Summer 8-2017

Client Deterioration in Individual Psychotherapy: A Systematic Review

Emily Lazar

CLIENT DETERIORATION IN INDIVIDUAL PSYCHOTHERAPY:

A SYSTEMATIC REVIEW

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Psychology

Emily Lazar

Indiana University of Pennsylvania

August 2017

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of Psychology

We hereby approve the dissertation of

Emily Lazar

Candidate for the degree of Doctor of Psychology

_____ _____
Derek Hatfield, Ph.D.
Professor of Psychology, Chair

_____ _____
Beverly Goodwin, Ph.D.
Professor of Psychology

_____ _____
Laurie Roehrich, Ph.D.
Professor of Psychology

ACCEPTED

_____ _____
Randy L. Martin, Ph.D.
Dean
School of Graduate Studies and Research

Title: Client Deterioration in Individual Psychotherapy: A Systematic Review

Author: Emily Lazar

Dissertation Chair: Dr. Derek Hatfield

Dissertation Committee Members: Dr. Beverly Goodwin
Dr. Laurie Roehrich

Client deterioration is a term that describes the worsening of a client's condition during treatment. The most recent, comprehensive review on deterioration of adult clients in individual psychotherapy is over 20 years old (Mohr, 1995). Contemporary reviews exist but are narrower in scope or lack systematic methodology. The current study provides a broad, updated synthesis of the empirical literature on client deterioration. Data collection involved a systematic search of outcome studies published between 2011 and 2016 that focused on individual psychotherapy with adults. In addition to an online database keyword search, a manual search of 19 pre-selected clinical journals was conducted. Unpublished data were also sought. Results include the frequency with which researchers reported client deterioration rates in psychotherapy outcome studies, what definitions and measures of deterioration were used, and what proportion of client samples deteriorated. Rates of client deterioration are presented across several categories, including client diagnosis. Factors that may contribute to differential rates of deterioration are also explored. Results indicate that the majority of published outcome studies failed to distinguish clients who deteriorated from clients who showed no change, or failed to report on client deterioration altogether. The current review additionally provides recommendations for future research and practice. The impact of study findings on the conceptual understanding of client deterioration, and how it is detected and managed, is discussed. Results highlight the need for a unified definition of client deterioration, in addition to clearer standards to guide how deterioration is reported in publications.

ACKNOWLEDGEMENTS

I offer sincere thanks to my family, partner, friends, and mentors for supporting me and this project. I also appreciate the collaboration and encouragement of the researchers I corresponded with during data collection. Finally, I am grateful to my therapy clients, without whom I would not fully recognize the importance of this topic.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In psychotherapy outcome research, deterioration is said to occur when a client's condition worsens during the course of treatment. Though psychotherapy is generally effective, rendering positive outcomes for most clients, deterioration takes place in a small but significant minority of cases (Berk & Parker, 2009; Boisvert & Faust, 2006; Hansen, Lambert, & Forman, 2002; Lambert, 2013[a], Lilienfeld, 2007; Linden & Schermuly-Haupt, 2014; Mohr, 1995). Deterioration appears in multiple treatment modalities (e.g. individual therapy, family therapy, group therapy) and across different theoretical approaches to treatment (Lambert, 2013[a]). Participating in psychotherapy is not always the direct cause of negative client outcomes; still, deterioration significantly impacts the treatment process (Lambert, 2013[b]; Mohr, 1995). It also raises legal and ethical concerns related to malpractice and to psychologists' commitment to avoid causing harm to those whom they serve (APA, 2002; Lilienfeld, 2007; Linden, 2012). In addition, deterioration can negatively impact a client's social, occupational, relational, economic, and psychological functioning, as well as willingness to seek and ability to benefit from future treatment (Lilienfeld, 2007; Linden & Schermuly-Haupt, 2014). Finally, client deterioration has negative financial implications for clients and managed care companies (Shimokawa, Lambert, & Smart, 2010). Although the phenomenon of deterioration has been researched for decades, it is difficult to study systematically, and much about it is still unknown.

Although he was not the first to observe it, Bergin (1966) coined the term "the deterioration effect" and was the first to synthesize data on this phenomenon (p. 235; Castonguay, Boswell, Constantino, Goldfried, & Hill, 2010; Lambert, Bergin, & Collins, 1977). He cited experimental research demonstrating that while many subjects receiving psychotherapy

benefited from participation, some fared worse than their control group peers. Bergin differentiated this unexpected negative outcome from a "temporary regression" caused by either the therapeutic process or environmental stressors (p. 237). In addition, he questioned which characteristics of therapists and their interventions may be responsible for deterioration, and called for clinicians to be aware of their potential to cause harm to their clients. Since then, it has become evident that deterioration can occur in many psychotherapeutic contexts. However, those aspects of the therapy process that may present risks for client deterioration (and factors outside treatment that may interact with and exacerbate these risks) warrant further investigation. The current discussion will center on deterioration of adult, nonpsychotic clients in individual psychotherapy.

CHAPTER 2

REVIEW OF RELATED LITERATURE

Theoretical Considerations

Beyond the general notion of poor client outcome (i.e. client leaves therapy worse off than when he or she began), there is no single, agreed upon definition of deterioration (Lambert, 2013[a]; Parry, Crawford, & Duggan, 2016). In fact, Strupp and Hadley (1977) noted that the concept is as difficult to define as mental health itself. This is one of the factors complicating the design, interpretation, synthesis, and communication of deterioration research.

There are several elements that make defining deterioration difficult. First, the existing literature is full of varying terminology. While some authors refer to deterioration, others discuss treatment "casualties," "negative change," "harm," "negative effects," "side effects," "adverse events," and "iatrogenic symptoms" experienced by psychotherapy clients. A few authors have attempted to make distinctions between these potentially overlapping terms, some influenced by the field of psychopharmacology. For example, Boisvert and Faust (2002) posited that the American Psychiatric Association's definition of iatrogenic illness, "a disorder precipitated, aggravated, or induced by the physician's attitude, examination, comments, or treatment," may be applied to explain negative outcomes in psychotherapy (Edgerton & Campbell, 1994, p. 103).

Linden (2012) differentiated between a number of terms in order to bring greater precision to the broad topic of "side effects" experienced by clients. The most general of these terms is "unwanted events" (UEs), non-positive events that a client faces during the course of or closely following his or her involvement in therapy, which can be but are not necessarily caused by the treatment. Linden referred to UEs thought to be caused by the treatment as "treatment-emergent reactions." He then explicated two additional categories of UEs: adverse treatment

reactions (likely caused by appropriately selected and implemented interventions) and malpractice reactions (likely caused by inappropriately selected or delivered interventions). Proposed reasons for poorly selected or implemented treatment include errors in the areas of diagnosis, case conceptualization/theoretical understanding of the client, treatment focus, use of specific techniques, and development of a therapeutic relationship (Linden, 2012). According to Linden, UEs can represent disruptions not only in psychological functioning, but in physical, social, and occupational realms. In his classification, both treatment non-response and deterioration fall under the category of UEs, with deterioration being defined as a treatment-concurrent worsening of the client's symptoms or "illness" (p. 288).

Similarly, Linden and Schermuly-Haupt (2014) made suggestions about how to capture the many different nuances of negative client outcome. These authors continued the use of the term UE and proposed that a ranking system (e.g. unrelated, probably unrelated, possibly related, probably related, or definitely related) be used to indicate clinician or researcher judgment of the likelihood of a causal relationship between UE and treatment. Like Linden (2012), they recommended that a distinction be made between UEs that occur in the face of a poorly selected and/or delivered treatment and those that do not. Furthermore, Linden and Schermuly-Haupt (2014) advocated categorizing the impact of negative events in therapy according to how long they last, their intensity, and the nature the damage the client experiences (e.g. mild/no consequences, moderate/distressing, severe/in need of corrective intervention, very severe/lasting consequences, extremely severe/requiring hospitalization, or life-threatening).

A complicating factor is that what Linden (2012) referred to as "treatment-emergent reactions," Linden and Schermuly-Haupt (2014) labeled "adverse treatment reactions," and what Linden (2012) dubbed "adverse treatment reactions," he and Schermuly-Haupt (2014) later

called “side effects.” The confusion that can be caused by these overlapping terms and competing semantics is evident. The interested reader may consult Roback (2000) for further examples of conflicting applications of terms used to describe a negative impact of therapy on a client.

Building upon the work of Linden (2012), Parry, Crawford, and Duggan (2016) proposed three preliminary definitions to further distinguish between different facets of what they deemed the “range of adverse effects of therapy” (p. 211). First, they referred to “adverse events” as discrete events that may be influenced by or directly caused by the treatment, occur in the midst of treatment or soon after it has concluded, and result in harm to the patient. The authors then stated that “statistically reliable” and “clinically significant” deterioration signify a “sustained” and “worsened mental state” that is often classified by means of a psychometric. They limited these terms to refer to a worsened state that is present at post-treatment and that is caused by the treatment. Finally, the authors acknowledged “patient-experienced harm,” which they defined as residual negative effects of treatment, subjectively experienced by the patient, although these may not have been detected through adverse event monitoring, clinician’s observations, or outcome measures.

