onset of action of treatments for MDD and GAD has generally been a secondary consideration derived retrospectively during the analyses of clinical trials.^{13–15} Although these reports have some value, fully creditable estimates and treatment comparisons of time to response must be based on prospectively designed, controlled, randomized clinical trials; a prespecified definition of onset; and an appropriate statistical analysis plan.¹⁶ Unfortunately, there have been only a few studies meeting these criteria.^{17–20} However, many articles have commented on the methodological issues attending the design and analysis of such trials.^{9,16,21–26}

To consider the clinical and scientific issues attending the assessment of onset, a consensus development conference

was organized by the International Society for CNS Drug Development (ISCDD), with experts from academia, the pharmaceutical industry, and regulatory agencies participating. The Onset Consensus Conference was held April 19, 2007, in Washington, District of Columbia. Participants' charge was to consider perspectives on methods for appraising onset of treatment benefit in MDD and GAD, including how to

- define measurable threshold events that would validly signal that a treatment has begun to provide clinically meaningful and sustained improvement;
- describe the onset properties of a treatment in terms that are relevant for clinical decision making; and
- perform statistical analyses that characterize onset properties and enable their comparison among different treatments.

Following the meeting, a work group was formed to review the presentations and discussions and to summarize the deliberations. All participants were subsequently invited to comment on the document, and the present article is the result of that effort.

Specific Criteria for Defining Onset

Onset is the time at which a patient treated with a drug begins to experience a meaningful clinical benefit. Not surprisingly, agreement was not reached at the conference on specific criteria necessary to define such an event. However, this does not undermine the importance of the consensus that was reached about what may and may not be considered valid approaches to characterizing onset. In fact, on this latter point, there was considerable consensus; participants generally agreed that as long as a reasonable definition of response is unambiguously defined in the study protocol, the data generated is of value, provided, of course, that the methods employed to characterize the time to response were valid.

An Invalid Approach

An important consensus reached at the conference concerns the invalidity of equating the time at which a drug's meaningful therapeutic effect begins with the first time the magnitude of the mean difference from placebo is statistically significant.

The efficacy of new drugs intended for the treatment of MDD and GAD are ordinarily assessed in short-term (4 to 8 weeks), double-blind, randomized controlled clinical trials. The group mean scores of patients assigned to the investigational drug treatment and a suitable control, often placebo, are compared on a protocol-specified primary measure of outcome (eg, change from baseline on a summary score of a multi-item rating scale) under the terms of a statistical analysis plan devised in advance of examination

of the data.²⁷ Statistical significance of group mean differences, of course, can also be determined at any of the time points at which observations are made in the study. This approach follows the general paradigm for analyzing a clinical trial. The treatment group distributions of an outcome measure, which is individually determined for every subject in the trial, are compared. Statistical significance is used to quantify the possibility that the observed differences were merely due to chance.

It is not uncommon for onset to be erroneously understood to be the time at which the difference between the mean scores of the drug and placebo first attain statistical significance. This interpretation, however, is neither warranted nor valid. Indeed, depending upon the response criteria, between-group differences could be statistically significant at each and every scheduled assessment time even if no patient ever attains a clinically meaningful response or if all patients in both groups meet onset criteria at the time of the first observation. The fundamental flaw in using the time of the first observed statistical significance as a measure of onset is that the contrasts are made in the effect domain rather than the time domain; and factors such as sample size and variance, which are unrelated to the intrinsic intensity of a drug's therapeutic activity, strongly influence the determination of statistical significance. As a result, a drug's onset could be incorrectly determined to be very rapid with the expenditure of sufficient funds to purchase a large sample size.²⁸

DETERMINING WHEN A PATIENT'S ONSET HAS OCCURRED

Setting the Criteria

It is no easy task to determine the time at which a meaningful improvement in the intensity of signs and symptoms, functional ability, or quality of life has commenced. Mallinckrodt et al²⁹ reviewed various approaches for assessing onset and made several specific recommendations. However, broad acceptance in the clinical research community of a definition of a clinically meaningful event has not yet occurred. Most studies have defined a *patient-specific onset* as the time of the first scheduled observation at which a predefined, sustained difference from baseline, in absolute or percentage change, in a rating scale score is achieved.³⁰ Two criteria that have been used are an absolute change of 3 points on the Hamilton Depression Rating Scale (HDRS) and 4 points on the Montgomery-Asberg Depression Rating Scale. Such definitions operationalize the determination of onset and introduce an apparent degree of rigor in the sense that comparisons between treatments are subject to the same process. In this sense, treatment contrasts are likely to be legitimate, but point estimates of population parameters, such as the median time to onset, are valid only to the extent that the criteria are appropriate. It is necessary to recognize that the chosen thresholds are inherently arbitrary,

and, whatever they are, there may be clinical circumstances in which a patient's status is mischaracterized.

