Journal of Consulting and Clinical Psychology

For Better or Worse: An Individual Patient Data Meta-Analysis of Deterioration Among Participants Receiving Internet-Based Cognitive Behavior Therapy

Alexander Rozental, Kristoffer Magnusson, Johanna Boettcher, Gerhard Andersson, and Per Carlbring

Online First Publication, October 24, 2016. http://dx.doi.org/10.1037/ccp0000158

Journal of Consulting and Clinical Psychology

0022-006X/16/$12.00 http://dx.doi.org/10.1037/ccp0000158

For Better or Worse: An Individual Patient Data Meta-Analysis of Deterioration Among Participants Receiving Internet-Based Cognitive Behavior Therapy

Alexander Rozental
Stockholm University

Kristoffer Magnusson
Karolinska Institutet

Johanna Boettcher
Freie Universität Berlin

Gerhard Andersson
Karolinska Institutet and Linköping University

Per Carlbring
Stockholm University

Objective: Psychological treatments can relieve mental distress and improve well-being, and the dissemination of evidence-based methods can help patients gain access to the right type of aid. Meanwhile, Internet-based cognitive–behavioral therapy (ICBT) has shown promising results for many psychiatric disorders. However, research on the potential for negative effects of psychological treatments has been lacking. ***Method:*** An individual patient data meta-analysis of 29 clinical trials of ICBT (N = 2,866) was performed using the Reliable Change Index for each primary outcome measures to distinguish deterioration rates among patients in treatment and control conditions. Statistical analyses of predictors were conducted using generalized linear mixed models. Missing data was handled by multiple imputation. ***Results:*** Deterioration rates were 122 (5.8%) in treatment and 130 (17.4%) in control conditions. Relative to receiving treatment, patients in a control condition had higher odds of deteriorating, odds ratios (*OR*s) = 3.10, 95% confidence interval (CI) [2.21, 4.34]. Clinical severity at pretreatment was related to lower odds, OR = 0.62, 95% CI [0.50, 0.77], and OR = 0.51, 95% CI [0.51, 0.80], for treatment and control conditions. In terms of sociodemographic variables, being in a relationship, OR = 0.58, 95% CI [0.35, 0.95], having at least a university degree, OR = 0.54, 95% CI [0.33, 0.88], and being older, OR = 0.78, 95% CI, [0.62, 0.98], were also associated with lower odds of deterioration, but only for patients assigned to a treatment condition. ***Conclusion:*** Deterioration among patients receiving ICBT or being in a control condition can occur and should be monitored by researchers to reverse and prevent a negative treatment trend.

What is the public health significance of this article?
Psychological treatments have been found to be successful in treating various psychiatric disorders and improving well-being for many patients. However, while investigating the positive effects of different methods for alleviating mental distress less focus has been given to the potential for negative effects, which is frequently reported in pharmacological research. In response, the current study investigated the degree to which patients receiving Internet-based cognitive–behavioral therapy deteriorated and whether there are any predictors of deterioration, indicating that almost 6% fared worse during the treatment period and 17% in a control condition, with higher symptom levels before treatment, being in a relationship, having at least a university degree, and older age lowering the odds of deterioration for patients receiving treatment.

Keywords: Internet-based cognitive behavior therapy, individual patient data meta-analysis, negative effects, deterioration, predictors

Alexander Rozental, Division of Clinical Psychology, Department of Psychology, Stockholm University; Kristoffer Magnusson, Division of Psychiatry, Department of Clinical Neuroscience, Karolinska Institutet; Johanna Boettcher, Department of Clinical Psychology and Psychotherapy, Freie Universität Berlin; Gerhard Andersson, Division of Psychiatry, Department of Clinical Neuroscience, Karolinska Institutet, and Department of Behavioural Sciences and Learning, Linköping University; Per Carlbring, Division of Clinical Psychology, Department of Psychology, Stockholm University.

We would like to thank the researchers, clinicians, and administrative staff who have been involved in the clinical trials that were included in the individual patient data meta-analysis, who have all been of great assistance while collecting the data. Michael J. Lambert and the research group at Brigham Young University in Salt Lake City, Utah, are also acknowledged for their great contribution to the research of deterioration in psychological treatments. In addition, the authors of the current study would like to thank the Swedish Research Council for Health, Working Life, and Welfare (FORTE 2013-1107) for their generous grant that permitted the investigation of deterioration among patients receiving Internet-based cognitive behavior therapy. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Correspondence concerning this article should be addressed to Alexander Rozental, Division of Clinical Psychology, Department of Psychology, Stockholm University, Frescati Hagväg 8, Stockholm, 106 91. E-mail: alexander.rozental@psychology.su.se

During the last few decades, a number of systematic reviews and meta-analyses have provided increasing support for the use of psychological treatments as a way of alleviating mental distress and enhancing well-being (Hofmann, Asnaani, Vonk, Sawyer, & Fang, 2012). In addition, great efforts have been made to improve the access to evidence-based methods, such as, cognitive–behavioral therapy (CBT), in an effort to disseminate effective psychological treatments to patients suffering from a variety of psychiatric disorders (McHugh & Barlow, 2010). Meanwhile, psychological treatments delivered via formats other than face-to-face, for instance, Internet-based cognitive–behavioral therapy (ICBT) and the use of different smartphone applications, have the potential of becoming an important and widely used addition to the health care system, delivering evidence-based methods to an even larger population at a significantly lower cost (Andersson & Titov, 2014), and with similar benefits for many patients (Cuijpers, Donker, van Straten, Li, & Andersson, 2010). However, although promising steps have been made for helping those who suffer from a psychiatric disorder, research of psychological treatments have focused almost entirely on its positive aspects, particularly, average treatment outcome and the number of patients who have achieved clinically significant change, while paying far less attention to the possible existence of negative effects (Barlow, 2010). Few clinical trials of psychological treatments tend to report adverse events occurring during the treatment period (Rozental et al., 2014), with a recent review indicating that information concerning risks was only described in 28 out of 132 (21%) published randomized controlled trials (Jonsson, Alaie, Parling, & Arnberg, 2014). In comparison to pharmacological research, clinical trials of psychological treatments were nine to 20 times less likely to mention possible or actual negative effects (Vaughan, Goldstein, Alikakos, Cohen, & Serby, 2014). The idea that some patients could experience adverse events due to the treatment they undergo has also largely been ignored throughout the history of psychological treatments, receiving little consideration by researchers and clinicians (Lilienfeld, 2007), even though the first empirical evidence on this issue was in fact published more than six decades ago (Powers & Witmer, 1951). One notable exception is Bergin (1966) who sparked a debate about the "client-deterioration phenomenon" (p. 236), or, the deterioration effect. In reviewing the results from seven clinical trials it was argued that, apart from those patients who improve and do not respond, some patients seem to get worse during the course of their psychological treatment. Albeit criticized for the difficulty of determining causality (cf. May, 1971), several investigations have since then indicated that deterioration appears to be relatively common and occurs across psychiatric disorders and treatment conditions (cf. Hansen, Lambert, & Forman, 2002), with an average number of deteriorated patients ranging between 5 and 10%, and with even higher rates among children, adolescents, and substance abuse patients (Lambert, 2013). With regards to ICBT, similar findings have been found in several randomized controlled trials (cf. Boettcher, Rozental, Andersson, & Carlbring, 2014).

Assessing Negative Effects

Deterioration, defined as a worsening in symptomatology, is not the only way to assess negative effects, and several other suggestions on how to define and monitor adverse events occurring during psychological treatments have been proposed (cf. Strupp & Hadley, 1977). For instance, decreased interpersonal functioning, dependency, and lowered self-esteem have all been put forward as detrimental effects of treatment, and could result in a better understanding of the mechanisms that may be responsible for negative effects if further scrutinized (Dimidjian & Hollon, 2010). For example, when reviewing the literature on negative effects possibly induced by psychological treatments, Boisvert and Faust (2002) discussed how labels, altered self-perceptions, and social roles might be related to a negative outcome. Similarly, Rozental, Boettcher, Andersson, Schmidt, and Carlbring (2015) conducted a qualitative content analysis on the responses to a number of open-ended questions concerning negative effects distributed in four clinical trials of ICBT, demonstrating that insight about what maintains a psychiatric disorder was experienced as distressing by some patients, as was the development of novel symptoms, such as insomnia and stress, difficulties implementing the treatment interventions, and a lack of feedback and guidance. However, although interesting from a theoretical perspective, such investigations often warrant the use of qualitative analyses that may prove hard to generalize. Furthermore, without the systematic use of standardized self-report measures explicitly probing for adverse events, as well as clearly defined and operationalized concepts of what constitutes different types of negative effects, the results would be difficult to interpret. Several suggestions on how to overcome some of these problems and enable the monitoring of negative effects other than deterioration have recently been put forward, including both therapist checklists and self-report measures distributed to the patient (cf. Linden, 2013), but their use is currently limited and their validity is not yet established. Deterioration is therefore still one of the most straightforward methods for detecting and examining negative effects, with the additional advantages of being easy for researchers and clinicians to comprehend as well as allowing comparisons across clinical trials.

Assessing Deterioration

Assessing deterioration can be a complex procedure that requires both theoretical and statistical considerations. Compared to investigating improvement, which typically involves the use of predefined cutoffs on a specific self-report measure or the calculation of what constitutes a clinically significant change, deterioration lacks a frequently used or agreed-upon approach for determining when the condition of a patient has declined. A negative change score from pre- to posttreatment may indicate that a patient has deteriorated, but by how many points that needs to be achieved is unclear (Mohr et al., 1990). This issue is further complicated by the fact that a patient cannot deteriorate indefinitely due to ceiling effects, as well as the problem of perceiving deterioration as a distribution of scores distinct from those of a dysfunctional and functional population (Martinovich, Saunders, & Howard, 1996). Jacobson, Follette, and Revenstorf (1984) were among the first to recognize these concerns, "There is no obvious counterpart to our distributional cutoff for clinical significance in the assessment of deterioration rates" (p. 350), suggesting that the investigation of deterioration is limited to the implementation of the Reliable Change Index (RCI), that is, inspecting whether the deterioration is reliable and not only caused by measurement error. The basic method for calculating the RCI was later refined and outlined by

Jacobson and Truax (1991) as the change score between pre- and posttreatment divided by the standard error of difference between the two test scores. If the resulting RCI is larger than $z = 1.96$, the change score would be considered unlikely ($p = .05$), without a true change really occurring. However, how the standard error of difference should be derived was never explicitly mentioned; only that it could be calculated from the standard error of measurement, creating some confusion on whether to use the internal consistency, such as, Cronbach's α, or the test–retest reliability of the self-report measure, for instance, Pearson *r*. This issue has later been shown to have implications for determining the improvement and deterioration rates among patients receiving psychological treatment (Speer, 1992). Most notably, the internal consistency reflects if the self-report measure being used consists of a single unidimensional construct, that is, assuming that it measures only one factor or concept, while the test–retest reliability introduces variation because of separate occasions of measurement, resulting in lower reliability. Subsequently, if a less reliable self-report measure is being administered, the greater the actual change score has to be in order for it to be regarded as reliable (Evans, Margison, & Barkham, 1998). Thus, if the idea were to assess a relatively stable trait or feature, internal consistency would probably suffice, but if some fluctuation is expected to occur, as when examining symptoms of a given psychiatric disorder, test–retest reliability is recommended (Edwards, Yarvis, Mueller, Zingale, & Wagman, 1978). Furthermore, Tingey, Lambert, Burlingame, and Hansen (1996) argued that the test–retest reliability should be derived from a normal population and cover a short time frame so that it represents the measurement error of scores from a specific interval, preferably 1 to 2 weeks, rather than any potential change that could be attributed to a psychological treatment. In the case of several cases of test-retest reliabilities of the same self-report measure, it is also suggested that the median number should be used. However, acquiring the necessary information for a specific self-report measure can often be difficult, especially for nonclinical samples. Furthermore, different alterations to the RCI have been presented over the years, taking into account regression to the mean as well as correcting for error at both pre- and posttreatment, complicating the issue further, although a comparison of various methods by Bauer, Lambert, and Nielsen (2004) still recommends the original suggestion by Jacobson and Truax (1991). In addition, the RCI has almost entirely been used for investigating the number of patients having improved rather than deteriorated. Hence, in determining deterioration the same standard deviations units have been used as for assessing recovery, that is, $z = 1.96$, even though it could be argued that a less strict criterion should be implemented to account for those who also experience milder forms of deterioration. Wise (2004), therefore, proposed several reliable change indexes to discriminate between different confidence levels related to deterioration, $z = 1.28$ for moderate deterioration ($p = .10$), as well as $z = 0.84$ for mild deterioration ($p = .20$), which could reveal negative effects that might otherwise have been overlooked. In other words, when examining deterioration among patients receiving psychological treatment, one need to recognize a number of issues related to the self-report measure being distributed, what standard error of measurement is available to obtain, as well as what reliable change index to use, before calculating the RCI.