Continuing attempts to specify aspects of negative occurrences in therapy may assist in arriving at a clearer consensus of what is meant by deterioration. For the purposes of the current discussion, the term deterioration will be used to describe a statistically or clinically significant worsening of a therapy client’s condition compared to baseline. The terms negative effect and negative outcome will also be used. However, they will describe a more general worsening of a client’s condition than the definition of deterioration outlined above, as well as adverse or unwanted events that occur during therapy.

A point made apparent when considering terminology is that one of the major questions in defining deterioration is whether it must be explicitly a result of the therapy (not a conclusion easily drawn, due to the complex, interacting nature of therapist, client, intervention, and environment) or whether it must merely occur during the course of therapy (Dimidjian & Hollon, 2010; Linden, 2012; Mash & Hunsley, 1993). For instance, the term "iatrogenic" implies that the treatment itself is directly responsible. Some authors, such as Mays and Franks (1985) separated deterioration from negative outcome, noting that the former is necessarily therapy-induced while the latter is a broader term encompassing all negative changes that take place in therapy, whether or not they can be readily attributed to the therapy or the therapist. However, in a number of more contemporary publications (e.g. Lambert 2013[b]; Mash & Hunsley, 1993; Mohr, 1995) the definition of deterioration does not exclude negative changes due to factors outside the immediate treatment environment.

The current discussion is primarily concerned with factors immediately present in the therapeutic encounter (i.e. conceptualization of the client, intervention selection and delivery, clinician and client characteristics, clinician and client behaviors) that may contribute to deterioration. This is because such factors can perhaps be more readily identified and manipulated by clinicians and researchers in the service of decreasing occurrence of deterioration (e.g. screening clients, selecting treatments, creating or employing new intervention strategies). However, external (e.g. methodological, environmental) factors play a role in negative treatment outcome as well, and it is valuable to understand these roles. For instance, additional influences on deterioration could include biases in reporting or outcome instruments, the "natural history and course" of the client's disorder—which may be constant or fluctuating, and client financial

stressors, poor physical health, and lack of social support (Bootzin & Bailey, 2005, p. 873; Dimidjian & Hollon, 2010).

Deterioration is also often addressed or reported under the general category of "treatment failure," presenting the complex problem of differentiating client worsening during the course of therapy from other undesirable outcomes such as early termination of treatment (which may co-occur with deterioration), relapse, or treatment non-response (Lambert, 2011; Lambert, 2013[a]; Lilienfeld, 2007; Linden, 2012; Linden & Schermuly-Haupt, 2014; Mash & Hunsley, 1993; Mohr, 1995; Ravitz, McBride, & Maunder, 2011). An example is found in a randomized controlled trial (RCT) by Schnurr and colleagues (2007) that examined the response of female veterans and active military personnel diagnosed with posttraumatic stress disorder (PTSD) to cognitive-behavioral treatment. The authors reported adverse events that occurred during the course of the trial (e.g. psychiatric hospitalizations, suicide attempts). However, the results of their analyses centered on the majority of participants who achieved clinically significant improvement in scores on a PTSD symptom scale. Readers are able to infer what percentage of participants did not enjoy such positive outcomes, yet lack the information to determine how many of the women deteriorated and how many simply experienced no significant change.

In contrast, some authors do distinguish between deterioration and its close yet separate counterparts. They do so by using available statistical methods, such as Jacobson and Truax's (1991) Reliable Change Index (RCI), which groups client outcomes into four categories: recovery, improvement, no change, and deterioration. For instance, in a RCT examining the effects of cognitive therapy and short-term dynamic therapy on individuals with personality disorders, Svartberg, Stiles, and Seltzer (2004) presented results using the RCI-specified categories. Vøllestad, Sivertsen, and Nielsen (2011) followed a similar approach when

describing the impact of a mindfulness-based stress reduction (MBSR) intervention on symptoms of depression and anxiety. The choice to report results in this precise manner serves readers by providing them with further clarity as they interpret study results and generate ideas for future research. Because the RCI and other means of statistically defining outcome are relevant to the question of operationalizing deterioration, they will be explored in greater detail in the Empirical Considerations section.

As seen, producing a circumscribed description of deterioration is complicated by the fact that its occurrence is not always reported separately from related phenomena like treatment non-response. Harkening back to Bergin (1966), it is also challenging to distinguish consistent decline in a client's condition from temporary discomfort, adversity, or fluctuations that may be expected in psychotherapy (e.g. a period of increased anxiety during exposure treatment) or due to stressors outside of, yet co-occurring with, the therapy (Dimidjian & Hollon, 2010; Lilienfeld, 2007; Mash & Hunsley, 1993). Similarly, distinguishing a stable negative outcome of treatment from minor negative influences or treatment risks is an important, albeit complex, separation to make (see Linden, 2012 and Mohr, 1995 for further discussion).

Difficulty in defining deterioration is not only limited to a wide range of terminology or attempts to specify a causal source. There are also many descriptions of the forms deterioration can take. These descriptions are often influenced by methods of measurement (which will be addressed in the Empirical Considerations section) and theoretical allegiance (Mash & Hunsley, 1993). The domains of mental health, and likewise the assessment and treatment targets of psychotherapy, are extremely broad (Linden & Schermuly-Haupt, 2014; see Ogles, 2013 for discussion of various attempts to organize possible domains of outcome assessment through conceptual models). As such, deterioration might appear as increased maladaptive cognitions;

risky behaviors; worsening of mood; deficits in coping, reality testing, and performing activities of daily living; and disturbance in physical health, relationships, work, cognitive abilities, emotion regulation, and identity (Mohr, 1995). Domains in which undesirable events in therapy, including deterioration, may manifest also involve worsening of current symptoms, new symptom development or symptom substitution, client non-compliance with treatment, course of treatment lasting longer than expected, client dependency on the therapist, need for a higher level of care, conflict in the client-therapist relationship, risk of harm to self or others, and disruption of social role functioning (Dimidjian & Hollon, 2010; Lilienfeld, 2007; Linden and Schermuly-Haupt, 2014; Mohr, 1995).

One of the most broad and popular ways of characterizing deterioration is as a general worsening of symptoms. Beyond this, however, separate theoretical schools of psychotherapy differ in the treatment goals they emphasize and thus, can differ in the way they conceptualize a successful versus a poor outcome (Mash & Hunsley, 1993; Ogles, 2013). Moreover, diverse theoretical approaches may be associated with differing forms of deterioration (Mohr, 1995). A 2011 special issue of the *Journal of Clinical Psychology* illustrated this by offering a series of articles on treatment failure, of which deterioration is one facet, written from a variety of theoretical viewpoints. One group of authors, Ravitz et al. (2011) discussed the treatment of depression using interpersonal therapy (IPT) and defined deterioration as a specific aspect of treatment failure characterized by increased symptom severity. Through case examples of clients who worsened during IPT, they described not only increased depressive symptoms, but also increased substance abuse and family and financial problems.

Another group—Hopko, Magidson, and Lejuez (2011)—addressed the treatment of depression through a behavioral lens. They examined poor outcomes among clients participating

in behavioral activation (BA). These authors stated that deterioration in depressed clients can be evidenced through "a range of content areas, including affective, verbal-cognitive, somatic, behavioral, and social symptoms of depression" (p. 1110). Attention was also given to increased symptoms of comorbid disorders such as anxiety and substance use, as well as cases in which client ability to perform behavioral targets specific to BA declined (e.g. behavioral avoidance, problems in social interactions or social isolation, reduced ability to take pleasure in or be reinforced by one's environment).

Watson (2011) reported on defining client outcome in the realm of humanistic and experiential treatments. Like Hopko et al (2011), she highlighted the idea that the definition of deterioration in the practice of these therapies is closely related to the theory-driven goals of the treatment. In humanistic and experiential therapies, such goals include experiencing safety and positive regard via a relationship with the therapist, which is proposed to facilitate self-acceptance, exploration and integration of various parts of the self, personal agency, improved self-care, and valuing and trusting one's own perceptions and experiences. Therefore, a client whose relationship with the therapist begins to suffer, who becomes less willing to self-disclose or is increasingly silent during sessions, more needy of support, more frequently engaged in self-defeating or self-neglectful behaviors (e.g. self-harms, restricts food intake), and experiences greater shame, loneliness, hopelessness, and self-dislike could be said to be deteriorating. In agreement with the authors writing from the interpersonal and behavioral perspectives, Watson also remarked that general worsening of symptoms is a key feature of deterioration.

