Measuring Onset of Antidepressant Action in Clinical Trials: An Overview of Definitions and Methodology

Andrew C. Leon, Ph.D.

This article discusses 4 critical questions that must be addressed in any valid analysis of comparative onset of therapeutic effect among antidepressant drugs: (1) how is onset defined, (2) how is onset measured, (3) how can the treatment groups be compared statistically with regard to onset of effect, and (4) how do these issues affect protocol design? A rigorous study of differential onset of effect should include an estimation of the portion of patients that respond, the timing of the onset of response, and the duration of response and an examination of the dynamic process of the onset of action. Methods for measuring symptom severity changes over time also will be reviewed.

(J Clin Psychiatry 2001;62[suppl 4]:12–16)

A valid assessment of differential onset of effect among antidepressants must address several questions. First, how is onset defined? Second, how is onset of effect measured? Third, how can the treatment groups be compared statistically with regard to onset of effect? Finally, how do these issues affect protocol design?

In addition to estimating the proportion of patients who respond, the timing of onset of response, and the duration of response, a well-designed study of differential onset should examine the dynamic process of the onset of action; that is, how symptom severity changes over time. This article reviews the key conceptual and methodological considerations of a study of comparative time to onset and applies various statistical approaches to existing data to illustrate the importance of selecting an appropriate analytic strategy.

DEFINING ONSET

What defines onset of antidepressant action, and how does it relate to response, remission, and recovery? As noted by Prien et al.,[1] these terms are used inconsistently in the psychopharmacology literature. Frank and colleagues[2] discussed these conceptual inconsistencies (although not in the context of early onset) and called for operational definitions that include both the degree of symptom reduction and duration of reduction.

Clearly, onset of antidepressant action occurs somewhere on the continuum between the fully symptomatic state and the asymptomatic state. But at what point? For example, can onset of action and response be considered synonymous? It seems reasonable to assert that onset commences before response and that the time to onset is shorter than the time to response.

In clinical trials, onset typically is defined categorically. For example, a responder can be defined as one who no longer meets DSM-IV criteria for major depression.[3] However, this definition ignores the body of evidence suggesting that subthreshold patients (i.e., those who do not meet full DSM-IV criteria) can be quite symptomatic and functionally impaired.[4–6] More commonly, response is defined by dichotomized severity measures, such as a Hamilton Rating Scale for Depression (HAM-D)[7] score less than or equal to 7 or a 50% reduction in the HAM-D score. The latter definition must be reconciled with the clinical trial inclusion criteria, such that no one labeled a responder at the end of the trial remains symptomatic enough to meet criteria for entering the trial.

Categorical definitions tend to be preferred because they are useful for clinical decision making. However, they are of less value for evaluating onset of action of antidepressants. A dimensional approach is more sensitive to the subtle changes in symptom severity that must be detected in a study of onset. Dimensional measures of symptom severity include the HAM-D, the Beck Depression Inventory,[8] the Montgomery-Asberg Depression Rating Scale,[9] and the Clinical Global Impressions scale.[10]

The timing and duration of improvement must be incorporated in a definition of onset. For instance, which assessment during the course of a clinical trial defines onset? Is it the first assessment that meets onset criteria? Is it the final assessment in the trial? Or should consecutive assessments that meet onset criteria be required? Consider the 2
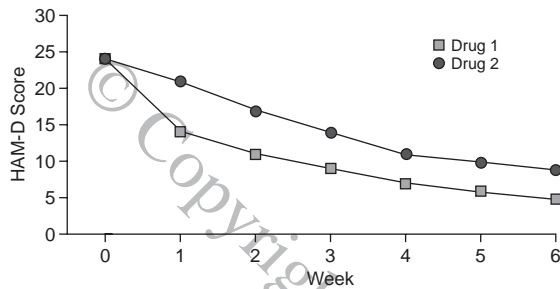
Table 1. HAM-D Scores of 2 Hypothetical Subjects