Individual Patient Data Meta-Analysis

Deterioration in itself is insufficient in determining what factors might be contributing to its occurrence. To understand why some patients fare worse during the course of their psychological treatment, research on possible predictors of deterioration is required (Castonguay, Boswell, Constantino, Goldfried, & Hill, 2010). However, due to the relatively small number of patients actually deteriorating in a single clinical trial, statistical analyses will often be underpowered to find meaningful differences. As mentioned by Edwards et al. (1978), "Very large samples of patients would have to be used to develop a large enough group for reliable determination of predictors of deterioration" (p. 286). Thus, to enable a more rigorous study of predictors of deterioration and discover potential subgroups that are at risk of becoming worse, large amounts of data are needed. Individual patient data meta-analysis (IPDM) is an approach to synthesize information from a number of clinical trials, using the raw scores from each patient for a more powerful examination of effects (Oxman, Clarke, & Stewart, 1995). By collecting data from several studies it is possible to undertake much more sophisticated statistical analyses and deal with some of the difficulties associated with investigating less frequently occurring events, such as, deterioration. In relation to ICBT, this approach has previously been used to assess the influence of the baseline severity level of depression on the effectiveness of treatment interventions delivered via the Internet, indicating that the more clinically severe patients benefit from their psychological treatment as much as those with less severity (Bower et al., 2013). Similarly, Karyotaki et al. (2015) found that dropout can be predicted by a number of variables, particularly, male gender, lower educational level, younger age, and comorbid anxiety. However, with regard to deterioration in ICBT, no previous attempt has been made to examine its occurrence or potential predictors using IPDM. The current study is therefore, to the knowledge of the authors, the first to assemble data from numerous clinical trials of ICBT for different psychiatric disorders to explore what factors might be related to deterioration during CBT delivered via the Internet. The overall aim is to determine the deterioration rates among patients receiving ICBT and to investigate plausible predictors of deterioration using variables that, theoretically or empirically, have been suggested to increase the risk of faring worse, as well as for those allocated to some form of control condition.

Method

Collection of Data

The procedure involved in performing IPDM is similar to that of conducting a meta-analysis, but with the essential difference of using the raw scores from each patient, rather than only group means and standard deviations. This includes the collection of data from different sources, either by implementing a systematic review approach and compiling all relevant clinical trials through a rigorous literature search to decide which will be included in the statistical analysis as well as assessing the potential risk of publication bias (Simmonds et al., 2005), or, alternatively, to pool together data from different research groups to investigate a specific question or hypothesis (Riley, Lambert, & Abo-Zaid, 2010).

The current study used the latter method by assembling data from three separate sites managed by the fourth and fifth author, comprising a vast range of clinical trials of ICBT. This was done because of the great availability of a large number of data sets and due to the heterogeneity in terms of patients, treatment modalities, and psychiatric disorders, which is essential for examining possible predictors of deterioration. Although this procedure affects the possibility of assessing the risk of bias that may be introduced by not being able to ensure that the data reflects all relevant clinical trials (Simmonds et al., 2005), it allowed the retrieval of a majority of all clinical trials of ICBT that have been performed in Sweden, which should be representative of how it is being delivered on a national level in both university settings and the regular health care system, that is, consistency in language, use of validated outcome measures, having established procedures, and conducting diagnostic interviews. In terms of collecting the data, predefined inclusion and exclusion criteria have also been used to guide the selection of clinical trials. Data retrieved from the three sites were chosen on account of (a) patients receiving some type of treatment condition involving ICBT, guided or unguided, consisting of treatment interventions that are theoretically linked to CBT, including applied relaxation and cognitive bias modification; (b) patients allocated to some form of control condition, for instance, waitlist control with/without limited support or a discussion forum; (c) psychiatric disorders or V-codes listed in the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth or Fifth Edition (American Psychiatric Association, 2000, 2013); (d) treatment as well as control condition lasting for at least 2 weeks, two sessions, or two modules; and (e) use of a validated outcome measure assessing the level of distress targeted by the treatment interventions, for instance, the Liebowitz Social Anxiety Scale–Self-Report (Liebowitz, 1987). In most cases a random allocation to either treatment/control condition, or different treatment conditions, had been implemented in each of the clinical trials, but this was not used as one of the predetermined inclusion and exclusion criteria for the current study. However, only three cases did not involve any randomization: two pilot studies (Carlbring et al., 2011; Dagöö et al., 2014) and one large-scale study of gambling disorder (Carlbring, Degerman, Jonsson, & Andersson, 2012). Clinical trials as well as raw scores not included in the current study were characterized by a treatment condition other than ICBT, such as, bibliotherapy with telephone support, or some form of psychological treatment not theoretically linked to CBT, namely, psychodynamic psychotherapy and interpersonal psychotherapy.

Data Extraction and Preparation

Data from three separate sites managed by the fourth and fifth author were extracted with the raw scores for each patient, while the clinical trials were assessed for eligibility using the predefined inclusion and exclusion criteria. The raw scores were then entered by the first author into one main data set while continuously checking that all nominal variables were coded with the same numbers, for example, sick leave (1 = yes, 0 = no). This included; name of the clinical trial, type of treatment condition or form of control condition, and including all available sociodemographic variables, outcome measures (primary and additional), ratings of satisfaction and credibility, previous use of psychological treatment and previous or ongoing use of psychotropic medication, sick leave (receiving disability checks, i.e., between 2 weeks and 1 year of absence from work due to a medical or psychiatric condition), number of completed modules, and time spent per week on the treatment interventions. To enable as many comparisons as possible, as different clinical trials used different coding systems, certain sociodemographic variables had to be collapsed. For instance, only single/relationship was retained in terms of civil status, while highest educational level and employment was restricted to fewer but more coherent groups. In addition, to even out large differences between some categories, a number of treatment and control conditions as well as psychiatric disorders were merged together; treatment/control were used instead of a diversity of different modalities, while psychiatric disorders were classified according to anxiety disorders (including posttraumatic stress disorder, panic disorder with/without agoraphobia, and anxiety with/without depression), depression (with/without comorbid dysthymia), as well as other (erectile dysfunction, relationship problems, and gambling disorder). Moreover, because the number of completed modules differed greatly as a consequence of the total number of modules included in each respective clinical trial, the weighted mean and standard deviation were calculated. In the event of unknown or ambiguous labels used among the raw scores, published articles as well as unpublished manuscripts were located and double-checked so that the data set was coded correctly and in accordance with the clinical trials. However, it should be noted that in a few cases, information concerning the coding of the original data sets had been lost and was not possible to retrieve even with the help of the principal investigators from the clinical trials.

Statistical Analysis

The RCI was calculated to investigate the occurrence of deterioration in each clinical trial, utilizing the change score on the primary outcome measure from pre- to posttreatment for a specific patient divided by the standard error of difference between the test scores (Speer, 1992). To compute the standard error of difference, the following formula derived from Evans et al. (1998) was applied: $SE_{diff} = SD_1\sqrt{2}\sqrt{1\text{-}r}$, where SD_1 refers to the standard deviation of the pretreatment scores, and r is the reliability estimate. The RCI was thus worked out separately for every clinical trial, and not according to each of the self-report measures, because this would have affected the standard deviation. Furthermore, in accordance with the recommendations by Edwards et al. (1978), test–retest reliability was used in the current study, preferably stemming from a normal population and with a relatively short time period between the two points of measurement (Tingey et al., 1996). Table 1 includes the references from which the test-retest reliabilities were obtained, the type of population that was used, and their reliability estimates and respective time frames. If Pearson r was unavailable to acquire, intraclass correlation coefficient or Spearman ρ were used instead, and in the event of separate estimates stemming from the same reference, such as, when having different subscales or time periods, the median value was calculated. Also, in line with the idea of several reliable change indexes, as put forward by Wise (2004), the RCI was calculated using 0.84 *SD* units, that is, for mild deterioration. Although resulting in a confidence level of 80% rather than

Table 1
Test–Retest Reliabilities Used for Calculating the Reliable Change Index

Primary outcome	Test–retest reliability	Time period	Population	Reference
Beck Anxiety Inventory	$r = .81$	2 weeks	Normal	Saemundsson et al. (2011)
Liebowitz Social Anxiety Scale–Self-Report	$r = .93$	8 weeks	Normal	Heeren et al. (2012)
Panic Disorder Severity Scale–Self-Report	$\rho = .94$	2 days	Patient	Lee, Kim, and Yu (2009)
Patient Health Questionnaire (nine items)	$r = .94$	2 weeks	Patient	Zuithoff et al. (2010)
International Index of Erectile Functioning (five items)[a]	$r = .84$	4 weeks	Patient	Rosen et al. (1997)
Beck Depression Inventory	$r = .77$[b]		Normal	Beck and Steer (1996)
Impact of Event Scale–Revised	$r = .89$–$.94$[c] $M = .92$	6 months	Patient	Sundin and Horowitz (2002)
Generalized Anxiety Disorder (seven items)	ICC = .83	1 week	Patient	Spitzer, Kroenke, Williams, and Löwe (2006)
Penn State Worry Questionnaire	$r = .84$	3 weeks	Normal	Pallesen, Nordhus, Carlstedt, Thayer, and Johnsen (2006)
Body Sensations Questionnaire	$r = .89$	3 months	Patient	Arrindell (1993)
Dyadic Adjustment Scale[a]	$r = .87$	2 weeks	Patient	Carey, Spector, Lantinga, and Krauss (1993)
Snake Anxiety Questionnaire	$r = .78$	1 month	Normal	Klorman, Weerts, Hastings, Melamed, and Lang (1974)
Montgomery–Åsberg Depression Rating Scale–Self-Report	ICC = .78	1 week	Patient	Fantino and Moore (2009)
Spider Phobia Questionnaire	$r = .94$	3 weeks	Normal	Muris and Merckelbach (1996)
The NORC Diagnostic Screen for Gambling Problems	$r = .98$–$.99$[d] $M = .98$	1 week	Patient	Gerstein et al. (1999)

Note. ICC = intraclass correlation coefficient; NORC = National Opinion Research Center at the University of Chicago.
[a] Reversed scales, higher scores indicate less problems. [b] Information regarding the time period was unavailable. [c] Separate estimates for the two subscales. [d] Lifetime test statistic and past year test statistic.

95% for 1.96, increasing the risk of the change scores being attributed to measurement error, this also enhances the prospect of detecting cases of deterioration, and not only those that are more severe. Any patient with a change score exceeding the RCI was then coded as deteriorated using a separate binary outcome (1 = yes, 0 = no).

To examine potential predictors of deterioration, variables were selected a priori based on theory or, if possible, prior empirical findings (L. Stewart & Tierney, 2002): (a) clinical severity at pretreatment, with higher symptom levels suggesting greater mental distress and possibly increasing the risk for deterioration; (b) civil status, indicating a presence or absence of social support; (c) previous psychological treatment as well as (d) previous or ongoing psychotropic medication, assumed to reflect more recurrent difficulties; (e) sick leave, implying more severe and persistent difficulties or decreased global functioning; (f) educational level, with lower education potentially being related to more stressors; and (g) age, suggesting a higher prevalence of some psychiatric disorders among younger individuals. Albeit not related to deterioration per se, previous findings give some credence to these assumptions, indicating that comorbidity, less education, and lower socioeconomic status are all factors related to higher dropout rates in face-to-face psychotherapy (Cooper & Conklin, 2015; Sharf, Primavera, & Diener, 2010; Swift & Greenberg, 2012). Similarly, male gender, low educational level, and clinical severity at pretreatment were also associated with increased dropout rates in psychological treatments via the Internet, while older age was linked to less dropout (Christensen, Griffiths, & Farrer, 2009; Karyotaki et al., 2015; Waller & Gilbody, 2009). However, as with all analyses involving possible predictors and subgroups, the results should at best be regarded as hypothesis generating, warranting further replication before any firm conclusions can be inferred (Clarke, 2005).