Representing a psychodynamic perspective, Gold and Stricker (2011) explained that symptom reduction and increased success in interpersonal and occupational spheres are typical benchmarks of treatment success in the eyes of the client. However, the authors posited that for a

psychodynamic practitioner, client self-understanding is the most valuable outcome, and that insight into one's symptoms is a mechanism that can lead to symptom relief. Correspondingly, these authors approached deterioration as a worsening of the overt presenting problem (e.g. depression, anxiety) driven by more covert processes such as the exacerbation of negative "self and object representations" (e.g. diminished sense of self-worth or autonomy, increased expectation of negative treatment by others, p. 1099). Relatedly, it is notable that conceptualizations of client personality often enter into a psychodynamic understanding of the mechanisms of deterioration. This is arguably less typical in formulations from other theoretical perspectives, such as behaviorism (Mohr, 1995).

Speculation about the *causes* of client deterioration may be guided by theoretical allegiance in the same way that attempts to describe the *nature* of deterioration can be. It is also possible that techniques associated with specific treatment theories or programs present greater risk of deterioration than others (Lilienfeld, 2007). These topics will be reported upon further in the Contemporary Reviews section.

When considering markers of deterioration, diversity exists not only in the judgments of clinicians; clients, significant others, and outcome measurement tools offer their own varying reports. It follows that another unresolved question related to deterioration concerns which stake-holding parties have a say in its definition (Mohr, 1995; Dimidjian & Hollon, 2010). Is deterioration a subjective phenomenon that one must rely on the client to report, either verbally or via an outcome measure (Duggan, Parry, McMurran, Davidson, & Dennis, 2014)? Can it be detected through other channels, such as actuarial models, personality testing, behavioral observations, or reports by relevant parties in the client's life? What is to be done when multiple sources of information present conflicting data or understand deterioration in deviating ways?

For instance, a therapist may note improvement in a client who becomes more assertive toward family members. However, the family may consider the new behaviors disruptive or contrary to their values, and the client may subsequently report decreased quality of family relationships. Strupp and Hadley (1977) proposed that mental health professionals, clients, and society (including people significant to the client) should all contribute to a comprehensive definition of what characterizes psychological deterioration versus psychological health. The authors stated that society and significant others offer a focus on maintaining stable behavior, interpersonal relationships, and social mores. The goals of the client, they argued, may or may not align with those of society. Typical client goals included experience of positive emotions, feeling one's individual needs are met, and ability to function in areas of personal importance (e.g. job, leisure, romantic pursuits). Finally, mental health professionals may be most interested in the client's personality structure, symptoms, diagnoses, adaptive functioning, coping/stress tolerance, and reality testing. Thus, worsening in any of the aforementioned areas as compared to how the client fared upon entering therapy has the potential to be construed as deterioration. The use of clinical judgment, outcome measures, and actuarial models to detect deterioration will be discussed further in the Empirical Considerations section.

As seen, the meaning of deterioration is not limited to increased frequency and severity of the symptoms that brought the client to treatment. The deterioration construct also may not fall on a single dimension; a client may improve in one area while worsening in another (Boisvert & Faust, 2002; Dimidjian & Hollon, 2010; Lilienfeld, 2007; Mash & Hunsley, 1993; Strupp & Hadley, 1977). Moreover, a client might show improvement in the symptoms he or she originally presented, but then deteriorate due to an emerging set of new symptoms (Linden & Schermuly-Haupt, 2014; Mohr, 1995).

Furthermore, it is apparent that determining the meaning of deterioration is not simply a matter of identifying specific occurrences, such as symptoms, life stressors, or ruptures in the therapeutic alliance. It is complicated by stakeholder, measurement, and theoretical allegiance effects, as well as by individual client context. Decline in a client's condition cannot be detected without making comparisons to a subjective standard of mental health, or to the client's individual baseline functioning (Strupp & Hadley, 1977). For instance, a client may begin treatment already on a deteriorating course, but the therapist may not immediately detect it (Dimidjian & Hollon, 2010; Lambert, 2013[b]; Linden, 2012; Mohr, 1995). Such a case begs the question of whether psychotherapy always contributes to deterioration, or sometimes merely fails to stand in its way (Mohr, 1995).

Moreover, it is possible for two clients to exhibit the same symptom profile, yet one could be deemed deteriorating and another improving (Dimidjian & Hollon, 2010). For instance, a person who begins to self-harm once a week during the course of therapy when he or she did not do so previously may be exhibiting deterioration. Simultaneously, another person who engages in the same behavior following a history of self-harming daily is showing improvement. Another example is that a client with anorexia nervosa who consistently gains weight during treatment is demonstrating desired progress, while a client who consistently gains weight secondary to increased depression or binge eating may be judged as worsening.

Empirical Considerations

Without a specific, unifying definition of client deterioration, researchers and theorists are left without clear ways of operationalizing this construct for scientific investigation. According to Linden and Schermuly-Haupt (2014), there is no universal method of detecting unwanted outcomes in clinical trials and other forms of research examining the effects of psychotherapy. Thus, measures of deterioration vary across treatment outcome studies,

complicating the interpretation of cross-study comparisons and creating confusion for clinicians who wish to select measures to use in their various practices (Ogles, 2013).

Study Design

Additional complexities are involved in selecting an overall design for a study in which deterioration will be examined. Studying client worsening in both an ethical and experimental way is not a straightforward task (Lambert, 2013[a]; Mash & Hunsley, 1993; Mohr, 1995). Lilienfeld (2007) expounded some of the difficulties. For instance, if a researcher tests a particular treatment to determine its potential to cause harm, this introduces concern about the implications of randomly assigning participants to the treatment condition. Furthermore, there is the question of when to discontinue an experimental treatment versus continuing to collect data after evidence of negative outcome appears, as well as what is to be done for the participants who have thus been harmed (Mohr, 1995).

If a treatment effect is found in a single study, it is typically recommended that these results be interpreted conservatively until independent parties have replicated the findings (Lilienfeld, 2007). Mohr (1995) contended that this recommendation rings especially true in cases of low-frequency treatment phenomena (which deterioration appears to be). Deterioration's low base rate can make interpreting research on it, and predicting when it will occur, a considerable feat. Yet this may offer some advantages for demonstrating the likelihood that a treatment has contributed to client deterioration. According to Lilienfeld (2007), "when low-base rate negative events…consistently appear shortly after the introduction of treatment, such evidence should be accorded considerable weight even when it does not derive from RCTs" (p.57). Without sound research designs and replication of deterioration studies, consumers of research are left with ambiguous results and limited information on which to base health care

decisions (Lilienfeld, 2007). However, if a treatment program, or at least one of its components, is associated with client deterioration in a single study, the ethicality of conducting replications is in question (Lilienfeld, 2007). Thus, researchers face an obvious dilemma.

Several authors have explored ways to circumvent some of the methodological and ethical problems presented by research on negative client outcomes. Dimidjian and Hollon (2010) urged further consideration of the replication problem, suggesting that such discussion not only involve the ideas of clinicians and researchers, but those clients and other stakeholders as well as experts on ethics, statistics, public policy, and law. Mohr (1995) called for researchers to consider how they may deviate from predetermined study procedures to intervene on behalf of deteriorating clients, and presented suggestions such as offering these clients additional treatments, extended sessions, or follow-up. Another potential solution is found in the form of adaptive treatment designs, in which a research participant's treatment course is tailored in response to his or her clinical progress during the study (Kendall, Comer, and Chow, 2013). Specifically, clients who experience decline in one study condition may be switched to another (e.g. from novel/experimental treatment to treatment as usual, or vice versa). Such methods may help reveal moderators and mediators of treatment response that can be incorporated into the tailoring strategies of future studies that employ adaptive treatment designs (Laurenceau, Hayes, & Feldman, 2007).

Mohr (1995) also suggested that it is more practical, given current methodological and ethical constraints, to focus on establishing markers of poor client prognosis (i.e. combinations of events that contribute to deterioration) instead of attempting to establish causality between deterioration and single, specified "agents" (p. 23). Relatedly, he warned against the study of deterioration being reduced to a mere "spitting war" between advocates of differing therapeutic

modalities (p. 20). In a similar vein, Dimidjian and Hollon (2010) pointed out that characteristics of the client, the therapist, and the quality of treatment delivery or technique must also be considered as moderators of treatment effects. Additionally, these authors noted that even explicating what is meant by "treatment" or "psychotherapy" is a complicated matter. This derives from the fact that clinicians vary in their level of skill and the content of the services they provide. Barlow (2010), however, remarked that there remains "heuristic value" in attempting to classify potentially harmful treatments (PHTs), despite the inability of this approach to address all questions about the causes of deterioration (p. 18). Such a task has already been taken up by Lilienfeld (2007). Castonguay and colleagues (2010) voiced support for this endeavor, adding that identifying and comparing PHTs may help to pinpoint common mechanisms they share. In his paper, Lilienfeld (2007) additionally proposed that reviewers and consumers of research on PHTs consider applying a ranking system to study designs in order to differentiate more convincing sources of evidence from less convincing sources. Representing the former end of the scale were RCTs. Representing the latter were naturalistic and case study designs. Quasi-experiments with matched comparison groups fell in the middle.