Ideally, the patient, perhaps in partnership with a clinician, should be the one to decide that a clinically meaningful change has occurred. Assessment of anxiety and depression severity in clinical trials has traditionally relied upon clinical interviewing skills and the judgments of trained raters. Patient-reported outcomes, which are finding increasing use in randomized controlled trials (RCTs), may help to advance and illuminate the essential elements of improvement and foster enhanced temporal resolution of meaningful change. Such measures are now gaining acceptance by the scientific community and regulatory agencies as valid assessments of outcome.^{31,32}

Single or Multivariate Onset Criteria

Mood and anxiety disorders are multidimensional constructs composed of many symptoms. Standard practice is to declare that a treatment is effective if an efficacy assessment at the end of the trial statistically separates it from placebo. The usual measure of outcome is a composite score derived from a rating scale whose purpose is to integrate the full spectrum of measured symptomatology. In contrast, onset determination is based on appraisals during the early phase of a trial, at which time only some of the symptoms may have improved.^{33,34} This raises the question of whether to determine onset globally or for individual symptom clusters. Use of a composite rating scale score is justified in part because it is consistent with end-of-study criteria. But it can also lead to questionable conclusions. For example, it has been suggested that purportedly early responses to some antidepressant medications may be an artifact of the fact that sedation, a side effect, is included in some well-known rating scales. Although transitory effects, such as improvements in sleep, may not alter core depressive symptoms, they do affect total scores, which may lead to biased estimates of onset.³⁵ There may be advantages to disassembling composite measures to better characterize the full time line of the onset of benefits.^{19,34} Estimating onset of benefit for outcomes such as weight change, menstrual symptoms, or sexual function may not be possible in the usual time frame of current clinical trials.

Duration of a Sustained Clinically Meaningful Event

If a patient achieves the criteria for onset of benefit, but subsequently deteriorates, has onset occurred? One view is that once having achieved the criteria, onset has occurred. This is the approach used in studying analgesics, where the patient is deemed to be the ultimate judge. Onset is registered when the patient clicks a stopwatch to record the time he or she subjectively determines that pain relief is meaningful.³

The inherent day-to-day variability of clinical manifestations in chronic conditions such as MDD and GAD implies that evidence from a single point in time is insufficient.

The response that nominally meets threshold criteria for a meaningful response must be sustained. How long the positive response must continue before onset can be declared is legitimately debatable. Theoretically, the required evidence ranges from a single crossing of an improvement threshold to sustained improvement for the remainder of the trial. Nierenberg et al¹¹ used the criteria for onset of response in depression as a 30% decrease from baseline in the total score on the HDRS followed by a 50% or greater decrease by week 8. A subject who crosses the latter threshold is commonly termed a responder. The justification for this definition is that onset must lead to at least 50% response or else a true therapeutic response did not commence. While there is no final word about the length of the minimum duration of meaningful improvement required to confirm that onset has occurred, it is clear that a single threshold crossing is inadequate.

When to Measure Clinical Status to Appraise Onset

In measuring onset of analgesic response, Laska and Siegel³ proposed a now widely used "stopwatch method" that asks patients to stop a watch when meaningful pain relief is first experienced. Neither complete relief nor, indeed, continuing relief is required. The threshold event is direct, has face validity and, notably, is determined by the patient. Perforce, different patients may have different perceptions on what amount of reduced pain intensity constitutes meaningful relief. However, since the purpose of administering the analgesic is to provide pain relief, it is meaningful relief if and only if the patient feels it to be so. While this approach has many appealing properties and is instructive, for MDD and GAD, where onset is measured in days or weeks, measurement tolerances of 1 or 2 days is probably adequate. Such precision may be achieved through the use of daily diaries, frequent telephone interviews, or clinic visits, which, it is generally believed, produce similar results. A series of scheduled postbaseline observations, with longer between-evaluation times, yield interval-censored observations of onset. The first threshold crossing that is subsequently sustained a sufficient number of times provides a time at which onset has already occurred, and the previous subthreshold observation provides a time at which onset has not yet occurred. Temporal resolution of diurnal, weekly, or longer cyclical variations is limited by the Nyquist frequency sampling distributions. Consequently, the estimate of a subject's actual onset can only be the time corresponding to the midpoint of the 2 observations. This approach is consistent with current practice in outpatient clinical trials, where weekly or biweekly study visits are the norm, particularly during the early phase of treatment.