Predictors of deterioration were summarized as odds ratios (*OR*s), reflecting an increase or decrease in odds of deterioration compared to a predetermined reference category. For dichotomous predictors, such as, sick leave, the *OR* indicates the adjustment in odds of deterioration when the patient goes from not being on sick leave (no) to being on sick leave (yes). For the two predictors using continuous scales, that is, clinical severity at pretreatment and age, the *OR* constitutes an increase of one standard deviation above their respective mean.

All statistical analyses were performed using R (version 3.2.3). The individual patient data were examined with generalized linear mixed models using the *lme4* (Version 1.1–10; Bates, Machler, Bolker, & Walker, 2015). Because the dependent variable consisted of binary outcomes, a binomial logistic model was used, allowing a random intercept at the level of the clinical trial to deal with clustering effects. This meta-analytic method is commonly referred to as a one-step approach, because all data is modeled in one analysis, as compared to the two-step approach where separate analyses are performed within each study and meta-analyzed using traditional aggregate methods (Debray, Moons, Abo-Zaid, Koffijberg, & Riley, 2013). The one-step approach was chosen for the current study because it offers greater flexibility in modeling the data and greater statistical power for testing predictors (G. Stewart et al., 2012). With regard to the predictors, these were analyzed separately in univariable generalized linear mixed models, and predictors with a p value less than .05 were kept and later examined in a multivariable model. Both the variable clinical severity at pretreatment and age were standardized and centered within each clinical trial, comparing patients against the average value for their respective clinical trial.

Missing Data

To handle varying degrees of data loss, multiple imputation (MI) was utilized, as is generally recommended when managing missing data in IPDM (Debray et al., 2015). Two separate imputation approaches were used to deal with the challenges that arise for different types of missing data in the current study; sporadically missing data and systematically missing predictors. Because several variables presented some data loss, MI was done by the method of fully conditional specification, as implemented in Mul-

tivariate Imputation by Chained Equations (Version 2.25; van Buuren & Groothuis-Oudshoorn, 2011). As for the sporadically missing data, imputations were stratified by clinical trial to avoid problems with heterogeneous measures and study populations, thereby making it possible to use as much available information as possible within each clinical trial in the imputation process. This imputation model was used to generate 50 data sets with the sporadically missing values imputed. Meanwhile, not every predictor was measured in all of the clinical trials, thus being missing in a systematic fashion. Naturally, it is not feasible to impute these predictors separately within each clinical trial, albeit one potential way of handling systematically missing predictors is to exclude the clinical trials with missing predictors. This method, however, excludes a lot of possibly important data, because only those clinical trials that include all relevant predictors will be retained in the model. A better way is therefore to use multilevel MI and impute over all of the clinical trials (Jolani, Debray, Koffijberg, van Buuren, & Moons, 2015). Missing predictors are subsequently imputed by borrowing information from clinical trials where the predictors are observed, allowing all to be retained in the final analysis. Although, because variables with sporadically missing data were imputed using a different method, a two-stage, or, nested imputation, was conducted for the missing predictors. Each of the 50 data sets, with sporadically missing data imputed, were consequently used to create five new data sets with the missing predictors imputed, leading to a total of 250 (5×50) complete data sets. This two-stage imputation approach leads to nonindependence among the imputed data, and different pooling rules are therefore needed for summarizing the nested data sets (Harel, 2007).

Ethical Considerations

The data used in the current study were derived from a number of clinical trials, all being granted ethical approval from the regional ethical review board in their respective study location (please refer to the original articles for more information). Hence, neither the collection of data or statistical analyses required any additional authorization. Furthermore, the data include only the raw scores from various outcome measures or numbers concerning different variables that have been coded according to certain categories, while leaving out qualitative data, for instance, responses to open-ended questions. Moreover, all patients were classified using the automatically assigned identification code derived in each clinical trial, for example, abcd1234, making it impossible to identify a specific individual. With regards to the ethical issue surrounding the assessment of deterioration among patients receiving ICBT, the current study used only the raw scores from already completed clinical trials, making it impossible to, in hindsight, detect and revert a potentially negative treatment trend. However, because the aim of the current study also was to examine possible predictors of deterioration, future clinical trials may be better able to monitor and assist those patients who fare worse during the treatment period.

Results

Study Characteristics

Available data from 29 clinical trials were reviewed according to the predefined inclusion and exclusion criteria and deemed eligible for the current study. Raw scores from all patients were then entered into the main data set, except for 27 and 19 patients who received either psychodynamic psychotherapy or interpersonal psychotherapy via the Internet as a control condition to ICBT. In total, 2866 patients were included, 2118 (73.9%) receiving some type of treatment condition involving ICBT, and 748 (26.1%) allocated to some form of control condition. The following psychiatric disorders were included (number of clinical trials, *k*): social anxiety disorder (nine), depression (with/without dysthymia; five), generalized anxiety disorder (three), anxiety disorder (with/without depression; three), mixed anxiety disorders (e.g., panic disorder and social anxiety disorder; two), specific phobia (two), posttraumatic stress disorder (one), panic disorder (with/without agoraphobia; one), gambling disorder (one), erectile dysfunction (one), and relationship problems (one). In terms of recruitment, self-referrals from the general population were in absolute majority, 27 clinical trials, while only one was conduced in primary care, and another in a university setting. In almost all of the cases, the Structured Clinical Interview for *DSM–IV*-Axis I Disorders (First, Gibbon, Spitzer, & Williams, 1997) was used for screening purposes, except for four clinical trials that used either the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998) or a structured clinical interview for a specific psychiatric disorder, for example, Clinician-Administered PTSD Scale (Blake et al., 1995). Furthermore, with regard to the length of the treatment and control conditions, the range was four to 10 modules ($M = 8.28$; $SD = 1.36$), 4–12 weeks ($M = 8.45$; $SD = 1.66$), and two to 10 sessions ($M = 5.40$; $SD = 3.58$), specific phobia being the absolute shortest, while various anxiety disorders and relationship problems were the longest. The total amount of missing data for all of the primary outcome measures at postassessment was 12.9%. For a complete overview of the clinical trials, please refer to Table 2.

Sociodemographic characteristics of the patients included in the current study can be obtained in Table 3, divided by treatment and control conditions, and including the percentage of missing data for each of the variables. Due to different coding schemes and open-ended questions asked at posttreatment among the clinical trials, such as, not inquiring about sick leave or treatment credibility, or coding errors, such as, nominal variables where information of how these were coded had been lost, some numbers were not available in the raw scores used in the current study, resulting in varying degrees of data loss. The lowest was gender, 27 cases (0.9%), and the highest was satisfaction with treatment, 1,867 cases (88.2%).

Deterioration Rates

In total, 252 patients exceeded a RCI of 0.84 in a negative direction and can thus be regarded as being reliably deteriorated. The deterioration rates for each clinical trial can be acquired in Table 2, divided by treatment and control. In general, a higher proportion of the patients deteriorated among those in a control condition, 130 (17.4%), compared to those who were receiving treatment, 122 (5.8%). The lowest numbers, that is, no deterioration at all, were found in one clinical trial for panic disorder, one for specific phobia of snakes, as well as one for social anxiety disorder. The highest numbers for the control condition were obtained in two clinical trials of social anxiety disorder, 10

Table 2
Characteristics and Deterioration Rates for the Clinical Trials Included in the Individual Patient Data Meta-Analysis

Study	Recruitment	Screening interview	Primary diagnosis	Treatment (*n*)	Control (*n*)	Modules/weeks or sessions	Primary outcome	Additional outcomes	Deterioration treatment *n* (%)	Deterioration control *n* (%)
IMÅ (Boettcher et al., 2014)	General population	SCID-I	Panic disorder, social anxiety disorder, generalized anxiety disorder, anxiety disorder NOS	Unguided mindfulness with FAQ (42)	Waitlist with discussion forum (46)	8 modules/8 weeks	BAI	BDI, QOLI, ISI	2 (5.1%)	5 (11.6%)
ACT Smart (Lindner et al., 2013)	General population	SCID-I	Panic disorder, social anxiety disorder	Unguided ACT (48), guided ACT (48)	Waitlist (47)	8 modules/10 weeks	LSAS-SR, PDSS-SR[a]	PHQ-9, GAD-7, QOLI	7 (7.2%)	13 (28.2%)
Actua (Carlbring et al., 2013)	General population	SCID-I	Depression	Guided physical activity (164), guided behavioral activation (158)[b]	n.a.	8 modules/12 weeks	PHQ-9	GAD-7, QOLI[c]	27 (8.4%)	n.a.
Adam (Andersson et al., 2011)	General population	Semistructured interview, IIEF-5	Erectile dysfunction	Guided CBT (39)	Waitlist with discussion forum (39)	7 modules/7 weeks	IIEF-5	IIEF, RAS, BDI, BAI	6 (14.2%)	3 (8.6%)
Challenger (Miloff, Marklund, & Carlbring, 2015)	General population	MINI	Social anxiety disorder	Unguided CBT with smartphone application (68), unguided bibliotherapy (70)	Waitlist (69)	9 modules/6 weeks	LSAS-SR	GAD-7, PHQ-9, QOLI, BBQ, Mini-SPIN	13 (9.4%)	6 (8.8%)
Stella (Andersson et al., 2013)	General population	SCID-I & II	Depression	Guided CBT (33), group therapy (36), guided CBT as preferred choice (16)	n.a.	8 modules/8 weeks, 8 sessions	BDI	MADRS-S, BAI, HAM-D, QOLI	4 (4.7%)	n.a.
Depressionshjälpen (Carlbring et al., 2013)	General population	SCID-I	Depression	Guided CBT (40)	Waitlist (40)	7 modules/8 weeks	BDI	MADRS-S, BAI, QOLI, WAI	2 (5.2%)	10 (23.8%)
Tellus (Ivarsson et al., 2014)	General population	CAPS	Posttraumatic stress disorder	Guided CBT (31)	Waitlist with support (31)[d]	8 modules/8 weeks	IES-R	PDS, BDI, BAI, QOLI	2 (6.5%)	6 (19.7%)
Klara (Vernmark et al., 2010)	General population	SCID-I	Depression	Guided CBT (29), guided CBT via email (30)	Waitlist (29)	7 modules/8 weeks	BDI	MADRS-S, BAI, QOLI	1 (2.1%)	4 (13.8%)
Oroshjälpen (Dahlin et al., 2016)	General population	SCID-I	Generalized anxiety disorder	Guided ACT (52)	Waitlist (51)	7 modules/9 weeks	GAD-7	PSWQ, GAD-Q-IV, BAI, MADRS-S, PHQ-9, QOLI	1 (2.1%)	11 (21%)
Nova 1 (Carlbring et al., 2011)	General population	SCID-I	Anxiety disorder with/without comorbid depression	Guided CBT (53)	n.a.	10 modules/10 weeks[e]	BAI	MADRS-S, CORE-OM, QOLI	2 (3.8%)	n.a.
Nova 2 (Nordgren et al., 2014)	Primary care	SCID-I	Anxiety disorder with/without comorbid depression	Guided CBT (51)	Waitlist (50)	10 modules/10 weeks[f]	BAI	MADRS-S, CORE-OM, QOLI	3 (6.4%)	2 (4.6%)
Origo 1 (Almlöv et al., 2011)	General population	SCID-I	Generalized anxiety disorder	Guided CBT (44)	Waitlist (45)	8 modules/8 weeks	PSWQ	GAD-Q-IV, STAI, BAI, BDI, MADRS-S, QOLI	4 (8.1%)	12 (27.5%)
Origo 2 (Andersson et al., 2012)	General population	SCID-I	Generalized anxiety disorder	Guided CBT (27)[g]	Waitlist (27)	8 modules/8 weeks	PSWQ	GADQ-IV, MADRS-S, BDI, BAI, STAI, QOLI	1 (5%)	4 (15.7%)

(*table continues*)

Table 2 (*continued*)

