Published in final edited form as:

Clin Trials. 2012 October ; 9(5): 596–604. doi:10.1177/1740774512450099.

The Illness Density Index (IDI) : A longitudinal measure of treatment efficacy

Mary E Kelley^a, Alexandre R Franco^{b,c}, Helen S Mayberg^b, and Paul E Holtzheimer^{b,d}

^aDepartment of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA

^bDepartment of Psychiatry, Emory University, Atlanta, GA, USA

^cDepartment of Electrical Engineering, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre RS, Brazil

^dDepartments of Psychiatry and Surgery, Geisel School of Medicine at Dartmouth, Lebanon, NH

Abstract

Background—A reliable and meaningful quantitative index of success is paramount in the trial of any new treatment. However, existing methods for defining response and remission for treatments tested for psychiatric disorders are limited in that they often minimize the variance in change over time among individual patients and generally use arbitrarily chosen levels of functioning at specified times during treatment.

Purpose—To suggest and determine the properties of an alternative measure of treatment success, the Illness Density Index (IDI), that may be more sensitive to fluctuations in symptoms over the course of treatment compared to existing measures.

Methods—We examined data from 64 depressed patients with multiple assessments of the Hamilton Depression Rating Scale (HDRS) over 12 weeks of randomized treatment in order to compare and contrast varying numerical definitions of response and remission, including percent change and linear slope over time.

Results—Examination of the indices comparing the within-sample rank of individual patients revealed that these indices agree in cases where patients have little or no response as well as clear and sustained response, while they differ in patients who have a slow (or late) response as well as relapse during the treatment course.

Limitations—The measure may not be useful for all types of studies, especially short-term treatment trials.

Conclusions—The IDI is highly correlated with both categorical (e.g., remission) and continuous (e.g., percent change) definitions of treatment success. Furthermore, it differentiates certain trajectories of change that current definitions do not. Thus, the proposed index may be a valuable addition to current measures of efficacy, especially when trying to identify biological substrates of illness or predictors of long-term outcome.

© The Author(s), 2012

Author for correspondence: Mary E Kelley, Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Rd NE, Atlanta, GA 30322, USA. mekelle@emory.edu.

Conflict of interest
None

Introduction

The demonstration of adequate treatment response is inherent to any trial of a potential new intervention. However, current measures of treatment success, while more formally and consistently defined than in the past [1,2], may minimize the variation among individual patients in severity over time and, most importantly, in fluctuations over the treatment period. In short-term studies (lasting only a few weeks), this may not be of significant concern in evaluating the success of treatment. However, in other studies, a measure of average severity over time, as opposed to categorically defined response or remission at an arbitrary time point, may be more important in judging efficacy. This is especially true for the studies of implanted devices for neuropsychiatric disorders (e.g., vagus nerve stimulation or deep brain stimulation (DBS) for epilepsy, Parkinson's disease, treatment-resistant depression (TRD), or obsessive-compulsive disorder (OCD)) where both short-term and long-term efficacies are to be evaluated, typically over months to years [3–6]. Given the invasiveness of these treatments, it is likely that success of these interventions will be based not only on short-term response but also on the degree to which patients achieve and *consistently* maintain improvement over the long term.

This article proposes an alternative metric, the Illness Density Index (IDI), for measuring illness severity over an extended period of time. By not relying on an arbitrary end point or cutoff value, the IDI is more sensitive to fluctuations in symptoms over time that affect the overall clinical picture or 'success' of a treatment. This metric is compared to the most common metrics used in depression treatment studies to measure efficacy: percent change over time, categorical response/remission, and model-based estimates of change. While the current discussion will focus specifically on the definition of successful treatment in major depressive disorder (MDD), the method proposed can be easily applied to any measure that is deemed adequate for tracking longitudinal aspects of illness. This method is suggested as a competitor to model-based or individually calculated slopes in assessment of long-term success. However, the measure may be of interest in other settings such as the identification of biomarkers of treatment response, where the ability to sustain improvement over time is the target.