The precision of estimates of the distribution of onset using prespecified assessment times depends on the chosen times of observation.³ While increasing the number of patient/rater interactions might provide finer temporal resolution, visits that are too frequent might also increase

placebo response because of enhanced nonspecific therapeutic interaction and negatively influence recruitment and patient retention rates. Further, the chance of unblinding raters might well increase, as might the costs of studies.²⁶

Electronic patient-reported outcomes may provide an alternative approach that somewhat resembles the stopwatch approach. They may permit more frequent assessments without incurring the problems associated with increased clinic visits.

Data Representation and Missing Values

The data on onset is the time from the start of the observation period until the clinical threshold event is observed or, if it has not occurred and there are no further observations to be obtained, the time of censoring. The censoring time of a nonresponder will often coincide with the end of the study or the time a subject leaves the trial. In current practice, ratings are assessed in a repeated-measures framework with many observations per subject. At each time point, the value of the rating scale can be examined and a binary indicator variable set to 1 if the threshold is crossed. *Onset for a patient* is defined as the first time at which the threshold is crossed and required number of subsequent binary indicators is 1.

One of the most vexing problems in clinical trials is what to do about missing observations. It is no less an issue when studying onset. If the study protocol requires that the threshold condition continues to hold for the remainder of the trial, how should an individual with an unbroken string of 1s whose last observation is missing be treated? How should an individual with zero indicators who is lost to follow-up after a few weeks be treated? These issues need to be considered and addressed in the statistical analysis.

DESCRIBING THE ONSET PROPERTIES OF A TREATMENT IN TERMS THAT ARE RELEVANT FOR MAKING CLINICAL DECISIONS

In order for a clinician to assess whether and how to use a treatment in caring for a patient, many of its properties need to be understood. Among these are the short- and long-term beneficial effects of the drug as well as its side-effect profile. As every clinician knows, when different treatment options are contemplated, it may be necessary to consider tradeoffs among these properties to best meet the particular needs of the patient. Here, we focus on only those relating to onset.

Clinical Considerations

Patients and clinicians should know

- What is the chance that a patient will achieve the desired threshold event that defines the response to treatment?
- For patients who experience the event, what is the distribution of time to onset?

- If a patient has not experienced onset after t days, what is the chance that onset will occur in the future?

In considering the onset properties of a treatment, 2 separate concepts must be considered—the chance that a drug will work (the response rate) and the time that it takes to work (the speed of onset). To illustrate, suppose 1 treatment has slow onset, but the probability that a patient will experience the threshold event is very high. A second treatment has a relatively rapid onset, but there is only a relatively small probability that a patient will cross the threshold. Among those experiencing onset with the second of these 2 treatments, most have achieved it by the second week; of those who have not, the chance of having onset thereafter is very small. In the next section, the formula to compute this last probability is presented. The clinician can use this information in the second week of treatment to decide whether to alter the dose or even to change therapies. These examples make clear the clinical need to know both the response rate and the speed of onset for those who respond.

Choice of Control Agents

An important element in the design of an RCT is the choice of control treatments. Comparisons with an active treatment without a placebo control leaves open the question of whether either treatment was an effective agent in the study. This reinforces the need for the study design to also include measurements later in the course of treatment and to include placebo as a control so that the validity of the onset comparisons can be established. Comparisons of the onset properties of active treatments with placebo usually lead to superior rates of response but, not infrequently, to undifferentiated conditional speed of response. As Quitkin et al³⁶ observed, when placebo works, it works quickly. The choice of which active comparator and dose to study needs careful consideration. If the overall effectiveness of the test and active control treatments overlap across their respective dose ranges, they should, if possible, be compared at equally effective doses. Alternatively, the comparison may be at the fixed doses believed to be therapeutically optimal for each agent. The relative overall effectiveness of the test and control treatments can be determined and the results taken into account when interpreting the onset contrasts.