Dimidjian and Hollon (2010) took an even more nuanced approach to design selection, explaining that the identification of negative outcomes in psychotherapy may begin with anecdotal evidence, individual case studies, and qualitative methods before proceeding to experimental, nomothetic, and quantitative methods. An example of qualitative work is found in Bystedt, Rozental, Andersson, Boettcher, and Carlbring, (2014). These authors employed thematic analysis to investigate therapist responses to open-ended questions about negative effects in their own clients. Three themes emerged in the self-reports. These included descriptions of the negative effects (e.g. increased dependency, emergence of new symptoms,

negative impact on multiple areas of life), speculations on its causes (e.g. PHTs, poor therapeutic alliance, unprofessional therapist behavior, client's external stressors), and criteria for assessing it (participants struggled to identify *criteria* as opposed to *methods* such as clinical judgment or outcome measures). Though no causal conclusions can be drawn from this study, it sheds light on the practices and thought processes of clinicians and presents a number of questions for future study.

Survey data can prove similarly useful despite their limitations (e.g. causal conclusions cannot be drawn). For instance, Crawford and colleagues (2016) conducted a large-scale survey of individuals receiving psychotherapeutic treatment in Wales and England, inquiring about any negative effects they might have experienced during their treatment. Five percent of the over 14,000 respondents endorsed "lasting bad effects" that they attributed to their treatment experiences. The authors' analysis of these data revealed that self-report of lasting bad effects was less likely to occur among respondents over the age of 65 as well as among respondents who agreed that they had received sufficient "information about treatment before it began." Self-report of lasting bad effects was more likely to occur among respondents who identified as members of sexual or ethnic minority groups as well as among respondents who indicated that they were not able to identify what type of therapy they had received (e.g. theoretical approach, modality). By identifying these potential protective and risk factors, the results of this study suggest variables for future research on negative effects to target.

Also relevant to the discussion of study design, Barlow (2010) has suggested the use of single-case experimental designs, which can assist in detecting variability in responding within and between recipients of particular interventions. In turn, RCTs may be used to support the notion that a treatment (as opposed to external factors alone) has caused a negative outcome, and

then to explain *why* by uncovering mechanisms of harm through mediation analysis. Component analyses and dismantling studies can further assist in isolating "active ingredients" contributing to deterioration—a necessary step in determining how treatments may be modified to reduce risk (Barlow, 2010; Dimidjian & Hollon, 2010). Furthermore, moderator analyses may clarify whether deterioration occurs uniformly for all clients receiving a certain treatment, or only for those with particular characteristics. Moderation can also aid the detection of possible harmful interactions between multiple within-treatment factors (Bootzin & Bailey, 2005; Dimidjian & Hollon, 2010).

Lambert (2013[a]) is another proponent of identifying specific treatments or treatment processes that play a role in deterioration. He noted that a baseline rate for the occurrence of deterioration in untreated individuals is unestablished, making it difficult to differentiate whether therapy itself is a key influence in negative client change, or whether this is primarily the result of external circumstances. Like Dimidjian and Hollon (2010), he recommended use of RCTs in order to elucidate whether treatment factors are responsible for decline in client functioning or whether this same result would have occurred without treatment. Lambert observed that in RCTs, rates of negative outcome are often lower among controls—results that encourage further inquiry into the specific role of the therapy process in client worsening.

According to Mash & Hunsley (1993), researchers sometimes turn to retrospective designs in order bypass ethical and methodological restrictions inherent in researching factors associated with deterioration. There are some notable limitations associated with studies that examine deterioration in this way, whether they rely on client report, clinician report, or analysis of archival data (e.g. progress notes). For instance, clients asked to report the results of their treatment following the conclusion of therapy are at risk of conflating deterioration during

treatment with relapse that occurred after treatment had ended (Mohr, 1995). Clinicians asked to recall client outcome and/or speculate on what contributed to a client's deterioration may be swayed by self-serving or hindsight biases. Relatedly, self-report is always influenced by the current state of the responder, and the accuracy and detail of one's memory of treatment decreases the further removed in time one is from when those experiences took place (Mohr, 1995).

Overall, there are many legitimate uses for retrospective designs in studying psychotherapy outcome. For instance, they are beneficial in the beginning stages of inquiry into novel topics, and sometimes they are the only available method. Nevertheless, researchers and consumers of research should be aware of the potential problems posed by use of retrospective studies in the context of trying to identify the mechanisms that produce deterioration.

Detecting Deterioration

After a research team has chosen a design for studying client outcomes, they must also choose how those outcomes will be measured. There are several barriers to successfully detecting deterioration in psychotherapy beyond the lack of a standard operational definition. Clinical judgment based on practitioner observation of the client is one proposed means. However, data suggest that this method alone, even when supplemented with base rate information, is not sufficient to detect client worsening (Hannan et al., 2005; Hatfield et al., 2010). Perhaps this finding can be attributed to the complex and multifaceted nature of deterioration or to the fear that evidence of deterioration in one's clients signifies that one is an ineffectual therapist (Mash & Hunsley, 1993). It may also be related to clinicians simply being unaware that a proportion of clients can and do worsen during treatment; they do not actively monitor for deterioration because they do not expect it to occur (Boisvert & Faust, 2006). Mohr (1995) also emphasized that clinicians tend to underestimate the extent of their clients' problems.

Clients can likewise minimize symptoms in self-report (Mohr, 1995). In addition, clients may experience deterioration yet not report it due to a respect for the therapist's authority, desire to please or avoid angering the therapist, or desire to feel as though their investment of time and resources in therapy has not been wasted (Gold & Stricker, 2011; Mash & Hunsley, 1993). Standardized outcome measures act as supplements to the judgments a clinician can form based on observations or informal client self-report. Here again, the imposing questions of how to operationalize client wellness versus deterioration, what items to include in a measure, what format to use, and whom to employ as raters or informants emerge—all elements that affect how the client's state is interpreted at any given time, as well as the calculation of effect sizes (Ogles, 2013; Ronk, Hooke, & Page, 2012). Once such questions are answered and a measure is selected or constructed, the next task is determining its utility and psychometric properties (Mash & Hunsley, 1993). Unfortunately, researchers' interest is more often captured by the opportunity to formulate new interventions than by the opportunity to design and test ways of assessing the success or failure of those interventions (Ogles, 2013).

Still, negative outcome can be identified by many of the same means used to measure positive outcome (Dimidjian & Hollon, 2010; Mash & Hunsley, 1993). Deterioration is a less frequent and less expected result. Yet outcome measures capable of capturing deterioration do exist, as do recommendations for their further development. For example, Mash and Hunsley (1993) proposed integration between research on factors that contribute to deterioration during treatment and the creation of means to assess those factors. These authors predicted that such collaboration can lead to an increasingly refined understanding of what deterioration is, what causes it, and how to monitor it. They then presented a framework to guide the pursuit of this ideal. It advocated for clear delineation of long- and short-term client goals at the outset of

treatment, specified markers to indicate progress or regression in terms of the client goals, and continual monitoring of the markers using theoretically-based instruments. The authors additionally specified that such instruments should possess good psychometric properties, utility and feasibility in applied clinical settings, and a structure sensitive enough to detect change throughout the therapy process. Similar ideas have since been applied to measuring psychotherapy outcome in general (see e.g. Ogles, 2013).

The previous section, Theoretical Considerations, highlighted several factors that may impact how deterioration is defined, including the theoretical orientation of the researcher or clinician, potential for clients to decline in areas other than those demarcated by their initial symptoms, and differing conceptualizations of deterioration by diverse stakeholders. Discussion returns to these concerns in light of their relevance to deterioration assessment. Mohr (1995) cited data favoring the position that outcomes differ significantly depending on who is providing the ratings. One example is a study by Yalom and colleagues (1971) in which a clinician's ability to detect deterioration among clients in group therapy was surpassed by that of fellow group members.

Given the potential for rater conflicts, the range of areas of client functioning in which deterioration might manifest, and the likelihood that there are many predictors of deterioration instead of a clear-cut few, a number of authors have supported the use of multifaceted approaches to outcome assessment (Dimidjian & Hollon, 2010; Lilienfeld, 2007; Mash & Hunsley, 1993). This may mean using a multidimensional measure, multiple instruments, or multiple modalities. Assessment modalities can include behavioral tests, self-reports, interviews, collateral informant data, and even physiological measures (Ogles, 2013). Mash and Hunsley (1993) added that because various theoretical techniques may differentially contribute to

particular kinds of deterioration (e.g. increased anxiety symptoms as a result of interoceptive exposure), professionals selecting assessment measures should be conscious to choose ones that cover domains relevant to the theory from which they practice. This notion has subsequently been supported by authors such as Mohr (1995) and Lilienfeld (2007), though neither went so far as to recommend specific outcome measure/therapy pairings. Ogles (2013), by contrast, observed that the field generally seems to be moving away from theory-based outcome measures and toward diagnostically-based ones.