Measures of treatment success for behavioral interventions

The clinical example we are using utilizes the Hamilton Depression Rating Scale (HDRS) [7] total score as the outcome of interest for treatment success in MDD. Treatment success can be quantified in a number of ways; however, we will focus on those measures used to summarize repeated measures over time or, more specifically, reduce the data from a multivariate (multiple measures per subject) framework to a univariate (one measure per subject) one. These are (1) change in the measure from baseline to end point and corresponding categorical groupings such as response and remission and (2) model-based methods such as time 'effects' or linear slope estimates.

Baseline to end point comparisons

The most obvious choice for the quantification of change in functioning, the absolute change from baseline, is easily calculated for each subject assuming that the measure has adequate variability (i.e., it is continuous rather than categorical), and there are no dropouts or missing data. However, it is well established that patients begin at varying levels of severity, thus the pre-post change is traditionally adjusted for the baseline value, either statistically using a covariate or more simply by dividing the pre-post change by the baseline value, that is, calculating percent change in the measure. The latter approach, percent change, is often of concern to statisticians because if the baseline measure approaches zero, the measure can exhibit a skewed (nonnormal) distribution and in some cases is not defined [8]. Due to its

clinical interpretability, however, it is the most widely used and provides useful benchmarks for some categorical classifications such as full (≥ 50%) and partial response (>25% but <50%). In addition, clinicians are often interested in comparing patients to one another and therefore desire a simple patient-specific measure, rather than a model estimate. Therefore, we present it here for completeness, with the statistical caveats discussed. More recently, the level of functioning at the end point has been the focus of many treatment assessments, with response defined as a minimum percent change from baseline (e.g., 50% decrease for depression studies) and remission defined by an absolute cutoff meant to indicate virtually no symptoms (e.g., a HDRS < 8 for depression studies). The concepts of response and remission can also be combined to form a classification based on the continuum (e.g., nonresponse, partial response, response, remission). These classifications allow for the comparison of response and remission rates across treatment groups, which can be particularly meaningful to the clinician.

While useful, these classifications also mask the variability in the actual amount of change within a group, calling into question whether we are indeed comparing like with like. For example, one patient may have a 51% decrease in the HDRS score and be labeled a responder, while another patient may have a 49% decrease and be labeled a nonresponder; however, these two patients might be virtually indistinguishable from one another, and the 'non-responder' may actually finish the study with a lower HDRS score than the 'responder' depending on the baseline score for each. Perhaps more problematic, however, is that all these measures rely only on the baseline and end point data and therefore ignore, or may be especially affected by, situational or random fluctuations in functioning that naturally occur during the course of treatment. More specifically, a patient who is doing rather well for most of the treatment period may have an unexpected increase at the end point (perhaps due to an unexpected stressor), which will cause the subject to be classified as a nonresponder, even though the treatment was for the most part successful. These measures, because they are based on specified points in time, may not reflect what would be considered the overall clinical picture or an assessment of how well the subject did over the entire treatment interval. This may be especially problematic when they are used to test potential pretreatment response predictors as they may not reflect the more comprehensive clinical picture.

Model-based methods

There are two main model-based approaches that can be used to quantify or compare treatment response over time, assuming that we are using linear mixed models for continuous outcomes. The primary difference in the approaches is whether we use time as a categorical repeated measure or 'effect' (sometimes referred to as 'mixed effects model repeated measures' or MMRM) [9] or as a continuous predictor by fitting functions of time to the data (commonly known as random coefficient or growth curve analysis). While both incorporate all available data rather than eliminate subjects with missing data, they summarize the data differently. While the MMRM allows for a broader definition of changes in functioning over time (e.g., unspecified covariance structure), it does not easily provide a per-subject summary of the response. In contrast, the concept of a within-subject change over time or slope is easily understood by most clinicians and can easily be calculated either using the raw data for each subject or a growth curve model with subject-level random effects for both intercept (severity at baseline) and slope (rate of change in symptoms over time). Although the comparison of slopes across groups is inherent in repeated-measures analysis of clinical trial data, the subject-level estimates have only recently been considered as possible indicators of treatment response [10,11].