Study	Recruitment	Screening interview	Primary diagnosis	Treatment (*n*)	Control (*n*)	Modules/weeks or sessions	Primary outcome	Additional outcomes	Deterioration treatment *n* (%)	Deterioration control *n* (%)
Panik 2 (Carlbring et al., 2005)	General population	SCID-I, CIDI	Panic disorder	Guided CBT (25), face-to-face CBT (24)	n.a.	10 modules/10 weeks, 10 sessions	BSQ	ACQ, MI, BAI, BDI, QOLI	0 (0%)	n.a.
Pia (Unpublished)	General population	SCID-I	Relationship problems	Guided CBT (80)	Waitlist with discussion forum (78)	10 modules/10 weeks	DAS	MSI, BDI, BAI, QOLI	14 (18%)	19 (24.5%)
Progredi, (Ström et al., 2013)	General population	SCID-I	Depression	Guided CBT with physical activity (24)	Waitlist (24)	9 modules/9 weeks	MADRS-S	BDI, BAI, QOLI[c]	2 (8.3%)	3 (12.5%)
Fobal orm (Andersson et al., 2013)	General population	SCID-I	Specific phobia	Guided CBT (13), face-to-face CBT (13)	n.a.	4 modules/4 weeks, 2 sessions[h]	SNAQ	ADIS, FSS, BAI, BDI	0 (0%)	n.a.
Sofie 13 (Boettcher et al., 2013)	General population	SCID-I	Social anxiety disorder	Guided CBT with CBM (61), Guided CBT without CBM (65)	n.a.	9 modules/9 weeks[i]	LSAS-SR	SIAS, SPS, MADRS-S, QOLI	7 (5.8%)	n.a.
Sofie 9 (Kuckertz et al., 2014)	General population	SCID-I & II	Social anxiety disorder	Guided CBT (40)	Attention bias modification (39)	9 modules/9 weeks	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADRS-S, QOLI	0 (0%)	1 (2.6%)
Sofie 12 (Dagöö et al., 2014)	General population	SCID-I, MINI	Social anxiety disorder	Guided CBT with smartphone application (24)[j]	n.a.	9 modules/9 weeks	LSAS-SR[k]		2 (8.3%)	n.a.
Sofie 1 (Andersson et al., 2006)	General population	SCID-I	Social anxiety disorder	Guided CBT (32)	Waitlist (32)	9 modules/9 weeks	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADRS-S, QOLI	0 (0%)	9 (28.1%)
Sofie 2 (Carlbring et al., 2007)	General population	SCID-I	Social anxiety disorder	Guided CBT (29)	Waitlist (28)	9 modules/9 weeks	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADRS-S, QOLI	1 (3.5%)	10 (35.7%)
Sofie 3 (Tillfors et al., 2008)	Students	SCID-I	Social anxiety disorder	Guided CBT (19), guided CBT with group sessions (18)	n.a.	9 modules/9 weeks, 5 sessions	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADR-S, QOLI	0 (0%)	n.a.
Sofie 4 (Furmark et al., 2009)	General population	SCID-I & II	Social anxiety disorder	Guided CBT with discussion forum (40), unguided bibliotherapy (40)	Waitlist (40)	9 modules/9 weeks	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADR-S, QOLI	7 (8.8%)	7 (17.5%)
Sofie 5 (Furmark et al., 2009)	General population	SCID I	Social anxiety disorder	Guided AR (29), guided CBT with discussion forum (29), unguided bibliotherapy (29), unguided bibliotherapy with discussion forum (28)	n.a.	9 modules/9 weeks	LSAS-SR	SIAS, SPS, SPSQ, BAI, MADR-S, QOLI	3 (2.6%)	n.a.
Elsa (In preparation)	General population	SCID-I	Anxiety disorder with/without comorbid depression	Guided CBT (33)	Waitlist with support (33)	8 modules/8 weeks[l]	BAI	MADRS-S, CORE-OM, PHQ-9, GAD-7, QOLI	1 (4.1%)	5 (15.2%)
Fobal spindel (Andersson et al., 2009)	General population	SCID-I	Specific phobia	Guided CBT (13), face-to-face CBT (14)	n.a.	5 modules/4 weeks, 2 sessions[h]	SPQ	ADIS, FSS, BAI, BDI	1 (3.7%)	n.a.

(*table continues*)

Table 2 (*continued*)

Study	Recruitment	Screening interview	Primary diagnosis	Treatment (*n*)	Control (*n*)	Modules/weeks or sessions	Primary outcome	Additional outcomes	Deterioration treatment *n* (%)	Deterioration control *n* (%)
Gamble (Carlbring et al., 2012)	General population	n.a.	Gambling disorder	Guided CBT with discussion forum (317)	n.a.	8 modules/8 weeks	NODS	HADS, QOLI	9 (2.9%)	n.a.
Total per condition									122 (5.8%)	130 (17.4%)

Note. IMÅ = Internetbaserad Mindfulnessbehandling av Ångest; ACT = Acceptance and Commitment Therapy; SCID-I = Structural Clinical Interview for *Diagnostic Manual of Statistical and Mental Disorders*, Fourth Edition (*DSM*-IV) Axis I Disorders; NOS = not otherwise specified; FAQ = frequently asked questions; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; QOLI = Quality of Life Inventory; ISI = Insomnia Severity Index; n.a. = not applicable; ACT = acceptance and commitment therapy; LSAS-SR = Liebowitz Social Anxiety Scale–Self-Report; PDSS-SR = Panic Disorder Severity Scale–Self-Report; PHQ-9 = Patient Health Questionnaire (nine items); GAD-7 = Generalized Anxiety Disorder (seven items); BA = behavioral activation; MADRS-S = Montgomery–Åsberg Depression Rating Scale–Self-Report; IIEF-5 = International Index of Erectile Functioning (five items); CBT = cognitive–behavioral therapy; IIEF = International Index of Erectile Functioning; RAS = Relationship Assessment Scale; MINI = The Mini-International Neuropsychiatric Interview; BBQ = Brunnsviken Brief Quality of Life Inventory; Mini-SPIN = Mini-Social Phobia Inventory; HAM-D = Hamilton Rating Scale for Depression; WAI = Working Alliance Inventory; CAPS = Clinician-Administered Posttraumatic Stress Disorder (PTSD) Scale for *DSM*-IV; IES-R = Impact of Event Scale–Revised; PDS = Posttraumatic Diagnostic Scale; PSWQ = Penn State Worry Questionnaire; GAD-Q-IV = Generalized Anxiety Disorder Questionnaire; CORE-OM = Clinical Outcome in Routine Evaluation–Outcome Measure; STAI = State-Trait Anxiety Inventory; CIDI = Composite International Diagnostic Interview; ACQ = Agoraphobic Cognitions Questionnaire; BSQ = Body Sensations Questionnaire; MI = Mobility Inventory; DAS = Dyadic Adjustment Scale; MSI = Marital Status Inventory; SNAQ = Snake Anxiety Questionnaire; ADIS = Anxiety Disorders Interview Schedule; FSS = Fear Survey Schedule; CBM = cognitive bias modification; SIAS = Social Interaction Anxiety Scale; SPS = Social Phobia Scale; SPSQ = Social Phobia Screening Questionnaire; AR = applied relaxation; SPQ = Spider Phobia Questionnaire; NODS = The National Opinion Research Center (NORC) Diagnostic Screen for Gambling Problems.
[a] Separate analyses of deterioration were conducted for the two primary outcome measures depending on the diagnosis of the patient. [b] Four treatment conditions were included in the study, with/without treatment rationale, respectively, but are pooled in the current analysis. [c] An additional outcome measure, the International Physical Activity Questionnaire, was used in the study, but is not included in the current analysis. [d] Passive control with the possibility to contact the research team if needed. [e] Patients were able to choose 10 out of 16 modules to be completed during 10 weeks. [f] Patients were able to choose 10 out of 19 modules to be completed during 10 weeks. [g] An additional treatment group, guided psychodynamic therapy, was also used in the study, but is not included in the current analysis. [h] One brief orientation session and one session of 3-hr prolonged exposure. [i] In addition to 2 weeks of CBM. [J] An additional treatment group, interpersonal psychotherapy, was also used in the study, but is not included in the current analysis. [k] Additional outcome measures were included in the original study but lost in the raw data file. [l] Patients were able to complete up to eight modules selected by the therapist.

(35.7%) and 9 (28.1%), and one clinical trial of both panic disorder and social anxiety disorder, 13 (28.2%). As for patients in some type of treatment condition, comparable numbers were detected in one clinical trial for relationship problems, 14 (18%), one for erectile dysfunction, 6 (14.2%), and one involving social anxiety disorder, 13 (9.4%).

Predictors of Deterioration

Using the pooled results from the imputed data sets, predictors of deterioration were explored and reported as *OR*. Because the primary aim of the current study was to investigate the deterioration rate and predictors of deterioration for ICBT, the *OR* was determined separately for patients in treatment and those allocated to some form of control condition, revealing slightly different results. In general, patients in a control condition had higher odds of deteriorating in relative to patients receiving treatment. For patients receiving some type of treatment, having symptom levels that were above average on the primary outcome measure at pretreatment were related to lower odds of deterioration. In addition, civil status was associated with lower odds, suggesting that patients in a relationship were less likely to deteriorate than those who were single. Also, educational level was linked to lower odds, which was similar for age, indicating that patients with at least a university degree or being of older age were less apt to deteriorate. Prior psychological treatment, prior or ongoing psychotropic medication, as well as sick leave did not yield any results. As for patients assigned to a control condition, only clinical severity at pretreatment was related to deterioration, which also implies that higher symptom levels were associated with lower odds of deteriorating. In addition, a post hoc analysis of the three categories of psychiatric disorders, that is, anxiety disorders, depression, and other, revealed no differences, 95% confidence interval (CI) [−0.72, 0.81] and [−0.67, 1.09] suggesting that none was related to higher odds of deterioration. Likewise, gender was not associated with faring worse either, 95% CI [−0.31, 0.31]. All predictors with their respective *OR* and 95% CI can be obtained in Table 4.

Discussion

The current study examined the occurrence of deterioration among patients with a variety of different psychiatric disorders receiving ICBT or being allocated to a control condition. Raw scores from 2,866 patients in 29 clinical trials were included, using the RCI as a measure of deterioration. Among those receiving some type of treatment, 122 (5.8%) deteriorated, compared to a larger proportion for those allocated to some form of control, 130 (17.4%). The numbers parallel those in investigations of face-to-face psychotherapy (cf. Whipple et al., 2003), and are comparable to recent research of ICBT (cf. Buntrock et al., 2015), thus resembling the deterioration effect originally proposed by Bergin (1966). In terms of the different clinical trials, most cases involving a treatment condition had similar deterioration rates, except for a few examples with notably higher proportions of deteriorated patients. For instance, both relationship problems and erectile dysfunction stood out in the analysis, and may be comprised of conditions that are more prone to result in deterioration during the treatment period, possibly due to difficulties of intervening in a dyadic constellation via the Internet, as well as somatic issues and

Table 3
Sociodemographic Characteristics of Patients Included in the Analysis

Baseline characteristic	Treatment (n = 2,118)	Control (n = 748)	Full sample (n = 2,866)	Missing data
Gender: n (% female)	1,299 (62.1)	501 (67)	1,800 (63.4)[a]	27 (.9)[b]
Age (years): M (SD)	38 (12.5)	40.6 (13.2)	38.7 (12.8)	29 (1)
Civil status: n (%)				744 (27)
Single	497 (33.5)	171 (28.1)	668 (31.9)	
Relationship	986 (66.5)	438 (71.9)	1,424 (68.1)	
Children: n (% yes)	554 (53.4)	226 (59)	780 (55)	1,446 (50.5)
Cohabitant: n (% yes)	306 (66.5)	48 (69.6)	354 (12.4)	2,337 (81.5)
Highest educational level: n (%)				1,099 (38.3)
Elementary school	53 (4.5)	33 (5.7)	86 (4.9)	
High school/college	361 (30.5)	169 (28.9)	530 (30)	
University	757 (64)	380 (65.1)	1,137 (64.3)	
Postgraduate	12 (1)	2 (.3)	14 (.8)	
Employment: n (%)				1,968 (68.7)
Unemployed	74 (10.8)	19 (9)	93 (10.4)	
Student	99 (14.4)	39 (18.6)	138 (15.4)	
Employed	469 (68.2)	138 (65.7)	607 (67.6)	
Other	13 (1.9)	11 (5.2)	24 (2.7)	
Retired	33 (4.8)	3 (1.4)	36 (4)	
Primary diagnosis: n (%)				88 (3.1)
Anxiety disorders	1,148 (55.8)	533 (74.1)	1,681 (60.5)	
Generalized anxiety disorder	141 (6.8)	138 (19.2)	279 (10)	
Social anxiety disorder	708 (34.4)	257 (35.7)	965 (34.7)	
Anxiety disorder NOS	11 (.5)	20 (2.8)	31 (1.1)	
Panic disorder (with/without agoraphobia)	86 (4.2)	30 (4.2)	116 (4.2)	
Posttraumatic stress disorder	32 (1.6)	32 (4.5)	64 (2.3)	
Anxiety disorder (with/without depression)	117 (5.7)	56 (7.8)	173 (6.2)	
Specific phobia	53 (2.6)	0 (0)	53 (1.9)	
Depression (with/without dysthymia)	475 (23.1)	69 (9.6)	544 (19.6)	
Other	436 (21.2)	117 (16.3)	553 (19.9)	
Erectile dysfunction	39 (1.9)	39 (5.4)	78 (2.8)	
Relationship problems	80 (3.9)	78 (10.8)	158 (5.7)	
Gambling disorder	317 (15.4)	0 (0)	317 (11.4)	
Sick leave: n (% yes)	42 (5.5)	25 (7.6)	67 (6.1)	1,768 (61.7)
Previous psychological treatment: n (% yes)	575 (54.1)	214 (56.5)	789 (54.7)	1,424 (49.7)
Previous or ongoing psychotropic medication: n (% yes)	366 (31.7)	156 (33.1)	522 (32.1)	1,239 (43.2)
Satisfaction with treatment: M (SD)[c]	2.9 (1)	n.a.	2.9 (1)	1,867 (88.2)[e]
Treatment credibility: M (SD)[d]	7 (2.4)	n.a.	7 (2.4)	1,535 (72.5)[e]
Modules completed: M (SD)[f]	6.5 (1.3)	n.a.	6.5 (1.3)	1,194 (56.4)[e]
Time per week: M (SD)[g]	3.6 (3.1)	n.a.	3.6 (3.1)	1,722 (81.3)[e]