A Model for the Survival Distribution of Onset

The examples above of two treatments—one with fast onset but low probability of response and one with slow onset but high probability of response—provide insight on how to formulate a probability model that corresponds to the clinical situation. In particular, the survival distribution of time to onset of treatment benefit in the whole population, which includes those who never achieve onset, is not the relevant function. The focus should be on the probability distribution of onset among those who do achieve

the threshold event. Patients who do not achieve onset contribute information about the rate of response, but they provide no information on the speed of response. Indeed, 2 treatments with different onset probabilities but identical onset distributions among those who obtain onset will have apparent differences in the onset distribution in the whole population that are entirely due to differences in the rate of response. A drug with the higher probability of onset will appear faster merely because more patients respond.³ It is not uncommon in the diagnoses considered here for more than half of the patients participating in a clinical trial to experience an inadequate response or no response at all,³⁶ even with drugs that have been determined to be effective by the US Food and Drug Administration. Including non-responders in the survival distribution of onset will lead to misperceptions about, for example, the median time to onset, which, in turn, can lead to erroneous judgments about how long a patient should be treated before concluding that a change in therapy is warranted. The confounding of the distribution of rapidity of action and the response rate is avoided by focusing on the conditional onset distribution, ie, on the distribution of time to onset among those patients who had onset of benefit.

The technical means to obtain valid estimates of the desired quantities is to assume that the distribution of onset follows a cure model. That is, the survival distribution has the form

$$H(t) = 1 - p + pS(T > t | \text{onset}).$$

$H(t)$ is the survival distribution of onset in the entire population, and it includes the possibility that onset will never occur. The quantity p is the response rate, the probability that a patient will respond. The last term on the right side of the equation, $S(T > t | \text{onset})$, is shortened to $S(t)$, and it is the conditional survival distribution of onset, which is the chance that onset will occur after time t , given that onset will occur. Under this model, the elements that need to be conveyed to clinicians may be estimated, and hypotheses about the individual components, rate and speed, can be tested.

An important consequence of this model is that it leads to a formal quantification of $Q(t)$, the probability that a patient who has not had onset by time t will obtain it in the future. From the statistical model of the survival distribution $H(t)$, we have

$$Q(t) = \frac{pS(t)}{1 - p + pS(t)}.$$

Since $S(t)$ is a decreasing function of time, the probability $Q(t)$ decreases as t increases and, eventually, becomes zero.

Comparing Treatments in Terms of Odds Ratios

The odds of obtaining onset after being randomly assigned to treatment A is $p_A/(1 - p_A)$, the probability of obtaining

onset divided by the probability of not obtaining onset. The relative odds of obtaining onset after receiving treatment A compared to treatment B, called the *odds ratio*, is

$$\frac{p_A(1 - p_B)}{(1 - p_A)p_B}.$$

The odds ratio at time t of obtaining onset in the future is

$$\frac{Q_A(t)[1 - Q_B(t)]}{[1 - Q_A(t)]Q_B(t)},$$

where $Q(t)$, defined above, is the probability of achieving onset at some time after t , given that it has not yet occurred. If the conditional survival distribution of treatments A and B are identical but their respective probabilities of onset differ, then for all values of t , the odds ratio is the odds ratio of obtaining onset after receiving treatment A compared with treatment B given above.

STATISTICAL ANALYSES

Statistical Analyses Based on Survival Methods

The time of onset is a survival random variable for which many statistical methods are available. If the event does not occur during the period of observation or if the patient is lost to follow-up, the observation is said to be right censored. If it is known that the event occurred in an interval, the observation is said to be interval censored. It is generally assumed that censoring times are independent of the event times. There are 3 distinct approaches used in survival analyses differing in the degree to which underlying assumptions are required: nonparametric, semiparametric, and parametric.

The most widely used nonparametric approach for estimating the survival function, defined as $H(t) = P(T > t)$, the probability that onset T occurs at time t or later, is the Kaplan-Meier product-limit estimator. Rank tests, such as the Wilcoxon or the log-rank test, are used to test for differences between the treatments' survival distributions. In the semiparametric framework, particularly when it is necessary to account for covariates, Cox proportional hazards regression is widely used. The name of the method celebrates its inventor and highlights the assumption that the groups being compared have hazard functions that are proportional. In parametric analyses, the shape of the survival distribution is assumed known and the data are used to estimate parameters that complete its specification. One common assumption is that the survival random variable follows a Weibull distribution. These methods are well-known in the clinical trials literature, are designed to accommodate data in which there are censored observations, and are easy to implement with commonly available software.

These methods all assume that every censored subject will eventually experience the event. For characterizing onset, clinical experience teaches that only a proportion, p , will experience the event. It is desired to estimate both p and the conditional survival distribution of those who do as described above in a cure model. Utilizing the fraction of subjects who obtain onset as an estimate of p leads to a biased estimate; subjects who were censored could well have gone on to cross the onset threshold. Similarly, it is also tempting to estimate the conditional onset distribution by utilizing only the subset of patients who had onset. This approach, too, is flawed because the censored observations must be taken into account.