Motivating example

The impetus to develop an alternative measure of antidepressant response in clinical trials was prompted by our experience with studies of DBS for TRD. In an ongoing study at our institution, surgery for DBS is followed by a 4-week sham period, a 24-week active stimulation period, and then longitudinal, naturalistic follow-up. In several patients, it was noted that the accepted measures for characterizing efficacy (detailed above) did not adequately reflect the clinical impression of benefit (or lack thereof). In some cases, patients were labeled as nonresponders at a priori end points, but had clearly done well for much of the study. In others, patients were labeled as responders or remitters at these same end points, but their overall course showed that they were not well for much of their time in the study. An example of this is illustrated in Figure 1. Therefore, it became clear that by measuring antidepressant response at a specific (and arbitrarily chosen) end point was not able to capture how a patient did *overall*. As the latter is likely much more relevant to assessing the effectiveness of treatments (especially implanted devices) intended for the continuous long-term management of patients, we concluded that a new metric was needed.

To illustrate this concept, we present an example from our study of DBS in TRD patients. Details of the protocol and management of these patients are provided elsewhere (clinicaltrials.gov (NCT00367 003)). For this patient, the percent change at 2 years, a decrease of 76%, would indicate substantial improvement in depression severity due to treatment. In contrast, the slope shows essentially no change at all, even though the patient did show improvement over baseline and did reasonably well during some of the treatment period. The patient had several relapses during the treatment period, the most severe and prolonged of which occurred at about the 1-year mark (Figure 1). The clear incongruity between the percent change in depression severity at 2 years, slope, and the patient's overall level of illness highlights the need for a new way to quantify response for such studies.

IDI: definition

The IDI is calculated as the area under the curve (AUC) of the plot of the repeated measure of interest (e.g., HDRS) by time within a subject, adjusted by time under observation in order to be comparable across subjects. This measure is used in a very broad range of applications, including summaries of pain intensity over time [12], bioavailability in pharmacokinetics [13], and as an indicator of physiological responsiveness to pharmacological challenge (e.g., hypothalamic–pituitary–adrenal (HPA) axis) [14]. What the AUC represents is quite different across applications and is clearly related to the aspects of the plot it is summarizing; thus, we will formally define it in the current setting of the longitudinal assessment of symptoms or functioning during treatment (i.e., response).

Given that there are a finite number of assessments during the treatment period on any particular subject, the AUC is measured exactly here using what is in mathematics referred to as the 'trapezoidal rule'. For two adjacent assessments at times t_j and t_{j-1} , the area would be calculated as follows

$$\text{Area} = (t_j - t_{j-1}) \frac{[f(t_j) + f(t_{j-1})]}{2} \quad (1)$$

We used this formula specifically to illustrate the use of this measure in the current context; that is, if $f(t)$ is the measure of symptoms at the assessment t , then the area represents the length of time between assessments multiplied by the average severity of the measure over that interval. We propose that this gives us an index, based on the measure f , of the amount of time spent at a particular functioning level, or 'illness momentum'. If we then sum each of

these measures for the entire period of treatment for each subject and divide by the total time in treatment (e.g., number of weeks), we have a number that is comparable across subjects. It is the summing of the total illness momentum (by analogy mass), divided by the total number of weeks on treatment (by analogy volume), that gives us the term 'density' in the naming of the index.

The remaining concern then is whether or not to standardize the metric to baseline, similar to percent change, to allow for the fact that those who begin at more severe levels are more likely to decrease over time (regression to the mean). For the current discussion, we will adjust for baseline, although the area without this adjustment may also be of interest, depending on how the investigator wishes to define success (see section 'Discussion'). Clearly, similar statistical concerns will exist with the adjusted IDI as with the percent change. However, we make the argument that for the current data, the main statistical issue of dividing by baseline, that is, it results in skewed data when the baseline is close to zero, is not a factor for the current data and in fact would rarely be a factor when using behavioral data such as that used for most clinical trials in psychiatry. Furthermore, even if that were the case, other methods of analysis can be employed [8] if we assume that it is still a clinically meaningful measure. It turns out that due to the nature of the calculation, the final result is equivalent if we divide each measure by the baseline and then calculate the index or if we calculate the raw index and then divide by the baseline at the end. However, for the purposes of graphing the data, plotting the ratio to the baseline gives us a more relevant view of what we are measuring (see Figures 1 and 2). In addition, the slope estimate fitted in this context (i.e., to percent of baseline over time) is then in a similar way 'corrected' for baseline severity allowing all the measures to be directly comparable. The formal definition of the IDI for each subject i , with $j = 1, \dots, n_i$ assessments, would then be as follows