Note. n.a. = not applicable; NOS = not otherwise specified.
[a] Valid percent, that is, percent of available data, excluding missing data. [b] Percent, that is, percent of complete dataset, including missing data. [c] Self-rated 0–5. [d] Self-rated 0–10. [e] Based on patients receiving treatment. [f] Weighted mean and standard deviation. [g] Number of hours per week.

the risk of stigmatization in sexual dysfunctions. With regard to one clinical trial of social anxiety disorder that also evidenced more deterioration, it was one of few instances that did not include any guidance from a therapist, which has been associated with a weaker treatment outcome, increased dropout, and negative effects for some patients (Andersson & Titov, 2014), suggesting that unguided ICBT should be accompanied by continuous symptom monitoring to detect and reverse a negative treatment trend, similar to what has been advised for face-to-face psychotherapy (Boswell, Kraus, Miller, & Lambert, 2015). In comparison, when inspecting cases involving a control condition, there was greater variability between the different clinical trials, with a majority being below the average deterioration rate, but as many as eight (44.4%) with at least one fifth of the patients deteriorating. However, no apparent pattern could be found in relation to what may induce this deterioration, except that most entailed patients with anxiety disorders.

Deterioration Rates

The large difference in proportion of deteriorated patients between ICBT and those assigned to a control condition can be considered an unexpected result, particularly given prior research indicating that a large number of patients on waitlist frequently seem to improve. Hesser, Weise, Rief, and Andersson (2011), in a meta-analysis of waitlist control groups in clinical trials of tinnitus distress, found that during a waiting period of 6–12 weeks, patients experienced a mean reduction of 3–8% on tinnitus related outcome measures, equivalent to Hedge's g of 0.17. Similarly, a meta-analysis by Posternak and Miller (2001) of patients with major depression revealed a mean decrease in symptoms of 10–15% during a time span of 2–20 weeks, arguing that up to one fifth of all waitlist patients might be achieving spontaneous remission without receiving any treatment. These findings are, however, based on the average change between two points of measurement,

Table 4
Odds Ratios for Each Predictor Variable Using the Full Imputed Model and Divided by Treatment and Control

Predictor (reference)	Treatment				Control			
	OR	Lower CI	Upper CI	*p*	*OR*	Lower CI	Upper CI	*p*
Clinical severity at pretreatment (lower severity)	.62	.50	.77	.00*	.64	.51	.80	.00*
Civil status, single/relationship (single)	.58	.35	.95	.03*	.76	.43	1.34	.34
Prior psychological treatment, yes/no (no)	1.28	.73	2.23	.39	1.30	.66	2.58	.45
Prior or ongoing psychotropic medication, yes/no (no)	1.05	.58	1.90	.86	1.41	.74	2.68	.30
Sick leave, yes/no (no)	2.24	.60	8.36	.23	2.14	.62	7.33	.23
Educational level, less than university/at least university (less than university)	.54	.33	.88	.01*	1.02	.60	1.75	.94
Age (lower age)	.78	.62	.98	.03*	.83	.66	1.05	.11

Note. *OR* = odds ratio; CI = 95% confidence interval.
* $p < .05$.

possibly missing individual cases of deterioration due to the variability of the scores. There is in fact evidence that some patients on waitlist can fare worse during a waiting period, for instance, partners receiving behavioral couple therapy (Baucom, Hahlweg, & Kuschel, 2003). Likewise, using data from primary care, Young (2006) discovered that, although most patients remain stable over time or even improve, 3.5% deteriorated on the Clinical Outcomes in Routine Evaluation (Evans et al., 2000) during a mean wait of 29.4 days. As for ICBT, Lancee, Eisma, van Straten, and Kamphuis (2015) revealed that proportionately more patients deteriorated on waitlist, 31.8%, compared to those receiving treatment for insomnia, 9.3%. Whether improvement or deterioration is more prevalent among patients on waitlist warrants further examination. However, the latter may occur and should be recognized by researchers, for instance, by supervising patients' progress and immediately initiate treatment if someone starts to deteriorate. Furthermore, it may be necessary to differentiate between different forms of control, in which more passive conditions, such as waitlist, result in greater rates of deterioration than those that are more active in nature. Because the current study included clinical trials that almost exclusively utilized waitlist as a comparator, this was not feasible to investigate and should be explored in the future. In addition, the mechanisms associated with deterioration among patients assigned to some form of control also need to be assessed. Furukawa et al. (2014), conducting a network meta-analysis to compare different control conditions to CBT for major depression, found far larger between group effect sizes when using waitlist than both no treatment and psychological placebo, suggesting that this could be caused by patients allocated to waitlist being more inclined to remain depressed to receive the treatment intervention after the waiting period is over. Repeated administration of outcome measures may also be a plausible explanation for deterioration, although it has yet been studied only with regard to improvement. Sharpe and Gilbert (1998) have, for instance, found a significant decrease in negative mood between completions of the Beck Depression Inventory (Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961), despite the absence of any treatment intervention. Similar results were obtained by Arrindell (2001), using several different outcome measures as well as both inpatients and outpatients, proposing that it could be caused by such factors as improved self-monitoring, social desirability, and habituation. However, with regard to deterioration this relationship is less clear, but might be attributable to a greater awareness of one's current situation and ongoing distress, which, in turn, results in deterioration when being assigned to waitlist instead of treatment. More research is required to comprehend what is causing patients to deteriorate while waiting, possibly by using more qualitative methods, such as, semistructured interviews regarding the experiences during the waiting period.

Predictors of Deterioration

The current study also examined possible predictors of deterioration using a set of variables that were selected a priori based on theory or prior empirical findings. The results indicated that for patients assigned to ICBT the odds of deteriorating was lower if the individual was in a relationship, had at least a university degree, or being of older age. This is in line with the findings of Karyotaki et al. (2015), where a lower educational level was related to an increased risk of dropout from psychological treatments delivered via the Internet, while every additional 4 years of age decreased this probability. The reason for why these sociodemographic variables might be associated with less deterioration is unclear, although it is reasonable to assume that they are related to greater resilience, such as, via greater social support, better problem solving skills, and higher socioeconomic status (cf. Davydov, Stewart, Ritchie, & Chaudieu, 2010). Older age has for instance been found to reduce the reported exposure to daily stressors as well as diminish their impact in terms of negative affect, as compared to younger individuals (Stawski, Sliwinski, Almeida, & Smyth, 2008). Educational level has also been found to improve adherence during ICBT (Waller & Gilbody, 2009), possibly by making the treatment content easier to comprehend. Martinez, Whitfield, Dafters, and Williams (2008), in a review of commonly prescribed English self-help manuals for depression revealed that the reading ability required to fully understand the texts is often much higher than the literacy level of many patients, which might result in difficulties understanding the treatment content provided, and, in turn, lowered self-esteem and deterioration. As for civil status, previous investigations indicate that having a significant other appears to serve as a protective factor for mental distress (cf. Scott et al., 2010), which might explain the reduced odds of

deterioration in the current study. However, in contrast to one of the initial hypotheses, clinical severity at pretreatment was not related to increased odds of deterioration, neither for patients in a treatment or control condition, instead revealing that higher symptom levels appear to lower the odds. This runs contrary to the findings of Christensen et al. (2009), stating that clinical severity at pretreatment was a predictor of dropout, and opposes the idea that distress may predict deterioration and, in turn, increase the risk of dropping out. One explanation for this might be that patients with higher symptom levels regress to the mean regardless of being in a treatment or control condition, which potentially explains the decreased odds for deterioration. Regression to the mean is, in fact, not uncommon in clinical trials, and some patients with clinical severity at pre assessment that is considerably above average might demonstrate a reduction of scores on a specific outcome measure attributable to measurement error rather than an improvement in symptomatology. Hsu (1995) has presented ways of handling regression to the mean that could be implemented when comparing both improvement and deterioration rates, although a later review of different methods for determining the RCI seems to suggest that the original proposition by Jacobson and Truax (1991) should yield similar results (Bauer et al., 2004). Assessments taken at follow-up could, however, be used to examine the implications of regression to the mean, as well as to infer whether the deterioration occurring during treatment is sustained over time or just a transient phenomenon. Although, another explanation is that higher symptom levels are not necessarily associated with a poorer or negative outcome. For instance, in an IPDM on depression by Bower et al. (2013), severely depressed patients were shown to benefit as much from their treatment as those that are less depressed. This does not preclude some patients with higher symptom levels at pretreatment from deteriorating, but the relationship might not be strong enough to predict deterioration among patients in general. Alternatively, ceiling effects could potentially account for this effect, which may seem likely given the fact that the scoring of a self-report measure prohibits the patient from deteriorating indefinitely. However, a closer inspection of the number of patients that could not exceed the RCI due to their pretreatment scores revealed that only 55 (1.9%) were close to hitting the ceiling, and that a majority of these cases improved, suggesting that ceiling effects were quite unlikely and should not have affected the results. In sum, a number of predictors of deterioration in ICBT seem to exist and are primarily related to such sociodemographic variables as educational level, age, and civil status, lowering the odds for patients receiving some type of treatment. Additionally, clinical severity at pretreatment also appears to be associated with deterioration, decreasing the odds for both treatment and control conditions. Further research is, however, required to replicate these findings and establish their clinical significance in terms of screening and monitoring procedures in ICBT. The current study, however, suggests that deterioration does occur and should be considered in future clinical trials as well as in the regular health care system using the Internet as a way of disseminating evidence-based methods to patients with different psychiatric disorders.