Relatedly, Mohr (1995) deliberated on the matter of global versus specific measures of outcome. He described global measures as capturing general symptoms or broad areas of functioning while specific measures are targeted to single symptom areas and may be more focused on the client's immediate state. Because deterioration can cast a wide net, affecting multiple areas of client life, the broad-based approach of global measures is advantageous in some contexts. Yet Mohr also argued that such measures may not be sufficiently sensitive to detect more subtle displays of client worsening. Additionally, he expressed concern about the potential to overlook or misinterpret changes in outcome scores due to covariance between particular symptoms or areas of functioning contained within the same measure. For example, a client might experience increased depressive symptoms, but this worsening could be masked by simultaneous positive changes in another domain (e.g. decreased anxiety symptoms). Thus, Mohr recommended the consideration of specific, as opposed to global, outcome measures for use in detecting negative effects. Options for combining aspects of both global and specific approaches might include selecting global measures containing specific subscales, or using multiple specific measures.

The frequency and timing with which outcome assessments are conducted is also significant. Choices surrounding this topic may depend on researcher or clinician resources, expected client reaction to participating in the assessment process, and the particular subjects of study. For example, some treatment programs (e.g. critical incident stress debriefing [CISD]) appear to produce benefits early in their course, but have been associated with negative outcomes later on (Dimidjian & Hollon, 2011). Lilienfeld (2007) suggested that it is also possible for the opposite effect to occur, wherein clients initially deteriorate during treatment but have ultimately improved at termination. Both cases call attention to the usefulness of assessment at regular intervals throughout the course of therapy.

This discussion dovetails with the idea of monitoring clients during the process of therapy as well as evaluating end-state results. Heretofore, this discussion has referred to any kind of assessment of a client's functioning during or after treatment as an "outcome measure." However, a distinction exists between using an outcome measure to track client *progress* continuously throughout therapy and administering a measure only pre- and post-treatment in order to capture *overall* outcome (Ogles, 2013; Overington & Ionita, 2012). A variety of authors (e.g. Hunsley & Mash, 1993; Mohr, 1995; Ogles, 2013) have promoted the former method. Some advantages they presented include the assertion that clients assessed during treatment will self-report more accurate data and that ongoing client monitoring can alert clinicians to deterioration promptly, allowing corrective action to be taken. One of the most well-known examples of an instrument used in this way is the Outcome Questionnaire (OQ), developed by Lambert and colleagues (1996).

The most commonly referenced version of the instrument, the OQ-45, is an empirically-supported, normed, and symptom-based self-report measure for adult psychotherapy clients (OQ

Measures, 2014). It was designed for use in routine practice but is also frequently used in research (Lambert, Gregersen & Burlingame, 2004). This measure is intended to be administered before the commencement of treatment and subsequently on a session-by-session basis until termination. To avoid making continuous assessment a time-consuming process, it is limited to 45 items and typically takes less than 20 minutes to complete (Lambert et al., 2004[a.]). In addition to highlighting endorsement of critical items (e.g. those assessing suicidality, violence toward others, and substance abuse), the OQ-45 tracks clients' overall progress as well as progress on three subscales. These subscales include Symptom Distress (capturing symptoms of mood and anxiety disorders as well as substance abuse), Interpersonal Relationships (capturing loneliness, withdrawal, interpersonal conflict, and marriage and family difficulties), and Social Role (capturing difficulties in the domains of work, academics, domestic responsibilities, and leisure; Lambert et al., 2004[a.]). Such a model corresponds with the previously discussed idea that clinical change may make itself apparent across various domains of client life. However, factor analysis has failed to support the notion that the subscales represent three distinct areas of client functioning (Kim, Beretvas, & Sherry, 2010; Lambert et al., 2004[b.]).

The OQ-45 also employs cut-off scores to allow clinicians to compare client data to that of normative, outpatient, and inpatient samples. A novel feature of the instrument is that it uses both these clinical cut-offs and a RCI developed using Jacobson and Truax's method (1991), described in further detail later in this section. This permits assessment of whether clinically significant change (including deterioration) in client functioning has occurred at any point during therapy (Lambert et al., 2004[a.]). Additionally, scores on the OQ-45 can be used as part of a "feedback system" or "outcome management system" (Lambert, 2013[b], p. 45, 48). One feature of such a system is a "signal-alarm" function that alerts clinicians to clients who are not making the

progress that is expected based on their pre-treatment symptom ratings and rate and direction of session-to-session change (Lambert, 2013[b]; Lambert et al., 2004[a.], p. 215).

Research has demonstrated that therapists' use of feedback reduces rates of deterioration in their clients (e.g. Lambert, 2013[b]; Lambert & Shimokawa, 2011; Shimokawa et al., 2010). Whipple, Lambert, and colleagues (2003) have also pioneered Clinical Support Tools (CSTs) to accompany the OQ-45. They consist of various problem-solving strategies recommended via use of a decision tree that can guide therapists in intervening with clients who are deteriorating or otherwise failing to make desired progress (Whipple et al., 2003). The problem areas targeted by the CSTs correspond with factors previously shown to predict treatment outcome, including the therapeutic alliance, client social support, appropriateness of diagnosis, client motivation for change, and need for psychiatric treatment (Lambert 2013[b]).

Though innovative and highly influential, the OQ-45 is not the only outcome measure that can be used in the detection of negative effects. Various others are noted by authors such as Linden and Schermuly-Haupt (2014) and further described and assessed in terms of their applications, costs, advantages, and disadvantages by authors such as Overington and Ionita (2012). Some of these measures include the Behavior and Symptom Identification Scale-24 (BASIS-24), Behavioural Health Measure-20 (BHM-20), Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM), Polaris Mental Health (Polaris-MH), Partners for Change Outcome Management System (PCOMS), Treatment Outcome Package (TOP), Inventory of Negative Effects, Vanderbilt Negative Indicator Scale, Unwanted Events and Adverse Treatment Reaction Checklist for Psychotherapy (UE-ATR), and Experience of Therapy Questionnaire (ETQ). References for the psychometric properties of a number of the measures can be found in Overington and Ionita (2012).

An additional instrument, the Negative Effects Questionnaire (NEQ) has been introduced by Rozental, Kottorp, Boettcher, Andersson, and Carlbring (2016). Its items were developed with input from researchers, client self-reports, and a review of the existing literature. An exploratory analysis of the measure's factor structure produced a six-factor solution that captured possible negative effects occurring in the context of psychotherapy. These factors included increased symptoms, poor quality of treatment (e.g. client struggles to understand the treatment content), excessive dependence on the therapist or treatment process, stigma (e.g. concern about being judged unfavorably by others due to one's involvement in treatment), lack of hope for recovery, and decreased self-esteem or sense of personal competence. The advent of this questionnaire may prove beneficial for researchers in that it allows for the targeted assessment of negative effects specifically. Additionally, the instrument's six factors raise awareness about what aspects of the client experience and the therapeutic process may be particularly salient in the continued study of negative effects in psychotherapy.

The diversity in clinical assessment tools that exists indicates that deterioration may be a growing area of interest in outcome research. Yet, as stated, an abundance of measures causes confusion as clinicians and researchers struggle to follow the literature on each and try to determine their costs and utility before implementing them in research or practice. In fact, some are in favor of developing a standard battery in order to facilitate replication studies and meta-analyses (Ogles, 2013).

As well as qualitative descriptors, quantitative formulations are needed to operationalize deterioration for empirical study. In fact, they act as a foundation for many of the outcome measures just discussed. There is no universal method for quantitatively defining deterioration, but several options exist, each possessing its own strengths and limitations (Dimidjian & Hollon,

2010). Scogin et al. (1996), for example, described a highly liberal means of classifying negative outcome: at least a 1-point change in a negative direction on any self-report outcome measures or clinician ratings. Such an approach could be useful in terms of minimizing false negatives (i.e. failing to correctly identify deterioraters) yet could also be overly sensitive, resulting in high rates of false positives and failing to discriminate between steady worsening and temporary fluctuations. These points highlight the importance of considering the population, setting, and other factors that make up the context in which deterioration is to be detected.

The concept of clinical significance or clinically significant change is often evoked in the discussion of how to measure client outcomes (Jacobson, Follette, & Revenstorf, 1984). Clinical significance, as opposed to the magnitude of a treatment effect (effect size) or statistical significance signified by results of inferential tests, refers to the practical, applied value of an effect as evidenced by changes in everyday client life (Campbell, 2005; Kazdin, 1999; Ogles, 2013). Jacobson, Roberts, Berns, and McGlinchey (1999) refer to clinically significant change as "returning to normal functioning" (p. 300). Though clinical significance may be measured through subjective reports, there have been various attempts to conceptualize in more objective, quantifiable terms (Ogles, 2013). Examples include the use of cut-off scores to determine whether a client's functioning has entered a normal (as compared to a clinical) range by the conclusion of treatment (Campbell, 2005; Jacobson & Truax, 1991).