$$IDI = \frac{\text{sum (all areas)}}{\text{total time in treatment} \times \text{baseline}} = \frac{\sum_{j=2}^{n_i} \left[(t_j - t_{j-1}) \frac{f(t_j) + f(t_{j-1})}{2} \right]}{(t_{n_i} - t_1) f(t_1)} \quad (2)$$

It is important to note that although there will most likely be a minimum number of assessments as well as frequency and length of assessment that will allow this to be a useful measure, we do not determine that in this article. Importantly, the IDI shares the added benefit of model-based measures in that it can be calculated in subjects who do not complete the final assessment. Furthermore, the IDI reflects the scale of the measure at hand in that if lower scores indicate better response and less severity over the treatment period, as with the HDRS, the IDI will have the same interpretation.

Comparison of IDI to existing measures of response and remission

Subjects

Our example is based on the long-term evaluation of depression symptom severity in TRD subjects receiving chronic brain stimulation delivered continuously via a permanently implanted surgical device. However, due to the nature of these studies and the relatively small number of subjects available, sample sizes would not permit quantitative comparisons of competing measures. Instead, we used data from a larger randomized clinical trial of depression treatment with both cognitive behavioral therapy (CBT) and escitalopram; the treatment period was 12 weeks. We realize that interpretation of patterns may differ considerably between a long-term evaluation and short-term evaluation such as 12 weeks; therefore, we stress that the data presented are for illustration purposes only. Given that a primary aim was to compare measures of response and remission, our comparison was

limited to the 64 subjects who completed the protocol and had data at the end point. Although the sample size is small for a clinical trial, the current presentation is meant to be descriptive, as it is an assessment of the association among different measures of response and not an attempt to compare any measure of response across treatments. We use data from the HDRS 17-item total score measured weekly by clinical staff. To distinguish this measure from other possible uses of the index, we will refer to this specific calculation as the IDI-D (for depression). Clinical characteristics of the sample are listed in Table 1.

Analytic approach

We first established the validity of the IDI-D by showing the associations with other measures of response on the HDRS. Because response ($\geq 50\%$) is based directly on percent change, we will limit our categorical comparisons to remission only, to provide benchmarks for clinicians who prefer these measures. We stress that the categorical data are presented for illustration only, as the issues with dichotomizing a continuous measure are well documented; statistically, we consider the relevant competitor here to be slope, adjusted for baseline in some way. In the presence of missing data, the slopes calculated from a model and those from the individual data are very highly correlated but not identical; we will use the raw data estimates for simplicity. More importantly, since both IDI and percent change are divided by baseline, it is necessary to use the calculated slope of percent of baseline over time (as in the graphs) in order for the rankings across the three measures to be meaningfully comparable.

We then contrast the measures by describing when they give similar and different profiles of treatment success. For this comparison, we limited ourselves to continuous comparison measures because although categorical groupings can be a meaningful summarization of the data, it is understood that the level of response is an underlying continuum and that any categorization is somewhat arbitrary. More importantly, any continuous measure should allow us to order subjects in any particular sample by their corresponding level of treatment success. A comparison of these orderings using the various measures will then allow us to identify groups of subjects, as well as particular subjects, for which the measures agree and disagree as to the level of success of the treatment. For our purposes, we will use an ascending order, that is, a rank of one would indicate the most reduction in severity or best response for all three comparison measures. Although it could be quantified in a number of ways, we decided first to assign each subject to a quartile (ranked grouping) based on each measure and then examine those cases where the measures were inconsistent in their placing of the subject in the quartile for the sample. Part of the reason for this is that any direct comparison of ranks would be limited by patient variability, that is, we would not expect absolute agreement as the measures are qualitatively different. Given the convergence in concepts and the correlation among measures, we used the absolute value of the difference in quartiles to quantify the classification differences, ignoring direction.