Study Limitations

The current study investigated the occurrence and predictors of deterioration in ICBT using IPDM as a method for collecting, assembling and performing statistical analyses on raw scores from a large number of clinical trials. Albeit a powerful way of examining effects that goes beyond group means and standard deviations, this approach also has several limitations that require careful consideration when reviewing the results. First, to the knowledge of the authors, there has been no similar attempt to explore deterioration in ICBT, making it difficult to compare the findings of the current study to those of others, warranting more research before any definite conclusions can be drawn. In terms of the deterioration rates reported, the numbers are, however, analogous to what has previously been observed. An average deterioration rate of 5–10% is often referred to as a reasonable estimate in face-to-face settings (Lambert, 2007). As for ICBT, the only systematic review of this issue indicates that worsening was present among 0–5% of patients receiving treatment, compared to 2–9% in the comparison groups, although only 8 of the 40 (20%) studied randomized controlled trials contained any information regarding deterioration (Arnberg, Linton, Hultcrantz, Heintz, & Jonsson, 2014). Given that individual patient data from numerous clinical trials was used in the current study, this should give a more precise estimate of deterioration than has previously been presented for ICBT. In relation to the predictors that have been proposed, the lack of comparable investigations specifically targeting deterioration warrants precaution as to their validity, wherefore more research is recommended to assess their clinical significance. Also, it should be noted that the selection of predictors, which was made a priori, could mean that some aspects that are related to deterioration were missed, which might warrant further investigation in the future. This includes assessments of other types of sociodemographic variables that were not explored, such as, socioeconomic status measured by average income, or more information regarding the patients' conditions, for instance, comorbidity. Likewise, dosage and adherence could also be related to deterioration. It has, for instance, been found that the completion of more modules in ICBT is associated with superior treatment outcome (cf. Hadjistavropoulos, Pugh, Hesser, & Andersson, 2016), indicating that the dosage may have relationship with deterioration too, or, at least with nonresponse. Second, data used in the IPDM were comprised of a wide range of patients, treatment modalities, and psychiatric disorders, which, in some respect, can be considered quite heterogeneous in nature and being suitable for the purpose of the current study. However, because it did not involve a systematic review approach, this also introduces the risk of availability bias with regard to what clinical trials were included (Clarke, 2005). Predefined inclusion and exclusion criteria were used to prevent this, and great measures were taken to assemble as many data sets as possible, but the risk of some clinical trials being left out of the analysis is still possible. Also, only the first author reviewed the original data and imputed it into the main data set used in the current study, potentially creating reviewer bias (L. Stewart & Tierney, 2002), which could have been circumvented by including an additional reviewer, preferably without any bindings to the authors of the current study. Furthermore, because all data were derived from sites managed by the fourth and fifth author, similar treatment content and means of delivery were used in most

of the clinical trials. Hence, the results might not be representative and hard to generalize to ICBT at large, as there exist a large number of treatments delivered via the Internet for different psychiatric disorders, at least when comparing different countries. There might, therefore, exist a variation in deterioration rates between sites that examine the efficacy of ICBT, which should be investigated further, for instance, by having researchers report deterioration rates at a more regular basis when providing information about the treatment outcome (Rozental et al., 2014). However, given that the data collected for the purpose of the current study involve a majority of all clinical trials of ICBT performed in Sweden, the results should be characteristic for how it is being implemented in both university settings and the regular health care system in this country, that is, consistency in language, use of validated outcome measures, having established procedures, and conducting diagnostic interviews. Third, even though the patients included in the IPDM were diverse in some regards, sociodemographic variables indicate that the mean age, gender, civil status, and education level may not be characteristic of patients in general. A majority were in their late thirties, $M = 38.7$ ($SD = 12.8$), female (63.4%), in a relationship (68.1%), and having at least a university degree (65.1%). On the one hand, reports on the characteristics of patients seeking treatment for mental health issues indicate that this is not particularly uncommon (Vessey & Howard, 1993), which would suggest that patients in ICBT are similar to most treatment-seeking individuals. On the other hand, it does merit some caution in relation to the results presented, as prior research implies that deterioration rates differ between age groups and psychiatric disorders (Lambert, 2013), possibly underestimating the number of patients that deteriorate along certain features. However, a greater limitation relates to the means of recruitment in the clinical trials included in the current study, which almost exclusively relied on self-recruitment in the general population. This may have resulted in the enrollment of patients with less severe problems as well as more motivation to complete the treatment program, compared to patients in either primary care or psychiatric outpatient clinics. Hence, the deterioration rates might not be representative of patients that are recruited in the regular health care system, warranting replications in these settings to assess its generalizability. Although, there is some evidence suggesting that patients recruited for ICBT may in fact have more severe conditions than the average population in epidemiological surveys, implying that the normal patient should not differ in terms of clinical severity (Titov, Andrews, Kemp, & Robinson, 2010). Fourth, even though deterioration is often conceived as a reliable way of determining whether or not negative effects have occurred during treatment, the concept could be criticized for lacking both construct and predictive validity. Early on, it was argued that deterioration does not have to be attributable to the treatment per se, making it difficult to establish a cause-effect relationship (May, 1971). It is therefore conceivable that the deterioration exhibited by some patients is in fact related to other circumstances, such as, interpersonal difficulties or occupational stressors, rather than some detrimental impact of the treatment content. Other methods for investigating negative effects is thus necessary, for example, using self-report measures with items conveying a variety of adverse events that are reasonable to arise during treatment, as well as making the patient report what could be responsible for their incidence, as has been proposed in a novel instrument, the Negative Effects Questionnaire (Rozental, Kottorp, Boettcher, Andersson, & Carlbring, 2016). Also, clinical ratings and structured diagnostic interviews may become useful in determining negative effects of psychological treatments that go beyond the use of self-report measures that are completed by the patients. For instance, trained clinicians should be able to evaluate the progress of a particular individual, or, lack thereof, by using the Clinical Global Impressions Scale (Busner & Targum, 2007), a measure that accounts for the change in the patient's condition compared to baseline on a Likert-scale. Both the Structured Clinical Interview for *DSM–IV*-Axis I Disorders (First et al., 1997) and the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998) may also be applied to check whether a psychiatric disorder still exist or not after the treatment period, reflecting improvement or nonresponse. The implementation of these methods could have affected the results in the current study, as subjective responses on a self-report measure may not always reflect true deterioration. However, both are also limited by the fact that they rely on the clinician's perception of progress, something that might not always correspond to the experience of the patient. As discussed by Strupp and Hadley (1977), improvement and deterioration in psychological treatments depend on what perspective you use; the patient's, the clinician's, or society's. Thus, it could be argued that the patient and the clinician regard the development during the treatment differently, which could perhaps be an interesting research question for the future, for example, by using parallel assessments of treatment outcome, that is, ratings from the clinician, the patient, and significant others. Also, to infer if deterioration occurring during the treatment period, or, alternatively, in a control condition, is in fact related to a negative outcome, follow-up measures are required to control for its long-term implication, which can be regarded as a limitation in the current study. It is possible that for some patients, deterioration could have a short and transitory effect that does not prevent the patient to benefit from treatment. Rozental et al. (2015), in a qualitative content analysis of patients' responses to open-ended questions distributed at posttreatment regarding negative effects in ICBT, found that certain patients experienced deterioration and novel symptoms in the beginning of the treatment period, related to becoming more aware of the factors responsible for one's condition and being presented with a sensible rationale. In retrospect, this was, however, not always regarded as something negative, possibly revealing a process through which the patient gains a better understanding of what is maintaining a psychiatric disorder and mourns the many years of distress. Furthermore, it could also be the case that deterioration in ICBT is related to the temporary yet negative experiences that many patients can have with specific interventions, such as, exposure and behavioral experiments, but that are expected to abate and become tolerable with repeated exercise. Hence, it may be the case that deterioration is not always clinically significant, making other methods for evaluating negative effects more important. It could also be argued that the use of 0.84 *SD* units as a cutoff for mild deterioration does not correspond to a change score that is in fact experienced as negative by the patient. On the one hand, this may have increased the risk of identifying patients as being deteriorated when no such effect has occurred. On the other hand, a less strict cutoff may help to reveal negative effects that could otherwise have been missed. More research is thus necessary to conclude if deterioration is a reliable construct

that is associated with enduring negative consequences, for instance, by interviewing patients with regard to their own perspective on deterioration. Also, clinical ratings, structured diagnostic interviews, quality of life, as well as composite measures that accounts for both benefits and risks might become important in differentiating the positive from the negative effects of psychological treatments (Wallace, Frank, & Kraemer, 2013). In other words, the method being used for assessing negative effects in the current study and its ensuing results does not provide enough evidence to refer to ICBT or psychological treatments in general as harmful, or that there are certain treatment interventions that can cause harm, that is, being iatrogenic. The findings only relate to deterioration, and, more specifically, a mild form of worsening that has been determined solely by statistical means. In addition, it does not account for both the benefits and risks in a way that conveys that a treatment can have a positive as well as negative impact. More research thus has to be conducted before it is possible to differentiate between the two, and to provide patients with adequate information to make an informed consent. One such future conceptualization could be the use of a value tree for displaying the benefit–risk balance of particular treatment interventions, as used in pharmacological research (Walker, Liberti, McAuslane, & Levitan, 2011). In such a procedure the advantages of a given treatment are considered together with its disadvantages, for instance, minimizing the chance of stroke but simultaneously increasing the chance of liver failure. A similar approach in relation to psychological treatments might be the use of assessments from various sources to determine the benefits and risks for a patient, for example, negative reactions from significant others but improved self-efficacy as a result of self-assertiveness training. This is, of course, also relevant when considering the predictors of deterioration presented in the current study, as they only relate to worsening and not harm, which is not the same as suggesting that they are prognostic for a negative outcome or can outweigh the many benefits of treatment. Fifth, the varying degree of missing data is a major limitation that may have affected the results in terms of both deterioration rates and predictors, especially for sick leave which had 61.7% of missing data, albeit it should be noted that the degree of missing data was never used as a criterion for the selection of predictors. Recommendations were followed with regard to handling missing data (Debray et al., 2015), but it is still possible that patients that are unable complete the primary outcome measure at postassessment could have done so in a systematic fashion. Assuming that it reflects deterioration, the numbers might thus have been higher than was estimated using imputed values. However, it is not clear if this is reasonable to expect, as noncompletion can be linked to other factors, for example, improvement or relocation. Additional points of measurements would potentially also have created more valid imputations, although this was only seldom used in the clinical trials that were included in the current study. As to the predictors that were selected a priori, it is possible that other variables would have resulted in a different interpretation of what might be causing deterioration in ICBT. Perhaps those that were chosen on theoretical or empirical grounds are not directly related to patients faring worse during treatment or control conditions; whereas, the inclusion of other sociodemographics variables or outcome measures could increase the understanding of what is causing some patients to deteriorate. Likewise, missing data for the predictors can also have affected the results, decreasing the power the find meaningful differences and influencing their validity when data had to be imputed. Further research is therefore warranted when it comes to possible predictors of deterioration, preferably by using additional data sets from even more sites than were possible to include in the current study.

Conclusion

The current study is, to the knowledge of the authors, the first IPDM investigating the occurrence and predictors of deterioration in ICBT, indicating that deterioration rates were higher in the control than treatment conditions. Variables examined in relation to predicting deterioration revealed that older age, being in a relationship, and having a higher education level lowered the odds of deteriorating during treatment, but not for patients in a control condition. Clinical severity at pre assessment was also associated with lower odds of deterioration, but only for patients receiving treatment. The results indicate that deterioration is to be expected to some extent and that monitoring the patients' progress is recommended to detect if someone is faring worse during treatment or being in a control condition. In addition, the predictors of deterioration could be helpful as a screening procedure as a way to identify patients at risk of deteriorating, although more research is needed to replicate the findings and determine their clinical significance.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Almlöv, J., Carlbring, P., Kallqvist, K., Paxling, B., Cuijpers, P., & Andersson, G. (2011). Therapist Effects in Guided Internet-Delivered CBT for Anxiety Disorders. *Behavioural and Cognitive Psychotherapy, 39,* 311–322. http://dx.doi.org/10.1017/S135246581000069X

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (DSM–IV–TR).* Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM–5).* Arlington, VA: Author.