One of the most well-known and utilized calculation methods is the Reliable Change Index (RCI; Jacobson & Truax, 1991), which refers to the magnitude of change needed to conclude that pre-post differences in client scores are due to meaningful alterations and not measurement error. The RCI is calculated by dividing the difference between outcome scores collected at two time points (e.g. pre- and post-treatment) by the standard error of the difference

between the two scores. A RCI of ±1.96 (the z-score required to achieve a 95% confidence interval) or more is interpreted as stable, meaningful change. The RCI can be applied to individuals or groups (Campbell, 2005). Its developers also demonstrated its usefulness in classifying treatment outcomes into four categories: recovery (scores within the "normal" range), improvement (scores moving toward the normal range), no change, and deterioration (scores moving away from the normal range; Jacobson & Truax, 1991; Ronk et al., 2012).

There is a concern that deterioration may at times go undetected by the RCI due to regression to the mean, and this is the most frequent criticism lobbied against it (Hiller, Schindler, & Lambert, 2012; Mohr, 1995; Ogles, 2013). Alternative methods for calculating the RCI and clinically significant change in general, some that attempt correct for regression to the mean, are available though they are less frequently used (Ogles, Lunnen, & Bonesteel, 2001). Several studies have compared their utility, generally finding that they perform similarly well (see Atkins, Bedics, McGlinchey, & Beauchaine, 2005 and Ronk et al., 2012). One of these alternate approaches is calculating percentage of improvement (or as applied to this discussion, decline) across time points (Ogles, 2013). This method has been adopted in psychopharmacology research and is recommended by Hiller and colleagues (2012) as a supplement to the RCI.

Notably, the examination of clinically significant change does not negate the value of investigating the magnitude of treatment effects. Negative effect sizes are an additional way to conceptualize deterioration numerically, although negative effect sizes can occur for reasons other than deterioration (e.g. random variation around a small mean effect size; Lilienfeld, 2007). As such, considering both effect size and clinical significance would be a beneficial way to clarify whether deterioration is present.

Reporting Deterioration

Practitioners can be hesitant to consider the possibility of deterioration (Hatfield & Ogles, 2004; Lambert, 2013[b]; Linden & Schermuly-Haupt, 2014). In the same way, researchers may be hesitant to assess and report it (Nutt & Sharpe, 2008). Granted, the state of the literature has expanded and improved notably since the existence of deterioration was first highlighted over 50 years ago. In fact, the majority of early outcome studies did not even include deterioration as a possible classification of client response (Lambert 2013[a]). Yet, as demonstrated in the Theoretical Considerations section, some modern outcome researchers still fail to explicitly report rates of deterioration within their samples. The RCT cited to illustrate this point (Schnurr et al., 2007) was, however, informative regarding possible manifestations of negative effects; the authors reported adverse events encountered by participants during the trial.

Unfortunately, a series of recent studies suggests that such negative effects of psychotherapy trials are likely under-reported in the literature. One study, conducted by Jonsson, Alaie, Parling, and Arnberg (2014) will be detailed in the next section, Contemporary Reviews. Another example is found in the work of Vaughan, Goldstein, Alikakos, Cohen, and Serby (2014). These authors performed a review of 45 RCTs in which they compared the frequency of adverse event reporting in psychopharmacological trials, psychotherapy trials, and combined treatment trials. Results revealed that those administering pharmacological trials were significantly more likely to make reference to negative outcomes, whether this meant reporting an actual occurrence or simply stating that participants were monitored for adverse reactions but none occurred.

A similar study by Meister and colleagues (2016), published two years after Vaughan and colleagues' review, produced comparable results. These authors also systematically reviewed randomized, controlled psychopharmacological trials, psychotherapy trials, and combined

treatment trials to determine how adverse events were reported in each. Samples in these reviewed trials consisted of adult participants diagnosed with persistent depressive disorder (PDD). Meister and colleagues reported that the majority of psychopharmacological trials (39/42) and combined treatment trials (7/9) provided information albeit variable in specificity, about adverse events (e.g. how adverse events were measured, how often participants were assessed). In contrast, only one of the nine psychotherapeutic trials they reviewed included any adverse event information.

Perhaps psychotropic medications inherently pose more risks than psychosocial treatments. However, in these reviews, even discussion of *potential* adverse reactions was more frequent among pharmacological papers. This suggests that attention given to worsening among clients receiving treatment for psychological difficulties is lacking in psychotherapy research compared to psychiatric/medical research, a field in which the discussion of side effects and risk-benefit analyses is more established (Vaughan et al., 2014). Vaughan and colleagues acknowledged that a major impediment to psychotherapists who wish to engage clients in a discussion of risks and benefits of treatment is the current lack of conclusive predictive knowledge in this area.

An additional similar study with similar results was conducted by Duggan, Parry, McMurran, Davidson, and Dennis (2014). These authors reviewed the protocols and final reports of all 82 trials funded by the National Institute for Health Research (NIHR) from 1995 to 2013, looking for reports of harm to participants. They found that drug treatment protocols were more likely to discuss anticipated adverse events than psychological treatment protocols. The protocols' discussions of potential adverse events were based primarily upon severe adverse event guidelines designed by the National Research Ethics Service (NRES) for use with

pharmacological interventions. Duggan and colleagues questioned whether such guidelines were transferrable to psychotherapeutic research and whether they were sufficient to detect poor outcomes in that realm. The authors encouraged the formulation of guidelines more specific and applicable to psychosocial treatments.

Vaughan et al. (2014) observed that negative effects that had actually transpired were reported in some of the psychotherapy trials they reviewed. Conversely, in Duggan et al.'s (2014) review, none of the final reports on psychological interventions even made mention of them. These authors noted three possible explanations: that no negative effects occurred, that such effects occurred but were not detected, or that such effects were observed but not included in the final report. Duggan and colleagues concluded that researchers, particularly those who are using clinical trials to evaluate a novel treatment, should give more precise attention to the monitoring and reporting of negative outcome. In addition, they called for renewed attention to Ioaniddis and colleagues' (2004) critique of "common poor reporting practices for harms-related data" in RCTs. Some examples include reporting only measures of central tendency instead of also remarking on extreme scores, and presenting general or summed data on adverse events instead of specifying the type and severity of these events.

In Vaughan and colleagues' study, an adverse event was defined as "a deleterious result attributed directly to a treatment intervention" and included such problems as worsening of mood symptoms and physical safety concerns (2014, p. 850). Duggan and colleagues' study, which examined the broader concept of "harm" deemed it a "sustained deterioration" ascribed to the choice of intervention or its inadequate application, although they allowed for the interaction of client factors as well (2014, p. 2). These descriptions of harm and adverse events are distinct from the definition of deterioration offered by a number of other authors (e.g. Lambert 2013[b];

Mash & Hunsley, 1993; Mohr, 1995) who allowed that deterioration can be due to factors other than the direct influence of treatment. In spite of the discrepancies between deterioration, adverse events, and harm, these different forms of negative outcome may share similar roots and consequences.

Under-reporting of negative results in psychotherapy is not simply a product of researchers' unwillingness to acknowledge or investigate decline in client wellbeing during data collection. As discussed throughout this paper, there are many challenges associated with measuring and detecting negative treatment phenomena. Vaughan and colleagues (2014) proposed that the lack of consensus regarding how to define negative outcome is a major culprit. Additionally, negative effects could be de-emphasized because they tend to occur infrequently; the field may value and focus primarily on the positive results experienced by the majority as compared to idiographic data provided by the unusual few who suffered undesirable reactions (Barlow, 2010; Mohr, 1995; Vaughan et al., 2014).

Furthermore, a reciprocally influential relationship between researchers and the standards of the journals in which they wish to publish may be implicated. Mohr (1995) speculated that journal editors can be wary of publishing studies in which negative outcomes are reported because such untoward events may require investigators to break from standard protocol, thus introducing confounds into the data and weakening the methodology. Mohr criticized this potential bias against research teams who choose to stray from study protocol in the service of reducing harm to clients, and then subsequently report these deviations in their manuscripts. He expressed that such a stigma prevents the dissemination of valuable information not only about what factors may contribute to negative effects, but also about what responses from researchers are most helpful when poor outcomes befall their participants.

Correspondingly, researchers may choose to under-report harm to participants or forego reporting it altogether out of embarrassment, shame, concern for their reputation, or fear that their submitted work will be rejected (Mohr, 1995). The under-publication of studies returning null results is commonly referred to as the "file-drawer problem" (Rosenthal, 1979). Within the realm of negative treatment outcome, a different kind of file-drawer problem may exist wherein studies demonstrating significant effects in the *opposite* direction than that which is expected (sometimes referred to as "Type III error") may be turned down for publication and filed away, too (Leventhal & Huynh, 1996, p. 278) . Thus, the responsibility lies not only with journal editors. If authors expect their papers to be rejected, they may not even seek to publish in the first place (Lilienfeld, 2007). The prevalence of this specialized type of file-drawer problem, and the extent of its impact, is largely unclear and warrants further investigation (Lilienfeld, 2007).