It is important to note that although some ' p -values' are presented for reference, we are not attempting any formal statistical comparison across measures.

Results

Calculation of the IDI-D using the adjustment for baseline reveals an approximately normally distributed measure (Figure 2) with a mean of 0.65 and a standard deviation (SD) of 0.20, illustrating that the use of linear models for analysis of this measure would be appropriate. The comparison among the various measures of treatment response indicates that all are highly related, as would be expected (Tables 2 and 3). None of the measures are significantly correlated with the baseline HDRS, indicating that the attempt to adjust for baseline was numerically successful. Most notably, percent change and the slope are nearly

indistinguishable statistically, with a correlation of 0.89. This is most likely because we used only 'completers' for this comparison, and thus, the change from baseline to end point is the largest component to the slope estimate for this particular sample. An assessment of the relationships among these measures when missing data is present (i.e., the 'available case' or mixed model setting) would be more complicated and is a topic for future evaluations of the method.

Usefulness of the measure for clinical studies then would be dependent upon the demonstration of the fact that it provides information beyond the traditional measures and the identification of those cases in which it might be preferable. More specifically, we wish to identify those cases and scenarios where the IDI and other measures do NOT agree. The comparison of quartiles as well as the low measures of kappa indicates that while all measures are highly correlated, they are far from identical in their rating of treatment success. For percent change and IDI-D, 55% of the subjects were in the same quartile, while slope and IDI-D differed more substantially, with only 39% agreement. More notably, the IDI-D differed from percent change by 2 quartiles in 11% of the subjects, while it differed similarly from slope in 16% of the subjects (Table 4). Thus, there is evidence that the IDI provides a different picture of response than existing measures and therefore can be a source of additional information. In order to illustrate specific cases where the measures differed, we examined the two most common scenarios: (1) where all measures were in agreement as to the success of treatment and (2) where both the percent change and slope gave a different picture from the IDI-D.

Figure 3 illustrates the two cases when all measures are in agreement, that is, clear and sustained response (subject A) and little or no response (subject B). In contrast, Figure 4 illustrates the discriminant value of the IDI-D over existing measures. For subject C, it is clear that there was a slight relapse at the end of the protocol, which reduced the magnitude of percent change although the patient had a reasonably good response. This uncharacteristic change at end point, which is also reflected for subject D but in the opposite direction, is one of the primary determinants of the differences between measures through inspection of all patients (data not shown).

Discussion

This article introduces the IDI, a measure of success of treatment over time. While no single behavioral measure completely captures the complexities of treatment response, this measure defines a longitudinal metric that is more indicative of the response seen across the entirety of the treatment interval, rather than at a specified end point. The use of AUC in this context of repeated behavioral measures allows us to apply a well-known technique to derive a different summary of the overall success of treatment that reflects severity throughout the duration of the treatment interval that is not as sensitive to changes at the end point. This index can easily be used with existing measures of symptoms or function as opposed to modifying the method of data collection (which may also be useful). Examination of short-term trial data indicates that the index agrees with other measures used to quantify treatment success, including percent change (response), remission, and linear slope estimates in cases where there is clear response or no response. In contrast, the index differs from other measures when there is a significant relapse during the treatment period or when the pattern indicates either slow response or an uncharacteristic drop at the end point. For the current example, slow response may be the more applicable interpretation due to the fact that the treatment interval was relatively short (12 weeks), and the pattern indicates gradual improvement. As stated in the 'Introduction' section, the IDI should be especially useful for summarizing clinical status over a much longer period (e.g., several months or years). This will be relevant for longitudinal studies of implanted devices, such as DBS for