Laska and colleagues¹ were the first to propose a cure model approach in the context of onset. Earlier, Berkson and Gage³⁷ had introduced the cure model to analyze time to death in cancer clinical trials, in which a proportion of the patients are cured. Since then, a large body of literature describing new methods and applications has emerged. Laska and Meisner¹⁰ gave a nonparametric generalized maximum likelihood estimator of p and of the conditional survival distribution, given the patient achieves onset under the cure model. Their estimator is based on the Kaplan-Meier estimator. They also gave a nonparametric test to compare p across treatments whose power compares well with its nonparametric alternatives. Orazem³⁸ proposed a nonparametric test to compare the conditional survival distributions of treatments.

Semiparametric approaches have been developed and perhaps the most widely used is the Cox proportional hazards mixture cure model. Kuk and Chen,³⁹ Peng and Dear,⁴⁰ Sy and Taylor,⁴¹ and Lam et al⁴² investigated methods to estimate model parameters. Large sample properties of estimators from the proportional hazards mixture cure model were investigated by Fang et al.⁴³ One widely used parametric cure model that was introduced by Farewell⁴⁴ is based on a logistic function to model p and a Weibull distribution to model $S(t)$. Both components allow the use of covariates to capitalize on prognostic patient characteristics. Koti⁴⁵ replaced the Weibull with the generalized γ distribution, which offers a broader range of shapes for the survival function.⁴⁰

As in the development of mixed models in the analysis of variance framework, it has been recognized in survival analysis that there may be considerable variation that cannot be explained by observed covariates. Failing to account for such heterogeneity is inefficient and may lead to invalid estimates. A frailty term is therefore included in the latency distribution to account for unobserved variation in the mixture cure model. In the approach of Price and Manatunga,⁴⁶ the cure fraction follows a nonnegative random variable and is regarded as a frailty, while the usual cure fraction is modeled with a logistic regression. Their approach describes how to analyze a cure-rate model when the data are interval censored.

Corbière and Joly⁴⁷ have published a SAS macro that can be used to estimate parametric and semiparametric mixture cure models with covariates. The cure fraction can be modeled by various binary regression models. Parametric and semiparametric models can be used to model the survival of uncured individuals. The maximization of the likelihood function is performed using SAS PROC NLMIXED for parametric models and through an expectation maximization algorithm for the Cox's proportional hazards mixture cure model.

Threshold-Crossing Analysis Based on Mixed Models

As discussed above, an operational definition of onset requires specification of a minimum sustained duration of effect evidenced by a minimum number of consecutive evaluations that meet the threshold demand. In this section we consider a special case in which a single crossing is conceptually sufficient to define onset and censored observations correspond to individuals who would never respond. Indeed, a subject may have crossed the threshold at 1 observation time, not at the next, and experience it again later. This is the reason it is called a threshold-crossing analysis.

In this special circumstance, the proportion of subjects who have crossed the threshold is an estimate of the unconditional survival distribution of onset at that time. A logistic regression model allows for inclusion of covariates and has a logical extension into the multivariate repeated-measures framework. The motivation for this framework is to use all the data to study the joint distribution of threshold crossings at all of the fixed observation times. This model can be used to test hypotheses about the main effect of threshold crossings of treatments and/or interaction effects and difference in slopes over time. These analyses have varying degrees of utility depending on the temporal profiles.

Multivariate analyses are more complex and more difficult to implement than simple logistic regression, but standard software is available to implement them in either a generalized estimating equation framework or a likelihood-based framework. This approach was utilized as the primary analysis in a clinical trial whose primary objective was to compare onset of action for 2 antidepressants,¹⁹ and results from this and many of the other methods discussed in this section were compared.²⁵ However, it must be recalled that examining the unconditional distribution of onset at any point in time confounds the probability of response and the time to response among the responders. A more desirable analysis separates these 2 parameters.