Week	0	1	2	3	4	5	6
Patient A	36	27	18	17	6	11	9
Patient B	36	26	16	10	6	5	5

Figure 1. Differential Slopes: Consistent Efficacy[a]

[a]Hypothetical data from the Hamilton Rating Scale for Depression (HAM-D) over the course of a 6-week trial in which treatment drug 1 has a steeper slope, i.e., quicker onset the first week, and maintains that advantage throughout the trial.
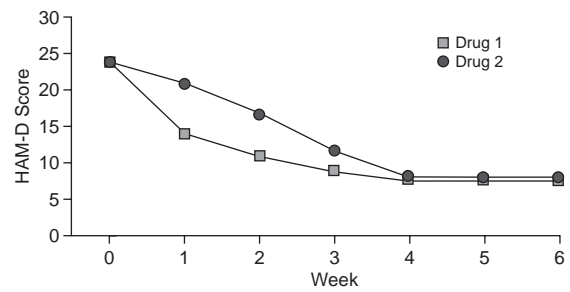
Figure 2. Differential Slopes: Undifferentiated Efficacy[a]

[a]Hypothetical data from the Hamilton Rating Scale for Depression (HAM-D) over the course of a 6-week trial in which the initial advantage of treatment drug 1 is lost by week 4.

Figure 3. Differential Slopes: Changing Efficacy[a]

[a]Hypothetical data from the Hamilton Rating Scale for Depression (HAM-D) over the course of a 6-week trial in which the apparent initial advantage of treatment drug 1 is lost by week 3, and comparator drug 2 is superior by week 3 through the end of the trial.

hypothetical cases shown in Table 1. Although both subjects meet the onset criterion (HAM-D score ≤ 7) at week 4, the first subject's symptom severity changes thereafter, while the second subject continues to meet the criterion. Are they both responders?
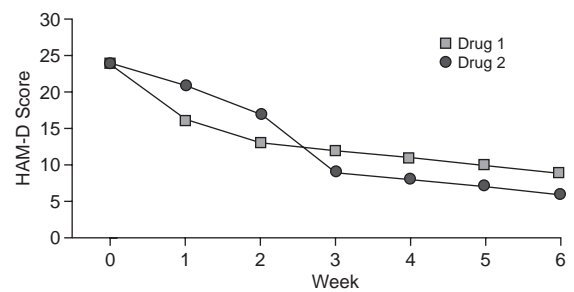
MEASURING ONSET

To detect subtle differences in timing of onset, psychometric assessments must be made with adequate frequency during the trial. Evaluations every 2 weeks clearly are too coarse. Daily assessments, on the other hand, are probably excessive or even counterproductive, since they may sensitize patients to the assessment instrument. Weekly or even twice-weekly assessments seem most appropriate for studies of onset. A prospective daily mood record, or a mood diary, also might be worthwhile.

Unfortunately, most of the standard instruments used in clinical studies of antidepressants are not designed to measure change over brief intervals of time. The HAM-D, for example, queries symptom severity over the past 2 weeks. To detect differential onset of antidepressant action, existing instruments would have to be modified to reflect a different time frame, and modified instruments would in turn need validation.

If standard instruments are altered to measure onset, the reliability and validity of the new scales must be examined empirically. Items that are not amenable to briefer assessment intervals (e.g., items to which subjects become sensitized with more frequent queries) must be revised further or deleted. The test-retest reliability of revised scales must be examined within the brief assessment intervals that are used. Likewise, the internal consistency among the revised items should be quantified using Cronbach's coefficient alpha.[11] This is particularly important because, holding other factors constant, within-group variability decreases as reliability increases. As a consequence, the between-group effect size is increased, and, most importantly, the clinical trial sample size requirements are reduced.[12]

DATA ANALYTIC STRATEGIES

Several issues must be considered when selecting a data analytic approach for a study of onset. Most importantly, the statistical strategy should capture the dynamic nature of the onset of action. Hypothetical data of the HAM-D over the course of a 6-week trial are plotted in Figures 1, 2, and 3. In Figure 1, one group is superior, has a steeper slope and a quicker onset the first week, and maintains that advantage throughout the trial. In Figure 2, the advantage is lost by the fourth week. In Figure 3, the advantage is lost in the third week, and the other drug is superior in the last several weeks in the trial, illustrating a trial in which the drug with quicker onset of action does not appear to be the drug of choice.