TRD where such a summary may be more meaningful in assessing effectiveness as opposed to a categorical outcome at an arbitrary time point. In a long-term study, the presented examples where the index differs from other measures (subjects C and D) could result in quite different conclusions than the same result in a short-term clinical trial. More specifically, for subject C, if the time frame was not 12 weeks, but several years, then we may be more willing to discount the rise at the end as random fluctuation, and in that case, the IDI reflects the fact that the patient spent more time 'well' than not well, whereas the percent change and slope are overly affected by the change at the end point. Similarly, for subject D, if the treatment period was again 2 years, we would be less willing to consider that pattern as 'delayed response' and more likely to put less weight on the drop at the end point of the interval and more weight on the time 'not well', which is indicated by the IDI, but not the percent change or slope.

The use of the IDI with adjustment for baseline facilitates the comparison of the index to other well-used measures, but it is not necessary for the calculation to be meaningful. The purpose of adjustment for baseline is to recognize that a 5-point change can be more substantial for someone starting at a score of 15 (33% change) rather than 30 (17% change). However, for the IDI, as a measure of overall severity over the treatment interval, it may be of interest to take into account that the subject who begins at a score of 30 has more time spent with a higher depression level, which would be better reflected in the uncorrected IDI. Which version to choose would depend on the goals of the study; however, it would appear from the current evaluation that the corrected IDI would make the most sense for short-term treatment studies, while the uncorrected IDI may be more applicable to long-term treatment evaluations.

Finally, the purpose of this article was to introduce the IDI as a proposed metric, using actual data with the 'best-case' scenario to define its characteristics, that is, in well-defined clinical cases with no missing data, as proof of principle. Although it is clear from the example that the IDI can provide additional information, clearly if it is to be useful as a measure of response in future work, the effects of missing data and dropout on the measure also need to be addressed. This includes but is not limited to (1) the effects of percentage of missing data on the estimate, (2) the effects of missingness of 'critical' time points, and (3) how to use the IDI in the presence of missing data. The third point is the most relevant for the next step in the development of this as a possible measure for clinical trials, thus it will also be addressed in a future article; however, we can address some of the issues here.

The IDI, although it represents a longitudinal trajectory or summary of the treatment response, is by definition a patient-specific measure. Thus, although it can be computed for any subject with any length of treatment in the presence of missing data and still be comparable across subjects, it is not clear to what extent missing data would affect the estimate. However, the mixed model growth curve methodology, either using time as a fixed factor or growth curves, relies on missing at random (MAR) assumptions to fit patient-specific profiles or slopes, shares the same concern. It is, however, still the state of the art for clinical trials; that is, it is less biased than 'completers only' analysis or last observation carried forward (LOCF) analysis. Thus, whether this particular measure will be able to make use of the same theoretical argument or will require imputation will need to be addressed with simulations.

Acknowledgments

We thank Boadie Dunlop MD, and C. Edward Craighead PhD for patient recruitment and treatment management, and Tanja Mletzko for study coordination.

Funding

This study was financially supported by NIMH IRO1 MH073719 (HSM), Dana Foundation (HSM), Stanley Medical Research Institute (HSM), Woodruff Foundation (HSM), and K23 MH077869 (PEH).