Logistic regression for repeated measures from data with a traditional assessment schedule at weeks 1, 2, 4, 6, and 8 was shown to provide the expected control of type I error and power comparable to or greater than survival analyses of data from either a traditional assessment schedule or from a frequent assessment schedule of twice weekly observations for 8 weeks.²⁶

In limited situations, it is possible to utilize a repeated-measures logistic regression analysis in the conditional framework. For example, a binary variable that indicates whether a patient has experienced the event can be added to the analytic model along with its interaction with treatment and time. Treatment comparisons in the visitwise probabilities of threshold crossing are then made within the subgroup that had at least 1 crossing. It must be remarked that since the binary indicator is an actually observed outcome variable, there is controversy as to the legitimacy of its use as an explanatory variable. Further, the value of the variable is not available for subjects who leave the trial before they have achieved onset. Therefore, it is necessary to assume that everyone who would eventually have had a successful outcome (sustained remission) did so during the trial. Nevertheless, the approach has been used to compare 2 antidepressants conditioning on whether patients had achieved a sustained remission at the end of the 8-month clinical trial and then comparing probabilities of onset during the 8-week acute-treatment period.¹⁹ This approach too has had very limited application in clinical trials.

SUMMARY

There is as yet no universally accepted best approach for assessing the onset of treatment benefit of antidepressants and anxiolytics. However, there are substantial areas where there is consensus. We have attempted to summarize here the underlying conceptual principles. Operationalizing them in specific situations requires a critical appraisal of the appropriateness of various reasonable options. These include choice of a rating scale or scales, an approach for obtaining ratings, a schedule of observations, a threshold value that signals the occurrence of a clinically meaningful event, the length of time it must persist, the control treatments, statistical methods for estimating the relevant parameters, and a method for testing hypotheses to compare treatments.

The Consensus Conference concluded that the assessment of onset is best obtained from a randomized, placebo-controlled study with an appropriate active comparator and relatively frequent assessments early in the trial utilizing the same rating scales validated for establishing overall effectiveness.

There was general agreement that it is not now possible to subscribe to a specific universal criterion for defining the level of response required to be clinically meaningful. On the other hand, a variety of definitions have face validity, and, as long as they are well formulated and documented in advance, useful clinical information and valid inference can be obtained from an RCT. Commonly used, not unreasonable criteria are based on a percent reduction from baseline in the 20% to 30% or even greater range, with maintenance of at least that degree of improvement over a series of subsequent visits. Another reasonable approach

adds to the demand that, at study end, the reduction from baseline scores must be at least 50%.

Regardless of the criteria used to define a meaningful clinical response, there was general agreement that if data from such trials were analyzed using a cure model, the results would provide information that could usefully help to inform clinical decisions, including selection of the most appropriate drug and guidance on whether to continue or alter treatment strategy for a patient who has not yet responded.

Published reports of clinical trials that investigate onset should include estimates of the probability of onset and summary measures, such as the median, of the conditional survival distribution of time to response. Also, a plot of the probability of obtaining onset in the future if it had not yet occurred would provide valuable clinical information. Prognostic baseline variables that affect the estimated probability of onset, median time to onset, or the chance of onset in the future should be reported.

Author affiliations: New York University School of Medicine, New York, and Nathan S. Kline Institute for Psychiatric Research, Orangeburg (Dr Laska), New York; Corporate Center, Eli Lilly and Company, Indianapolis, Indiana (Dr Mallinckrodt); Healthcare Technology Systems, Inc. (Drs Mundt and Greist) and University of Wisconsin School of Medicine and Public Health (Dr Greist), Madison, Wisconsin; Neuro-Pharm Group, LLC, Potomac, Maryland (Dr Leber); Ontario Cancer Biomarker Network, Toronto, Canada (Dr Vaccarino); and Quintiles Inc., San Diego, California (Dr Kalali).

Onset Consensus Conference participants: Joseph Cappelleri, PhD (Pfizer); David DeBrot, MD (Eli Lilly); Douglas Feltner, MD (Pfizer); John H. Greist, MD (Healthcare Technology Systems); Amir H. Kalali, MD (Quintiles, University of California, San Diego); Eugene M. Laska, PhD (New York University, Nathan S. Kline Institute for Psychiatric Research); Thomas Laughren, MD (US Food and Drug Administration); Paul Leber, MD (Neuro-Pharm Group); William Lenderking, PhD (Pfizer); Craig H. Mallinckrodt, PhD (Eli Lilly); James C. Mundt, PhD (Healthcare Technology Systems, Inc); William Potter, MD (Merck); Richard Shader, MD (Tufts University); David V. Sheehan, MBA, MD (University of South Florida); Saul Shiffman, PhD (University of Pittsburgh); Michael Thase, MD (University of Pennsylvania); and Anthony L. Vaccarino, PhD (The International Society for CNS Drug Development).