Contemporary Reviews

Given the complexity of client deterioration and its numerous subtopics, critical literature reviews can play a valuable role in clarifying extant research on the subject. One of the most recent comprehensive attempts to synthesize this literature is found in a review by Mohr (1995). His investigation centered on 42 studies that were published between 1953 and 1993 and made mention of negative outcome for psychotherapy clients. Mohr treated the term "negative outcome" as including reliable deterioration, as well as related but separate constructs, such as any increase in client's symptoms, treatment non-response, and qualitative judgments documented by clinicians in the client chart or self-reported by the client. The designs of the studies that Mohr reviewed ranged from RCTs to single-subject case studies. A detailed explanation of the methods used to select the papers was not provided. Mohr, however, did note the instruments that authors employed to measure the criterion of negative outcome, information about the therapists who were providing treatment to the study participants, the nature of the

treatment (e.g. group, individual, cognitive-behavioral, insight-based, interpersonal), the frequency with which negative outcomes occurred in each study's sample, and further specified results. Additionally, he recorded the diagnoses represented within each study (e.g. anorexia nervosa, anxiety disorders, major depressive disorder, conversion disorder) as well as some demographic information (e.g. inpatients, outpatients, college students) for each sample. He limited his examination to studies of adult populations not initially presenting with psychotic symptoms.

Across the studies, a number of methods were utilized to detect negative outcomes. These included clinical judgment/therapist ratings, client self-reports, ratings by a third party (e.g. therapist's supervisor), structured interviews, and other instruments such as the Minnesota Multiphasic Personality Inventory (MMPI), Beck Depression Inventory (BDI), Symptom Check List-90 (SCL-90), Thematic Apperception Test (TAT), Q-Sort, and Hamilton Depression Rating Scale (HDRS or HAM-D). The type of negative outcomes observed included new or returned symptoms (e.g. anxious, depressive, or psychotic symptoms; binge eating; weight loss or gain; suicidality; increased muscle tension; increased substance use), interpersonal and occupational problems, and changes in personality and ego functioning. The percentage of participants who experienced worsening with treatment varied between and within studies, ranging from 0-80% (single-case studies and studies failing to report deterioration separately from no change excluded). The vast majority of studies, however, reported rates between one and 20 percent.

In addition to determining the rates at which negative outcome occurs in psychotherapy, Mohr reported on methodological difficulties involved in studying the topic. Several of these have been cited in previous sections of the current discussion. For instance, Mohr observed that in some studies slight variation occurred depending on whether clients or therapists were

providing ratings, although across studies one party did not consistently report negative outcome more frequently. Furthermore, Mohr sought to explore "patient, therapist, and therapy variables," including therapy process variables, suggested by his cross-study comparisons to be associated with negative outcome (1995, p. 1). He chose not to include variables external to therapy that could affect client wellbeing, noting that to do otherwise would create too large a scope for the article. Also excluded was discussion of the most extreme form of negative outcome: suicide. Mohr justified this by stating that a large body of literature on predicting and preventing client suicide already existed at the time.

Though he noted that client improvement in psychotherapy is generally the norm, Mohr's review identified several potential risk factors for deterioration and other negative outcomes. Included among the factors stemming from the client were diagnoses of borderline personality disorder or obsessive-compulsive disorder, "severe interpersonal difficulties" (e.g. marked distrust or hostility toward the therapist, becoming overly dependent on the therapist, limited interpersonal skills), low motivation or expectations for treatment, initial symptom severity, and the expectation that treatment will not entail a degree of difficulty and discomfort (p. 1). An additional factor suggested in Mohr's findings is a tendency for some clients to genuinely or seemingly deteriorate as a means of interfering with or attempting to remain in treatment. Among the therapist- and therapeutic relationship-based factors were failure to provide warmth and empathy (although among clients who are highly suspicious, provision of warmth and empathy by the therapist may lead to negative outcome), failure to correctly judge the severity of the client's problems, errors in delivery or technique (e.g. providing too little structure in treatment), negative countertransference reactions (e.g. anger, disappointment), unethical behavior (e.g. sexual relationship between client and therapist), and disagreement with the client

about aspects of the therapeutic process (e.g. goals of therapy, degree of severity attributed to client's concerns).

When considering factors originating from treatment selection, Mohr focused on expressive-experiential therapies (e.g. emotion-focused therapy, gestalt therapy) and therapies that are largely client-directed (e.g. bibliotherapy), which he argued show broadly higher deterioration rates than do cognitive, behavioral, and interpersonal therapies. However, this statement was qualified by findings that sudden improvements can also occur within the context of gestalt/expressive-experiential treatments, and that client-directed therapies are most typically associated with deterioration when applied to clients in severe distress who likely require a higher level of care. Mohr interpreted this to mean that negative outcomes linked to such treatments likely occur because of interactions between the treatment and one or more other variables (e.g. symptom severity, errors in technique or treatment delivery). Moreover, Mohr conceded that negative outcome can take place regardless of a clinician's theoretical approach. Some examples cited are increased anxiety following progressive relaxation training (a behavioral technique) and negative outcome in psychodynamic therapy for clients with severe psychopathology. He cautioned against his work being used as evidence of the global superiority of one treatment modality and stated that practitioners should instead consider which modality is best suited for a particular type of client or presenting problem.

Another conclusion reached by Mohr is that therapy may not be the treatment of choice for all people with psychological concerns, shedding further light on the need for more informed cost-benefit analyses by clinicians. This position was echoed later by Dimidjian & Hollon (2010). While recognizing that most psychotherapy clients do not experience negative outcomes,

Mohr argued that it remains a worthwhile topic of inquiry in order to better determine which types of treatments are suitable for which clients, and to refine treatment delivery in general.

Mohr's review is acknowledged as increasing interest in negative psychotherapy outcomes within the field of psychology. Others have followed his example, providing their own research and commentary, both theoretical and empirical, on the topic. Among them are Boisvert & Faust (2002), who reviewed research discussing potential negative effects of therapist language, conceptualization, and diagnostic labeling. Using these data, they argued that a "pathology-oriented" framework or belief system used by therapists can significantly bias the judgments they make about clients and how they communicate with them once they have diagnosed them with a psychological disorder (p. 248). For instance, a therapist might too hastily interpret a given client behavior as stemming from his or her diagnosed psychopathology rather than considering the contributions of environmental factors (e.g. family dynamics, financial difficulties, experience of discrimination). In addition, the authors provided evidence from multiple studies that psychiatric labeling contributes to "iatrogenic symptoms" (p. 244). Such iatrogenesis includes client self-stigma and social rejection. Like Mohr (1995), these authors pointed out that clients may appear to worsen in therapy when there is a secondary gain for doing so (e.g. receiving attention, remaining in treatment). Relatedly, they suggested that an overly pathologizing formulation of a client may induce a self-fulfilling prophecy.

Boisvert and Faust also proposed ways of reducing the potential for the aforementioned negative outcomes, not by doing away with diagnostic nosology, but by avoiding use of pejorative language and by utilizing pathologizing terms or jargon (e.g. "self-defeating," "enabler") with caution. Emphasizing client strengths and agency, employing a solution-focused approach, and considering multiple explanations before making interpretations of client behavior

were also recommended. In addition, the authors suggested that highlighting supplemental sources of support for the client outside of therapy (e.g. friends, family, religious leaders) can assist with normalizing client concerns and can protect against overreliance on the therapist.

Relatedly, Nolan, Strassle, Roback, and Binder (2004) performed an exploratory review in order to generate ideas for prevention of and intervention with negative treatment effects. Among the suggested strategies were continuing education, supervision, and peer consultation for therapists; improving client-treatment match; and the use of assessment tools, such as outcome measures, feedback systems, and recovery curves. Similarly, Castonguay and colleagues (2010) proposed guidelines to help educate clinical trainees about the potential for harmful effects to occur in psychotherapy, prevent such effects when possible, intervene successfully when they do occur, and monitor treatment process and outcome. These guidelines stemmed from an examination of both theoretical and empirical literature on various topics pertaining to therapy process and outcome. The topics encompassed the therapeutic relationship, choice of treatment for particular presenting problems, client and therapist characteristics, and therapist techniques. A notable finding cited is the potential for relaxation training to cause increased anxiety in some clients with generalized anxiety disorder. According to Castonguay and colleagues, additional factors suggested by research to pose risks for poor outcome (though not necessarily deterioration) include perfectionism among depressed clients, client diagnosis of borderline personality disorder, lack of client motivation, therapist's frequent use of transference interpretations, therapist's overly rigid adherence to a particular theoretical rationale or technique when the client is resistant to it, mismatch between client and therapist on the variables of religion and age, and treatment of lesbian, gay, and bisexual individuals by a heterosexual male therapist. The authors additionally called for further research on client and therapist

characteristics that may negatively impact the treatment relationship or process, either generally or in combination. Furthermore, they advocated for the development of a list of potentially harmful errors in treatment technique similar to Lilienfeld's (2007) list of PHTs.