References

1. Nierenberg AA, Wright EC. Evolution of remission as the new standard in the treatment of depression. *J Clin Psychiatry*. 1999; 60(Suppl. 22):7–11. [PubMed: 10634349]
2. Nierenberg AA, DeCecco LM. Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: A focus on treatment-resistant depression. *J Clin Psychiatry*. 2001; 62(Suppl. 16):5–9. [PubMed: 11480882]
3. Rush AJ, Sackeim HA, Marangell LB, et al. Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: A naturalistic study. *Biol Psychiatry*. 2005; 58:355–63. [PubMed: 16139581]
4. Nahas Z, Marangell LB, Husain MM, et al. Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. *J Clin Psychiatry*. 2005; 66:1097–104. [PubMed: 16187765]
5. Lozano AM, Mayberg HS, Giacobbe P, et al. Subcallosal cingulate gyrus deep brain stimulation for treatment-resistant depression. *Biol Psychiatry*. 2008; 64:461–67. [PubMed: 18639234]
6. Greenberg BD, Gabriels LA, Malone DA Jr, et al. Deep brain stimulation of the ventral internal capsule/ventral striatum for obsessive-compulsive disorder: Worldwide experience. *Mol Psychiatry*. 2010; 15:64–79. [PubMed: 18490925]
7. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56–62. [PubMed: 14399272]
8. Berry DA, Ayers GD. Symmetrized percent change for treatment comparisons. *Am Stat*. 2006; 60:27–31.
9. Mallinckrodt CH, Kaiser CJ, Watkin JG, Molenberghs G, Carroll RJ. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin Trials*. 2004; 1:477–89. [PubMed: 16279288]
10. Lutz W, Stulz N, Kock K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *J Affect Disord*. 2009; 118:60–68. [PubMed: 19217669]
11. Uher R, Muthen B, Souery D, et al. Trajectories of change in depression severity during treatment with antidepressants. *Psychol Med*. 2010; 40:1367–77. [PubMed: 19863842]
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
13. Heldens JG, Weststrate MW, van den Hoven R. Area under the curve calculations as a tool to compare the efficacy of equine influenza vaccines – a retrospective analysis of three independent field trials. *J Immunol Methods*. 2002; 264:11–17. [PubMed: 12191504]
14. Fekedulegn DB, Andrew ME, Burchfiel CM, et al. Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosom Med*. 2007; 69:651–59. [PubMed: 17766693]

Figure 1.

Course of HDRS total score in a patient treated with DBS over a 2-year period. Treatment evaluation begins at week 1 following a 4-week sham phase.

HDRS: Hamilton Depression Rating Scale; DBS: deep brain stimulation; IDI: Illness Density Index.

Figure 2.
Distribution of the IDI for depression ($n = 64$).
IDI-D: Illness Density Index for depression.

Figure 3. Patients where IDI profile agrees with percent change and slope. Subject A – little or no response – and subject B – clear and sustained response.
HDRS: Hamilton Depression Rating Scale; DBS: deep brain stimulation; IDI: Illness Density Index.

Figure 4.

Patients where IDI profile differs from percent change and slope. Subject C – relapse at end point – and subject D – slow response or uncharacteristic change at end point.

Table 1Characteristic of the patient sample ($n = 64$)

Characteristic	Mean (SD) or n (%)
Age	42.3 (8.3)
Male	29 (45.3)
Duration current episode (weeks)	190.3 (370.2)
2+ Prior episodes	45 (71.4)
Baseline HDRS	19.2 (3.4)
Percent change	-52.9 (29.0)
Response	36 (56.3)
Remission	24 (37.5)
IDI-D	0.65 (0.20)

SD: standard deviation; HDRS: Hamilton Depression Rating Scale; IDI-D: Illness Density Index for depression.

Table 2Associations among response criteria – Pearson correlation coefficients ($n = 64$)

	IDI-D	Percent change	Slope
IDI-D	–	0.795 ($p < 0.0005$)	0.615 ($p = 0.0005$)
Percent change	–	–	0.890 ($p < 0.0005$)
Slope	–	–	–
Baseline HDRS	–0.156 ($p = 0.219$)	–0.135 ($p = 0.287$)	–0.176 ($p = 0.163$)

HDRS: Hamilton Depression Rating Scale; IDI-D: Illness Density Index for depression.

Table 3

Remission versus other measures of response

Variable	Remitters	Nonremitters	<i>t</i> (<i>df</i> = 62)	<i>P</i>
IDI-D	0.505 (0.167)	0.729 (0.167)	5.21	<0.0005
Percent change	−78.88 (11.60)	−37.25 (24.79)	7.72	<0.0005
Slope	−0.066 (0.013)	−0.031 (0.020)	7.78	<0.0005

IDI-D: Illness Density Index for depression.

Table 4

Evidence of disagreement among measures of response

Comparison	Difference	Percent change	Slope
IDI-D	None	55%	39%
	1 Quartile	34%	45%
	2 Quartiles	11%	16%
Kappa with IDI-D		0.396	0.188

IDI-D: Illness Density Index for depression.