The data analytic strategies must be appropriate for the form of the definition of onset (i.e., categorical or dimensional measures). On the surface, a chi-square test comparing response rates in each group might appear desirable, but it oversimplifies the research question. Although the chi-square test provides a cross-sectional snapshot of the percentage of responders at a particular week, at the end of the study, or at the final assessment, it fails to capture the dynamic nature of symptomatic change. Survival analysis is somewhat dynamic in that it examines the cumulative rates of onset over time.[13–15] Nevertheless, it examines a dichotomized definition of onset. Furthermore, in this context, an implicit assumption of survival analysis is that once a subject has responded, the subject remains a responder.

Analyses of dichotomized outcomes ignore the distinction among subjects with a wide range of symptom severity, all of whom are classified as responders. For instance, if nonresponse is defined as ≥ 8 on the 17-item HAM-D, it is implicitly assumed that an 8 on the HAM-D is more similar to a 36 on the HAM-D than it is to a 7. In addition to being unreasonable, such arbitrary divisions introduce misclassification and measurement error into the analysis. Furthermore, the pooling of heterogeneous subjects reduces statistical power.

Another consideration for selecting a data analytic strategy is the ubiquity of missing data in clinical trials. Unfortunately, the mechanisms that give rise to missing data are not random, although it is convenient to assume so. Thus, subjects with missing data are not necessarily representative. Moreover, it cannot be assumed that responders and nonresponders are equally likely to drop out of a trial. Subjects receiving active medication may be more likely to drop out if they are feeling very well. In contrast, subjects receiving placebo are more likely to drop out if they are doing very poorly. Observed treatment differences often attenuate during the course of a clinical trial owing, in part, to these treatment-specific, nonrandom dropout mechanisms.

In a valid analysis of efficacy, the complexity of missing data cannot be ignored under the cloak of the last-observation-carried-forward approach. The implicit assumption of this approach—that a subject's severity rating at the time of dropout would be the same as his or her rating at the end of the trial—is fundamentally flawed. In an onset study, the statistical procedure must take into account the week number that corresponds with each assessment. More importantly, every effort should be made to assess subjects even after they drop out of the study.[16]

The data analytic procedure must use the available data without completely excluding subjects who are missing data. A mixed-effects regression model examines the dynamic process of onset of action and can include available data on subjects who are missing data at other points in the trial.[17–19] It can also include a varied number of observa-

tions per subject. The random-effect terms can account for the correlation among subject observations.

There are mixed-effect models for dimensional, dichotomized, and ordinal categories. A mixed-effect logistic regression analysis can be used for dichotomous data such as weekly ratings of responder status over the course of a clinical trial.[20] A mixed-effect linear regression analysis can be used to examine weekly severity ratings such as the HAM-D. Linear regression is more sensitive than a dichotomous approach and more flexible than analysis of variance or multiple analysis of variance. Perhaps most importantly, it does not require a full data set or imputation of missing data.

With a mixed-effect linear regression model, there are several ways to test for differential onset of symptom reduction. First, the treatment-by-time interaction can be examined to determine whether or not the slopes of symptom reduction are parallel for the 2 groups. This addresses the question, "Are symptom ratings declining more quickly for 1 group than for the other?" Second, polynomial terms can be used to examine the change in slopes over time for each group. Finally, the model can incorporate a spline function to compare early and late aspects of the trial.