*Andersson, G., Carlbring, P., Holmstrom, A., Sparthan, E., Furmark, T., Nilsson-Ihrfelt, E., . . . Ekselius, L. (2006). Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 677–686. http://dx.doi.org/10.1037/0022-006x.74.4.677

*Andersson, G., Hesser, H., Veilord, A., Svedling, L., Andersson, F., Sleman, O., . . . Carlbring, P. (2013). Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. *J Affect Disord, 151,* 986–994. http://dx.doi.org/10.1016/j.jad.2013.08.022

*Andersson, G., Paxling, B., Roch-Norlund, P., Östman, G., Norgren, A., Almlov, J., . . . Silverberg, F. (2012). Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. *Psychotherapy amd Psychosomatics, 81,* 344–355. http://dx.doi.org/10.1159/000339371

Andersson, G., & Titov, N. (2014). Advantages and limitations of Internet-based interventions for common mental disorders. *World Psychiatry: Official Journal of the World Psychiatric Association, 13,* 4–11. http://dx.doi.org/10.1002/wps.20083

*Andersson, G., Waara, J., Jonsson, U., Malmaeus, F., Carlbring, P., & Ost, L. G. (2009). Internet-based self-help versus one-session exposure in the treatment of spider phobia: a randomized controlled trial. *Cogn Behav Ther, 38,* 114–120. http://dx.doi.org/10.1080/16506070902931326

*Andersson, G., Waara, J., Jonsson, U., Malmaeus, F., Carlbring, P., & Öst, L.-G. (2013). Internet-Based Exposure Treatment Versus One-Session Exposure Treatment of Snake Phobia: A Randomized Controlled Trial. *Cognitive Behaviour Therapy, 42,* 284–291. http://dx.doi.org/10.1080/16506073.2013.844202

*Andersson, E., Walen, C., Hallberg, J., Paxling, B., Dahlin, M., Almlov, J., . . . Andersson, G. (2011). A Randomized Controlled Trial of Guided Internet-delivered Cognitive Behavioral Therapy for Erectile Dysfunction. *Journal of Sexual Medicine, 8,* 2800–2809. http://dx.doi.org/10.1111/j.1743-6109.2011.02391.x

Arnberg, F. K., Linton, S. J., Hultcrantz, M., Heintz, E., & Jonsson, U. (2014). Internet-delivered psychological treatments for mood and anxiety disorders: A systematic review of their efficacy, safety, and cost-effectiveness. *PLoS ONE, 9,* e98118. http://dx.doi.org/10.1371/journal.pone.0098118

Arrindell, W. A. (1993). The fear of fear concept: Stability, retest artifact and predictive power. *Behaviour Research and Therapy, 31,* 139–148.

Arrindell, W. A. (2001). Changes in waiting-list patients over time: Data on some commonly-used measures. *Beware! Behaviour Research and Therapy, 39,* 1227–1247. http://dx.doi.org/10.1016/S0005-7967(00)00104-2

Barlow, D. H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist, 65,* 13–20. http://dx.doi.org/10.1037/a0015643

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Baucom, D. H., Hahlweg, K., & Kuschel, A. R. (2003). Are waiting-list control groups needed in future marital therapy outcome research? *Behavior Therapy, 34,* 179–188. http://dx.doi.org/10.1016/S0005-7894(03)80012-6

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment, 82,* 60–70. http://dx.doi.org/10.1207/s15327752jpa8201_11

Beck, A. T., & Steer, R. A. (1996). *Beck Depression Inventory manual, Svensk version (Swedish version)*. Fagernes, Norway: Psykologiförlaget AB.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571. http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

Bergin, A. E. (1966). Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology, 71,* 235–246. http://dx.doi.org/10.1037/h0023577

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a Clinician-Administered PTSD Scale. *Journal of Traumatic Stress, 8,* 75–90. http://dx.doi.org/10.1002/jts.2490080106

*Boettcher, J., Astrom, V., Pahlsson, D., Schenstrom, O., Andersson, G., & Carlbring, P. (2014). Internet-Based Mindfulness Treatment for Anxiety Disorders: A Randomized Controlled Trial. *Behavior Therapy, 45,* 241–253.

*Boettcher, J., Hasselrot, J., Sund, E., Andersson, G., & Carlbring, P. (2013). Combining Attention Training with Internet-Based Cognitive-Behavioural Self-Help for Social Anxiety: A Randomised Controlled Trial. *Cognitive Behaviour Therapy, 43,* 34–48. http://dx.doi.org/10.1080/16506073.2013.809141

Boettcher, J., Rozental, A., Andersson, G., & Carlbring, P. (2014). Side effects in Internet-based interventions for social anxiety disorder. *Internet Interventions, 1,* 3–11. http://dx.doi.org/10.1016/j.invent.2014.02.002

Boisvert, C. M., & Faust, D. (2002). Iatrogenic symptoms in psychotherapy. A theoretical exploration of the potential impact of labels, language, and belief systems. *American Journal of Psychotherapy, 56,* 244–259.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25,* iii. http://dx.doi.org/10.1080/10503307.2014.942531

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., . . . Liu, E. T. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: Meta-analysis of individual patient data. *British Medical Journal (Clinical Research Ed.), 346,* f540. http://dx.doi.org/10.1136/bmj.f540

Buntrock, C., Ebert, D., Lehr, D., Riper, H., Smit, F., Cuijpers, P., & Berking, M. (2015). Effectiveness of a web-based cognitive behavioural intervention for subthreshold depression: Pragmatic randomised controlled trial. *Psychotherapy and Psychosomatics, 84,* 348–358. http://dx.doi.org/10.1159/000438673

Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry, 4,* 28–37.

Carey, M. P., Spector, I. P., Lantinga, L. J., & Krauss, D. J. (1993). Reliability of the Dyadic Adjustment Scale. *Psychological Assessment, 5,* 238–240.

*Carlbring, P., Degerman, N., Jonsson, J., & Andersson, G. (2012). Internet-based treatment of pathological gambling with a three-year follow-up. *Cognitive Behaviour Therapy, 41,* 321–334. http://dx.doi.org/10.1080/16506073.2012.689323

*Carlbring, P., Gunnarsdóttir, M., Hedensjö, L., Andersson, G., Ekselius, L., & Furmark, T. (2007). Treatment of social phobia from a distance: A randomized trial of internet delivered cognitive behavior therapy (CBT) and telephone support. *British Journal of Psychiatry, 190,* 123–128

*Carlbring, P., Hägglund, M., Luthström, A., Dahlin, M., Kadowaki, Å., Vernmark, K., & Andersson, G. (2013). Internet-based behavioral activation and acceptance-based treatment for depression: A randomized controlled trial. *J Affect Disord, 148,* 331–337. http://dx.doi.org/10.1016/j.jad.2012.12.020

*Carlbring, P., Lindner, P., Martell, C., Hassmen, P., Forsberg, L., Strom, L., & Andersson, G. (2013). The effects on depression of Internet-administered behavioural activation and physical exercise with treatment rationale and relapse prevention: Study protocol for a randomised controlled trial. *Trials, 14*. http://dx.doi.org/10.1186/1745-6215-14-35

*Carlbring, P., Maurin, L., Törngren, C., Linna, E., Eriksson, T., Sparthan, E., . . . Andersson, G. (2011). Individually-tailored, Internet-based treatment for anxiety disorders: A randomized controlled trial. *Behaviour Research and Therapy, 49,* 18–24. http://dx.doi.org/10.1016/j.brat.2010.10.002

*Carlbring, P., Nilsson-Ihrfelt, E., Waara, J., Kollenstam, C., Buhrman, M., Kaldo, V., . . . Andersson, G. (2005). Treatment of panic disorder: live therapy vs self-help via the Internet. *Behaviour Research and Therapy, 43,* 1321–1333. http://dx.doi.org/10.1016/j.brat.2004.10.002

Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist, 65,* 34–49. http://dx.doi.org/10.1037/a0017330

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in Internet interventions for anxiety and depression: Systematic review. *Journal of Medicine Internet Research, 11,* e13.

Clarke, M. J. (2005). Individual patient data meta-analyses. *Best Practice & Research Clinical Obstetrics & Gynaecology, 19,* 47–55. http://dx.doi.org/10.1016/j.bpobgyn.2004.10.011

Cooper, A. A., & Conklin, L. R. (2015). Dropout from individual psychotherapy for major depression: A meta-analysis of randomized clinical trials. *Clinical Psychology Review, 40,* 57–65. http://dx.doi.org/10.1016/j.cpr.2015.05.001

Cuijpers, P., Donker, T., van Straten, A., Li, J., & Andersson, G. (2010). Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychological Medicine, 40,* 1943–1957. http://dx.doi.org/10.1017/S0033291710000772

*Dagöö, J., Asplund, R. P., Bsenko, H. A., Hjerling, S., Holmberg, A., Westh, S., . . . Andersson, G. (2014). Cognitive behavior therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: A randomized controlled trial. *Journal of Anxiety Disorders, 28,* 410–417. http://dx.doi.org/10.1016/j.janxdis.2014.02.003

*Dahlin, M., Andersson, G., Magnusson, K., Johansson, T., Sjögren, J., Håkansson, A., . . . Carlbring, P. (2016). Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy, 77,* 86–95. http://dx.doi.org/10.1016/j.brat.2015.12.007

Davydov, D. M., Stewart, R., Ritchie, K., & Chaudieu, I. (2010). Resilience and mental health. *Clinical Psychology Review, 30,* 479–495. http://dx.doi.org/10.1016/j.cpr.2010.03.003

Debray, T. P., Moons, K. G. M., Abo-Zaid, G. M. A., Koffijberg, H., & Riley, R. D. (2013). Individual participant data meta-analysis for a binary outcome: One-stage or two-stage? *PLoS ONE,* 8.

Debray, T. P., Moons, K. G., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H., . . . the GetReal Methods Review Group. (2015). Get real in individual participant data (IPD) meta-analysis: A review of the methodology. *Research Synthesis Methods, 6,* 293–309. http://dx.doi.org/10.1002/jrsm.1160

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist, 65,* 21–33. http://dx.doi.org/10.1037/a0017299

Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test taking and the stability of adjustment scales: Can we assess patient deterioration? *Evaluation Quarterly, 2,* 275–291. http://dx.doi.org/10.1177/0193841X7800200206

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health, 1,* 70–72. http://dx.doi.org/10.1136/ebmh.1.3.70

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health, 9,* 247–255. http://dx.doi.org/10.1080/713680250

Fantino, B., & Moore, N. (2009). The self-reported Montgomery-Åsberg Depression rating scale is a useful evaluative tool in major depressive disorder. *BMC Psychiatry, 9,* 26.

First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. (1997). *Structured clinical interview for DSM–IV Axis I Disorders (SCID-I)*. Washington, DC: American Psychiatric Press.

*Furmark, T., Carlbring, P., Hedman, E., Sonnenstein, A., Clevberger, P., Bohman, B., . . . Andersson, G. (2009). Bibliotherapy and Internet delivered cognitive behaviour therapy with additional guidance for social anxiety disorder: randomised controlled trial. *British Journal of Psychiatry, 195,* 440–447.

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., . . . Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica, 130,* 181–192. http://dx.doi.org/10.1111/acps.12275

Gerstein, D. R., Volberg, R. A., Toce, M. T., Harwood, H., Johnson, R. A., Buie, T., . . . Sinclair, S. (1999). *Gambling impact and behavior study: Report to the National Gambling Impact Study Commission*. Chicago, IL: NORC at the University of Chicago.

Hadjistavropoulos, H. D., Pugh, N. E., Hesser, H., & Andersson, G. (2016). Predicting response to therapist-assisted internet-delivered cognitive behavior therapy for depression or anxiety within an open dissemination trial. *Behavior Therapy, 47,* 155–165. http://dx.doi.org/10.1016/j.beth.2015.10.006

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9,* 329–343. http://dx.doi.org/10.1093/clipsy.9.3.329

Harel, O. (2007). Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology, 4,* 75–89. http://dx.doi.org/10.1016/j.stamet.2006.03.002

Heeren, A., Maurange, P., Rossignol, M., Vanhaelen, M., Peschard, V., Eeckhout, C., & Phillippot, P. (2012). Self-report version of the Liebowitz Social Anxiety Scale: Psychometric properties of the French version. *Canadian Journal of Behavioural Science, 44,* 99–107.

Hesser, H., Weise, C., Rief, W., & Andersson, G. (2011). The effect of waiting: A meta-analysis of waitlist control groups in trials for tinnitus distress. *Journal of Psychosomatic Research, 70,* 378–384. http://dx.doi.org/10.1016/j.jpsychores.2010.12.006

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research, 36,* 427–440. http://dx.doi.org/10.1007/s10608-012-9476-1

Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 141–144. http://dx.doi.org/10.1037/0022-006X.63.1.141

*Ivarsson, D., Blom, M., Hesser, H., Carlbring, P., Enderby, P., Nordberg, R., & Andersson, G. (2014). Guided Internet-delivered cognitive behaviour therapy for post-traumatic stress disorder: A randomized controlled trial. *Internet Interventions, 1,* 33–40. http://dx.doi.org/10.1016/j.invent.2014.03.002

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352. http://dx.doi.org/10.1016/S0005-7894(84)80002-7

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19. http://dx.doi.org/10.1037/0022-006X.59.1.12

Jolani, S., Debray, T. P. A., Koffijberg, H., van Buuren, S., & Moons, K. G. M. (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: A generalized approach using MICE. *Statistics in Medicine, 34,* 1841–1863. http://dx.doi.org/10.1002/sim.6451

Jonsson, U., Alaie, I., Parling, T., & Arnberg, F. K. (2014). Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials, 38,* 1–8. http://dx.doi.org/10.1016/j.cct.2014.02.005

Karyotaki, E., Kleiboer, A., Smit, F., Turner, D. T., Pastor, A. M., Andersson, G., . . . Cuijpers, P. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An "individual patient data" meta-analysis. *Psychological Medicine, 45,* 2717–2726. http://dx.doi.org/10.1017/S0033291715000665

Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric properties of some specific-fear questionnaires. *Behavior Therapy, 5,* 401–409.