Financial disclosure: Dr Greist has received grant/research support from AstraZeneca, Eli Lilly, Forest, GlaxoSmithKline, Janssen, Merck, National Institute of Mental Health, Novartis, Pfizer, Solvay, UCB, and Wyeth-Ayerst; has been a consultant to GlaxoSmithKline, Eli Lilly, Ortho-McNeil, Pfizer, Solvay, and Wyeth-Ayerst; and is a principal in Healthcare Technology Systems, Inc., which has licensed interactive voice response assessments for phase 1 through 4 pharmaceutical trials to ClinPhone. Drs Laska, Mallinckrodt, Mundt, Leber, Vaccarino, and Kalali report no additional financial or other relationships relevant to the subject of the article.

Funding/support: Funding and material support was received under the auspices of the International Society for CNS Drug Development (ISCDD [www.iscdd.org]), San Diego, California. The ISCDD funded the consensus conference with support from Astra-Zeneca, Bristol-Myers Squibb, GlaxoSmithKline, Janssen, Forest, Eli Lilly, Novartis, Organon, Pfizer, and Sepracor.

REFERENCES

1. Laska EM, Siegel C, Sunshine A. Commentary—onset and duration: measurement and analysis. *Clin Pharmacol Ther.* 1991;49:1–5.
2. Laska EM, Siegel C. Characterizing onset in psychopharmacological clinical trials. *Psychopharmacol Bull.* 1995;31(1):29–35.