As an alternative, however, severity scores can be examined as ordered categories, but without the assumption of equal intervals between scores.[21] For instance, one need not assume that the difference in illness severity between subjects with HAM-D ratings of 30 and 20 is the same as that between subjects with ratings of 10 and 0. Instead, with ordered categorical data one simply acknowledges the ordered levels of illness severity. (That is, a subject with a HAM-D score of 30 is sicker than one with a 20, who in turn is sicker than a subject with a 10, and so on.) An ordinal mixed-effect logistic regression model could be used for such data.[20]

Finally, there are data analytic procedures not described here that are worthy of consideration. For instance, some of the recent advances in pharmacokinetic and pharmacodynamic modeling also use a mixed-model approach.[22] Pharmacokinetic models examine dose-concentration relationships. Pharmacodynamic models examine concentration–clinical response relationships. Other approaches have been discussed by Laska et al.[23] and Siegel et al.[24] in studies of analgesics.

APPLICATION

Keller et al.[25] described a 3-site, 16-week, randomized, double-blind clinical trial of alprazolam, imipramine, and placebo for patients who met criteria for both panic disorder and depression. Assessments were conducted at baseline and weeks 2, 4, 6, 8, 12, and 16. Klerman et al.[26] reported results of the same trial with mean HAM-D score reduction as the outcome variable. There was differential dropout over the course of the trial: by week 16, 31 (72.1%)

Figure 4. Survival Analysis: Time Until 24-Item Hamilton Rating Scale for Depression (HAM-D) Score ≤ 8[a]
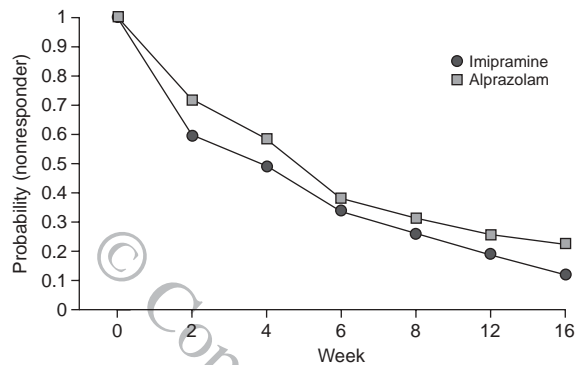
[a]Survival analysis examining the time to reach a 24-item HAM-D score ≤ 8 in a 16-week trial described by Keller et al.[25] failed to detect differential timing in onset of action.

Figure 5. Mean Hamilton Rating Scale for Depression (HAM-D) Scores by Treatment Over 16-Week Trial[a]

[a]Mixed-effect ordinal logistic regression analysis of the course of symptom severity over the course of a 16-week trial described by Keller et al.[25] Weekly means, derived from the 81 subjects who were assessed at least once postbaseline (a total of 521 observations), are plotted. There was a significant treatment-by-time interaction (–2 log likelihood = 10.92, df = 1, p < .001).

remained in the alprazolam group and 35 (77.8%) in the imipramine group. No clear differences were noted between the 2 active treatments.
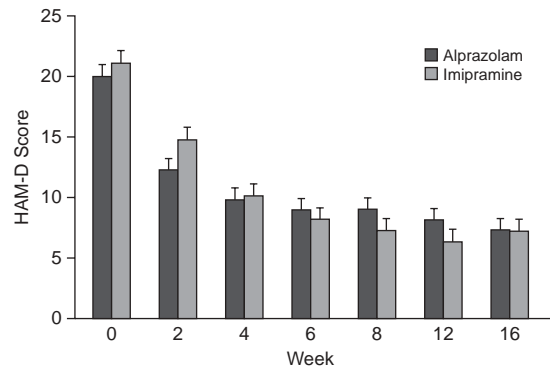
To illustrate how various data analytic strategies can be used to study onset of action, I have applied 2 of the approaches described above to data from the study by Keller and colleagues. The first approach is a survival analysis examining the time to reach a 24-item HAM-D score ≤ 8. The second is a mixed-effect ordinal logistic regression analysis of the course of symptom severity over the 16-week trial. Both analyses are limited to subjects assigned to one of the active medications.