*Kuckertz, J. M., Gildebrant, E., Liliequist, B., Karlstrom, P., Vappling, C., Bodlund, O., . . . Carlbring, P. (2014). Moderation and mediation of the effect of attention training in social anxiety disorder. *Behaviour Research and Therapy, 53,* 30–40. http://dx.doi.org/10.1016/j.brat.2013.12.003

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17,* 1–14. http://dx.doi.org/10.1080/10503300601032506

Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50,* 42–51. http://dx.doi.org/10.1037/a0030682

Lancee, J., Eisma, M. C., van Straten, A., & Kamphuis, J. H. (2015). Sleep-related safety behaviors and dysfunctional beliefs mediate the efficacy of online CBT for insomnia: A randomized controlled trial. *Cognitive Behaviour Therapy, 44,* 406–422. http://dx.doi.org/10.1080/16506073.2015.1026386

Lee, E. H., Kim, J. H., & Yu, B. H. (2009). Reliability and validity of the self-report version of the Panic Disorder Severity Scale in Korea. *Depression and Anxiety, 26,* E120–E123.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry, 22,* 141–173. http://dx.doi.org/10.1159/000414022

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2,* 53–70. http://dx.doi.org/10.1111/j.1745-6916.2007.00029.x

Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology & Psychotherapy, 20,* 286–296. http://dx.doi.org/10.1002/cpp.1765

*Lindner, P., Ivanova, E., Ly, K., Andersson, G., & Carlbring, P. (2013). Guided and unguided CBT for social anxiety disorder and/or panic disorder via the Internet and a smartphone application: study protocol for a randomised controlled trial. *Trials, 14,* 437. http://dx.doi.org/10.1186/1745-6215-14-437

Martinez, R., Whitfield, G., Dafters, R., & Williams, C. (2008). Can people read self-help manuals for depression? A challenge for the stepped care model and book prescription schemes. *Behavioural and Cognitive Psychotherapy, 36,* 89–97.

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on "assessing clinical significance." *Psychotherapy Research, 6,* 124–132. http://dx.doi.org/10.1080/10503309612331331648

May, P. R. A. (1971). For better or for worse? Psychotherapy and variance change: A critical review of the literature. *Journal of Nervous and Mental Disease, 152,* 184–192. http://dx.doi.org/10.1097/00005053-197103000-00004

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments. A review of current efforts. *American Psychologist, 65,* 73–84. http://dx.doi.org/10.1037/a0018121

*Miloff, A., Marklund, A., & Carlbring, P. (2015). The Challenger app for social anxiety disorder: New advances in mobile psychological treatment. *Internet Interventions, 2,* 382–391. http://dx.doi.org/10.1016/j.invent.2015.08.001

Mohr, D. C., Beutler, L. E., Engle, D., Shoham-Salomon, V., Bergan, J., Kaszniak, A. W., & Yost, E. B. (1990). Identification of patients at risk for nonresponse and negative outcome in psychotherapy. *Journal of Consulting and Clinical Psychology, 58,* 622–628. http://dx.doi.org/10.1037/0022-006X.58.5.622

Muris, P., & Merckelbach, H. (1996). A comparison of two spider fear questionnaires. *Journal of Behavior Therapy and Experimental Psychiatry, 27,* 241–244.

*Nordgren, L. B., Hedman, E., Etienne, J., Bodin, J., Kadowaki, A., Eriksson, S., . . . Carlbring, P. (2014). Effectiveness and cost-effectiveness of individually tailored Internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: A randomized controlled trial. *Behaviour Research and Therapy, 59,* 1–11. http://dx.doi.org/10.1016/j.brat.2014.05.007

Oxman, A. D., Clarke, M. J., & Stewart, L. A. (1995). From science to practice. Meta-analyses using individual patient data are needed. *Journal of the American Medical Association, 274,* 845–846. http://dx.doi.org/10.1001/jama.1995.03530100085040

Pallesen, S., Nordhus, I. H., Carlstedt, B., Thayer, J. F., & Johnsen, T. B. (2006). A Norwegian adaption of the Penn State Worry Questionnaire: Factor structure, reliability, validity and norms. *Scandinavian Journal of Psychology, 47,* 281–291.

Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: A meta-analysis of outcomes from studies using waitlist control groups. *Journal of Affective Disorders, 66,* 139–146. http://dx.doi.org/10.1016/S0165-0327(00)00304-9

Powers, E., & Witmer, H. L. (1951). *An experiment in the prevention of delinquency: The Cambridge Somerville youth study*. New York, NY: Columbia University Press.

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *British Medical Journal (Clinical Research Ed.), 340,* c221. http://dx.doi.org/10.1136/bmj.c221

Rosen, R. C., Riley, A., Wagner, G., Osterloh, I. H., Kirkpatrick, J., & Mishra, A. (1997). The international index of erectile function (IIEF): A multidimensional scale for assessment of erectile dysfunction. *Urology, 49,* 822–830.

Rozental, A., Andersson, G., Boettcher, J., Ebert, D. D., Cuijpers, P., Knaevelsrud, C., . . . Carlbring, P. (2014). Consensus statement on defining and measuring negative effects of Internet interventions. *Internet Interventions, 1,* 12–19. http://dx.doi.org/10.1016/j.invent.2014.02.001

Rozental, A., Boettcher, J., Andersson, G., Schmidt, B., & Carlbring, P. (2015). Negative effects of internet interventions: A qualitative content analysis of patients' experiences with treatments delivered online. *Cognitive Behaviour Therapy, 44,* 223–236. http://dx.doi.org/10.1080/16506073.2015.1008033

Rozental, A., Kottorp, A., Boettcher, J., Andersson, G., & Carlbring, P. (2016). Negative effects of psychological treatments: An exploratory factor analysis of the negative effects questionnaire for monitoring and reporting adverse and unwanted events. *PLoS ONE, 11,* e0157503. http://dx.doi.org/10.1371/journal.pone.0157503

Saemundsson, B. R., Borsdottir, F., Kristjansdottir, H., Olason, D. B., Smari, J., & Sigurdsson, J. F. (2011). Psychometric properties of the Icelandic version of the Beck Anxiety Inventory in a clinical and a student population. *European Journal of Psychological Assessment, 27,* 133–141.

Scott, K. M., Wells, J. E., Angermeyer, M., Brugha, T. S., Bromet, E., Demyttenaere, K., . . . Kessler, R. C. (2010). Gender and the relationship between marital status and first onset of mood, anxiety and substance use disorders. *Psychological Medicine, 40,* 1495–1505. http://dx.doi.org/10.1017/S0033291709991942

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy, 47,* 637–645. http://dx.doi.org/10.1037/a0021175

Sharpe, J. P., & Gilbert, D. G. (1998). Effects of repeated administration of the Beck Depression Inventory and other measures of negative mood states. *Personality and Individual Differences, 24,* 457–463. http://dx.doi.org/10.1016/S0191-8869(97)00193-1

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for *DSM–IV* and ICD-10. *The Journal of Clinical Psychiatry, 59,* 22–33.

Simmonds, M. C., Higgins, J. P. T., Stewart, L. A., Tierney, J. F., Clarke, M. J., & Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. *Clinical Trials, 2,* 209–217. http://dx.doi.org/10.1191/1740774505cn087oa

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402–408. http://dx.doi.org/10.1037/0022-006X.60.3.402

Spitzer, R. L., Kroenke, K., Williams, J. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092–1097.

Stawski, R. S., Sliwinski, M. J., Almeida, D. M., & Smyth, J. M. (2008). Reported exposure and emotional reactivity to daily stressors: The roles

of adult age and global perceived stress. *Psychology and Aging, 23,* 52–61. http://dx.doi.org/10.1037/0882-7974.23.1.52

Stewart, G., Altman, D. G., Askie, L. M., Duley, L., Simmonds, M. C., & Stewart, L. A. (2012). Statistical analysis of individual participant data meta-analyses: A comparison of methods and recommendations for practice. *PLoS ONE, 7,* e46042

Stewart, L. A., & Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions, 25,* 76–97. http://dx.doi.org/10.1177/0163278702025001006

*Ström, M., Uckelstam, C.-J., Andersson, G., Hassmén, P., Umefjord, G., & Carlbring, P. (2013). Internet-delivered therapist-guided physical activity for mild to moderate depression: a randomized controlled trial. *PeerJ, 1,* e178. http://dx.doi.org/10.7717/peerj.178

Strupp, H. H., & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcomes: With special reference to negative effects in psychotherapy. *American Psychologist, 32,* 187–196. http://dx.doi.org/10.1037/0003-066X.32.3.187

Sundin, E. C., & Horowitz, M. J. (2002). Impact of Event Scale: Psychometric properties. *British Journal of Psychiatry, 180,* 205–209.

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology, 80,* 547–559. http://dx.doi.org/10.1037/a0028226

*Tillfors, M., Carlbring, P., Furmark, T., Lewenhaupt, S., Spak, M., Eriksson, A., . . . Andersson, G. (2008). Treating university students with social phobia and public speaking fears: Internet delivered self-help with or without live group exposure sessions. *Depression and Anxiety, 25,* 708–717. http://dx.doi.org/10.1002/da.20416

Tingey, R., Lambert, M., Burlingame, G., & Hansen, N. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research, 6,* 109–123. http://dx.doi.org/10.1080/10503309612331331638

Titov, N., Andrews, G., Kemp, A., & Robinson, E. (2010). Characteristics of adults with anxiety or depression treated at an Internet clinic: Comparison with a national survey and an outpatient clinic. *PLoS ONE, 5,* e10885.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45,* 1–67.

Vaughan, B., Goldstein, M. H., Alikakos, M., Cohen, L. J., & Serby, M. J. (2014). Frequency of reporting of adverse events in randomized controlled trials of psychotherapy vs. psychopharmacotherapy. *Comprehensive Psychiatry, 55,* 849–855. http://dx.doi.org/10.1016/j.comppsych.2014.01.001

*Vernmark, K., Lenndin, J., Bjärehed, J., Carlsson, M., Karlsson, J., Öberg, J., . . . Andersson, G. (2010). Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. *Behaviour Research and Therapy, 48,* 368–376. http://dx.doi.org/10.1016/j.brat.2010.01.005

Vessey, J. T., & Howard, K. I. (1993). Who seeks psychotherapy? *Psychotherapy: Theory, Research, & Practice, 30,* 546–553. http://dx.doi.org/10.1037/0033-3204.30.4.546

Walker, S., Liberti, L., McAuslane, N., & Levitan, B. S. (2011). Refining the benefit-risk framework for the assessment of medicines: Valuing and weighting benefit and risk parameters. *Clinical Pharmacology and Therapeutics, 89,* 179–182. http://dx.doi.org/10.1038/clpt.2010.290

Wallace, M. L., Frank, E., & Kraemer, H. C. (2013). A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *Journal of the American Medical Association Psychiatry, 70,* 1241–1247. http://dx.doi.org/10.1001/jamapsychiatry.2013.1960

Waller, R., & Gilbody, S. (2009). Barriers to the uptake of computerized cognitive behavioural therapy: A systematic review of the quantitative and qualitative evidence. *Psychological Medicine, 39,* 705–712. http://dx.doi.org/10.1017/S0033291708004224

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50,* 59–68. http://dx.doi.org/10.1037/0022-0167.50.1.59

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82,* 50–59. http://dx.doi.org/10.1207/s15327752jpa8201_10

Young, C. (2006). What happens when people wait for therapy? Assessing the clinical significance of the changes observed over the waiting period for clients referred to a primary care psychology service. *Primary Care Mental Health, 4,* 113–119.

Zuithoff, N. P. A., Vergouwe, Y., Kling, M., Nazareth, I., van Wezep, M. J., Moons, K. G. M., & Geerlings, M. I. (2010). The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: Consequences of current thresholds in a crossectional study. *BMC Family Practice, 11,* 98.

Received February 19, 2016
Revision received August 28, 2016
Accepted August 29, 2016 ■