3. Laska EM, Siegel C. Assessing the onset of relief of a treatment for migraine. *Cephalalgia*. 2000;20(8):724–731.
4. Tamura RN, Faries DE, Feng J. Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer-von Mises test. *Stat Med*. 2000;19(16):2169–2184.
5. Adell A, Castro E, Celada P, et al. Strategies for producing faster acting antidepressants. *Drug Discov Today*. 2005;10(8):578–585.
6. Artigas F, Romero L, de Montigny C, et al. Acceleration of the effect of selected antidepressant drugs in major depression by 5-HT_{1A} antagonists. *Trends Neurosci*. 1996;19(9):378–383.
7. Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. *J Clin Psychiatry*. 2001;62(suppl 15):12–17.
8. Blier P. Possible neurobiological mechanisms underlying faster onset of antidepressant action. *J Clin Psychiatry*. 2001;62(suppl 4):7–11.
9. Briner K, Dodel RC. New approaches to rapid onset antidepressants. *Curr Pharm Des*. 1998;4(4):291–302.
10. Laska EM, Meisner M. Nonparametric estimation and testing in a cure model. *Biometrics*. 1992;48(4):1223–1234.
11. Nierenberg AA, Farabaugh AH, Alpert JE, et al. Timing of onset of antidepressant response with fluoxetine treatment. *Am J Psychiatry*. 2000;157(9):1423–1428.
12. Stahl SM, Nierenberg AA, Gorman JM. Evidence of early onset of antidepressant effect in randomized controlled trials. *J Clin Psychiatry*. 2001;62(suppl 3):17–23.
13. Dremencov E, Gispán-Herman I, Rosenstein M, et al. The serotonin-dopamine interaction is critical for fast-onset action of antidepressant treatment: in vivo studies in an animal model of depression. *Prog Neuropsychopharmacol Biol Psychiatry*. 2004;28(1):141–147.
14. Miller FE. Strategies for the rapid treatment of depression. *Hum Psychopharmacol*. 2001;16(2):125–132.
15. Montgomery SA. Fast-onset antidepressants. *Int Clin Psychopharmacol*. 1997;12(suppl 3):S1–S5.
16. Scorza C, Silveira R, Nichols DE, et al. Effects of 5-HT-releasing agents on the extracellular hippocampal 5-HT of rats: implications for the development of novel antidepressants with a short onset of action. *Neuropharmacology*. 1999;38(7):1055–1061.
17. Entsuah R, Derivan A, Kikta D. Early onset of antidepressant action of venlafaxine: pattern analysis in intent-to-treat patients. *Clin Ther*. 1998;20(3):517–526.
18. Gorman JM, Korotzer A, Su G. Efficacy comparison of escitalopram and citalopram in the treatment of major depressive disorder: pooled analysis of placebo-controlled trials. *CNS Spectr*. 2002;7(4 suppl 1):40–44.
19. Leon AC, Blier P, Culepper L, et al. An ideal trial to test differential onset of antidepressant effect. *J Clin Psychiatry*. 2001;62(suppl 4):34–36.
20. Quitkin FM, Taylor BP, Kremer C. Does mirtazapine have a more rapid onset than SSRIs? *J Clin Psychiatry*. 2001;62(5):358–361.
21. Behnke K, Sogaard J, Martin S, et al. Mirtazapine orally disintegrating tablet versus sertraline: a prospective onset of action study. *J Clin Psychopharmacol*. 2003;23(4):358–364.
22. Derivan A, Entsuah AR, Kikta D. Venlafaxine: measuring the onset of antidepressant action. *Psychopharmacol Bull*. 1995;31(2):439–447.
23. Fabre LF. Treatment of depression in outpatients: a controlled comparison of the onset of action of amoxapine and maprotiline. *J Clin Psychiatry*. 1985;46(12):521–524.
24. Jouvent R, Le Houezec J, Payan C, et al. Dimensional assessment of onset of action of antidepressants: a comparative study of moclobemide vs clomipramine in depressed patients with blunted affect and psychomotor retardation. *Psychiatry Res*. 1998;79(3):267–275.
25. Montgomery SA, Bech P, Blier P, et al. Selecting methodologies for the evaluation of differences in time to response between antidepressants. *J Clin Psychiatry*. 2002;63(8):694–699.
26. Nierenberg AA, Griest JH, Mallinckrodt CH, et al. Duloxetine versus escitalopram and placebo in the treatment of patients with major depressive disorder: onset of antidepressant action, a noninferiority study. *Curr Med Res Opin*. 2007;23(2):401–416.
27. De Paula AJM, Omer LMO. Diclofenac (Ro 84650), a new psychoactive drug: its efficacy and safety in non-psychotic depression under double-blind placebo-controlled conditions. *Curr Ther Res*. 1980;28:837–844.
28. Leber P. Speed of onset. *Psychopharmacol Bull*. 1995;31(1):37–40.
29. Mallinckrodt CH, Detke MJ, Christopher J, et al. Comparing methodologies for assessing onset of action in neuropsychiatric clinical trials. *Stat Med*. 2006;25(14):2384–2397.
30. Huitfeldt B, Montgomery SA. Comparison between zimeldine and amitriptyline of efficacy and adverse symptoms—a combined analysis of four British clinical trials in depression. *Acta Psychiatr Scand Suppl*. 1983;308:55–69.
31. US Department of Health and Human Services. Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. <http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf> Published February 2006. Accessed May 12, 2009.
32. Jefferson JW. Strategies for switching antidepressants to achieve maximum efficacy. *J Clin Psychiatry*. 2008;69(suppl E1):14–18.
33. Mundt J, Gelenberg AJ, Greist JH. Patient reported symptoms of depression: which symptoms improve first from the patients' perspectives? Poster abstract presented at: Annual Meeting of the New Drug Clinical Evaluation Unit; May 27–30, 2008; Phoenix, AZ.
34. Santen G, Danhof M, Della Pasqua O. Evaluation of treatment response in depression studies using a Bayesian parametric cure rate model. *J Psychiatr Res*. 2008;42(14):1189–1197.
35. Thase ME. Methodology to measure onset of action. *J Clin Psychiatry*. 2001;62(suppl 15):18–21.
36. Quitkin FM, Rabkin JD, Markowitz JM, et al. Use of pattern analysis to identify true drug response: a replication. *Arch Gen Psychiatry*. 1987;44(3):259–264.
37. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *J Am Stat Assoc*. 1952;47:501–515.
38. Orazem J. *Rank Tests to Test Equality of Two Survival Distributions in a Cure Model* [PhD thesis]. New York, NY: Columbia University; 1990.
39. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992;79:531–541.
40. Peng Y, Dear K. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000;56(1):237–243.
41. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000;56(1):227–236.
42. Lam KF, Fong DYT, Tang OY. Estimating the proportion of cured patients in a censored sample. *Stat Med*. 2005;24(12):1865–1879.
43. Fang HB, Li G, Sun J. Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics: Theory and Applications*. 2005;32:59–75.
44. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*. 1982;38(4):1041–1046.
45. Koti K-M. Gamma failure-time mixture models: yet another way to establish efficacy. *Pharm Stat*. 2003;2:133–144.
46. Price DL, Manatunga AK. Modelling survival data with a cured fraction using frailty models. *Stat Med*. 2001;20(9–10):1515–1527.
47. Corbière F, Joly PA. SAS macro for parametric and semiparametric mixture cure models. *Comput Methods Programs Biomed*. 2007;85(2):173–180.