Like the original cross-sectional analyses described by Klerman et al.,[26] the survival analysis of Keller and colleagues'[25] data failed to detect differential timing in onset of action (Figure 4; log-rank χ^2 = 1.19, df = 1, p = .27). The mixed-effect ordinal logistic regression, on the other hand, did detect a difference. Weekly means, derived from the 81 subjects who were assessed at least once postbaseline (a total of 521 observations), are plotted in Figure 5. There was a significant treatment-by-time interaction (–2 log likelihood = 10.92, df = 1, p < .001). The parameter estimates from the mixed model indicate that there was quicker onset of action for subjects assigned to alprazolam, but that advantage diminished during the course of the trial.

SUMMARY

Although data from standard clinical trials might be useful for a preliminary evaluation of onset of action, a study that is specifically designed to detect timing of onset will be more informative. The design of the onset study should be guided by the clinical objective of the clinical trial. The definition of onset must account for the dynamic nature of onset, the symptom reduction, and the duration of reduction over time. The protocol must include more frequent assessments than are typically used in standard clinical trials for psychotropic agents. To detect subtle differences in timing, the instruments must be modified to correspond to the time frame of these assessment intervals. The psychometric properties of the revised instruments must be examined to determine if they are reliable and valid.

The data analytic procedures must be sensitive to the differential change over time in order to detect differential onset. The analyses must be flexible enough to incorporate available data without excluding subjects who fail to complete all assessments. The mixed-effect model approach addresses each of these issues and captures the dynamic process of onset of action of antidepressants.

Drug name: alprazolam (Xanax and others).

REFERENCES

1. Prien RF, Carpenter LL, Kupfer DJ. The definition and operational criteria for treatment outcome of major depressive disorder: a review of the current research literature. Arch Gen Psychiatry 1991;48:796–800
2. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991;48:851–855
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association; 1994
4. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. JAMA 1989; 262:914–919
5. Johnson J, Weissman MM, Klerman GL. Service utilization and social morbidity associated with depressive symptoms in the community. JAMA 1992;267:1478–1483
6. Judd LL, Akiskal HS, Maser JD, et al. A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch Gen Psychiatry 1998;55:694–700
7. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry

1960;23:56–62

8. Beck AT, Ward CH, Mendelson M. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571

9. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389

10. Guy W. ECDEU Assessment Manual for Psychopharmacology. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, Md: National Institute of Mental Health; 1976

11. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334

12. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. Arch Gen Psychiatry 1995;52:867–871

13. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York, NY: John Wiley & Sons; 1980

14. Lawless JF. Statistical Methods for Lifetime Data. New York, NY: John Wiley & Sons; 1982

15. Lee ET. Statistical Methods for Survival Data Analysis. Belmont, Calif: Lifetime Learning Publications; 1980

16. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? Neuropsychopharmacology 1992;6:39–48; discussion 49–63

17. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods 1997;2:64–78

18. Gibbons RD, Hedeker D, Elkin I, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data: application to the NIMH Treatment of Depression Collaborative Research Program dataset. Arch Gen Psychiatry 1993;50:739–750

19. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38:963–974

20. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. Biometrics 1994;50:933–944

21. Agresti A. Tutorial on modeling ordered categorical response data. Psychol Bull 1989;105:290–301

22. Machado SG, Miller R, Hu C. A regulatory perspective on pharmacokinetic/pharmacodynamic modeling. Stat Methods Med Res 1999;8:217–245

23. Laska EM, Siegel C, Sunshine A. Onset and duration: measurement and analysis. Clin Pharmacol Ther 1991;49:1–5

24. Siegel C, Sunshine A, Richman H, et al. Meptazinol and morphine in postoperative pain assessed with a new method for onset and duration. J Clin Pharmacol 1989;29:1017–1025

25. Keller MB, Lavori PW, Goldenberg IM, et al. Influence of depression on the treatment of panic disorder with imipramine, alprazolam and placebo. J Affect Disord 1993;28:27–38

26. Klerman GL, Shear MK, Leon A, et al. Treating the depression associated with panic disorder. Giorn Ital Psicopat 1996;2:1–8