Berk and Parker (2009) also reviewed evidence from what they deemed a "sparse literature base" regarding the role of particular therapist, client, technique, and treatment (i.e. long-term psychodynamic versus time-limited evidence-based therapies) factors play in client harm (p. 788). They used the term "side effects" to describe any negative impact of treatment on the client, including deterioration. Examples of negative side effects of psychotherapy offered by the authors included increased anxiety, anger, and alcohol use; self-harm; hospitalization; decreased self-efficacy; and family problems. One major finding discussed was increased anxiety and symptoms of PTSD in individuals who received CISD. Additionally, the authors urged caution in the implementation of "high-risk" treatment techniques (e.g. confrontation, critical feedback, and arousal of strong emotions) or any interventions that could make a client feel high levels of stigmatization or blame (p.788). Risk for harm from CISD has been similarly addressed by Bootzin and Bailey (2005), who briefly reviewed its outcomes along with two other therapies associated with "iatrogenic" effects: group treatment of adolescents diagnosed with conduct disorder and therapies for dissociative identity disorder (DID).

Another narrative review comes from Lambert (2011), who addressed "treatment failure" during empirically supported therapy. Lambert described treatment failure as a broad category consisting of several suboptimal results of treatment. These include treatment nonresponse and deterioration/negative change, which he emphasized "are not the same problem and do not have the same implications for improving practice" (p. 414). His discussion went on to cover findings on client, therapist, and therapy factors associated with treatment failure. These included low

client motivation, multiple client comorbidities or otherwise complicated presentations, and therapist mistakes in treatment delivery. However, non-response rather deterioration was the primary type of treatment failure being investigated in the studies this author reviewed.

The post-Mohr (1995) reviews just discussed illustrate the wide array of variables and subtopics pertinent to the subject of client deterioration. Some of these reviews are speculative or theoretical/conceptual in nature. All lack a specified methodology by which the literature under review was searched, selected, compared, and synthesized. However, additional reviews on negative effects of psychotherapy exist and have employed a more systematic approach.

First, it is notable that Lambert and colleagues' development of the OQ-45 and their strong support of the use of outcome measures and feedback systems have assisted in making the case for continued study on deterioration. For instance, Lambert (2007) summarized the results of six studies conducted between 2002 and 2006, selected because they tested the ability of empirical algorithms based on OQ-45 scores to predict client deterioration. He concluded that such methods are effective, and that the essential pieces of information needed to achieve a prediction about treatment outcome are a client's self-reported symptom severity (as represented by the client's OQ-45 score) at 1) baseline and 2) following at least one session of treatment. According to Lambert, considering additional variables (e.g. therapist level of experience, choice of treatment, client diagnosis, age, sex, or ethnicity) beyond initial symptom severity and treatment response does not significantly improve prediction of outcome. These findings provoke questions about how frequently the OQ-45 (and similar instruments) is being used in current research and how its definition of deterioration in terms of global symptom severity compares to others'.

Previously mentioned in the Empirical Considerations section, another well-recognized and systematic review is found in Lilienfeld (2007). This author sought to create a provisional list of potentially harmful psychological treatments (PHTs). To qualify as PHTs, he specified that treatments had to meet the following criteria: "demonstrated harmful psychological or physical effects in client or others" (i.e. significant others, such as relatives), the harm was enduring rather than a temporary exacerbation of symptoms, and the harmful effects had been replicated by an independent party (p. 57). Further, he clarified that "psychological harm" referred to effects that appeared to be the direct result of treatment and could include both deterioration and slowed or non-response. Lilienfeld used several different data collection/selection methods. These included a keyword (i.e. *psychotherapy, therapy,* and *treatment* in combination with *harmful, harm, iatrogenic, worse,* and *deterioration)* search of the PsycINFO online database as well as consulting existing reviews on the topic (e.g. Mohr, 1995; Rhule, 2005 [who studied outcomes in youth with conduct problems]; and Werch and Owen, 2002 [who examined effects of prevention programs targeting drug and alcohol use]).

Studies consulted by Lilienfeld represented a range of designs, including RCTs, quasi-experiments, meta-analyses, and replicated single-case formats. PHTs were grouped, based on strength of evidence (i.e. replication, internal validity of the studies, and consistent emergence of new symptoms soon after the commencement of treatment), into two categories: those that *probably* cause harm to some of their recipients and those that *possibly* cause harm. Inclusion in the former category indicated a higher degree of evidence was available. The PHTs relevant to adult populations that were contained in that category included CISD (associated with increased risk of PTSD symptoms), recovered-memory techniques (generation of false memories for traumatic events), DID-focused therapy (introduction of "alter" personalities), grief counseling

for normal bereavement (increased depressive symptoms), and expressive-experiential therapies (increased experience of painful or hostile emotion). One PHT relevant to adults fell in the possible harm category: relaxation training for panic disorder (increased anxiety symptoms, including panic attacks). The author suggested that additional PHTs may exist; however, the evidence currently available to make the case for their status as a PHT was not sufficient to meet his inclusion criteria.

Castonguay and colleagues (2010) praised Lilienfeld's work and pointed out potential commonalities across the treatments he identified. For instance, they observed that CISD and grief counseling for normal bereavement could entail placing pressure on the client to confront emotion quickly and intensely. Similarly, recovered-memory interventions and therapies for DID may involve "use of powerful persuasion and suggestion" (p. 39). Encounters with intense emotion and suggestibility, the authors posited, are not necessarily harmful in themselves. Yet when paired with particular clients, presented too early or quickly, or facilitated by an unskilled therapist, they may contribute to deterioration.

While Lilienfeld (2007) focused on treatments that appear to systematically cause harm, a review by Ackerman and Hillsenroth (2001) focused on the therapist as a potential source of negative effects, specifically what personal characteristics and elements of treatment delivery may negatively impact the therapeutic alliance. They found that therapist rigidity, lack of confidence, tenseness, and distractibility, as well as excessive or inappropriate use of structure, self-disclosure, transference interpretations, and silence influenced the alliance in a negative way. Disruption in the client's relationship with the therapist is an area of interest closely related to client deterioration; unsurprisingly, many of these authors' findings are similar to those of Mohr (1995). The study did not address reliable or clinically significant deterioration per se. Its

methodology, however, could readily be applied to conduct a comprehensive review of factors related to deterioration. Specifically, Ackerman and Hillsenroth went beyond a standard online keyword search in order to identify papers to review, additionally consulting texts such as Bergin and Garfield's *Handbook of Psychotherapy and Behavior Change*, reviewing the reference sections of selected relevant articles, and manually reviewing the previous year's contents of selected journals. Manually reviewing journals as an adjunct to keyword searching is beneficial, especially when features such as full-text searching are not available in certain online databases. This is because particular topics may be contained in a paper yet are not part of the title or abstract, and thus may be overlooked by keyword searching.

A brief discussion of unwanted effects of psychotherapy by Linden & Schermuly-Haupt (2014), for instance, could have benefited from Ackerman and Hillsenroth's (2001) approach. Linden and Schermuly-Haupt conducted a keyword search of papers published between 1954 and 2014 listed in the PsycINFO and PubMed online databases. The authors used the keyword *psychotherapy* paired with *side effects, negative effects*, or *adverse events*. The PsycINFO search returned only 21 studies and the PubMed search 16. The authors used these results as evidence of a scant literature on these topics. While their line of thinking was not inaccurate, a significant limitation is that they searched for keywords in the article titles only, likely excluding a considerable number of relevant studies.

Addressing the question of how frequently harm (i.e. adverse events, side effects, deterioration) to participants is monitored and reported in psychotherapy outcome research, Jonsson and colleagues (2014) conducted a review limited to examination of RCTs only. Using the PsycINFO and PubMed databases, they identified studies published during 2010 investigating the effectiveness of psychological interventions for individuals with mental and

behavioral health disorders. Once eligible papers were collected, the authors searched the corresponding electronic files using the "find" feature of a PDF reader program. Keywords used at this level of the search included *harm, deteriorat* (truncated to include variations of the word), *side effect, side-, worse, safe, adverse,* and *impair*. Additionally, results sections of all the papers were reviewed manually to ensure that no reporting of harms had been overlooked. If such a report was found, the methods section of that paper was also reviewed to identify how harm had been defined and measured. Results demonstrated that just 28 (21%) of 132 trials meeting inclusion criteria explicitly stated that monitoring for any kind of harm to participants took place. Of these, nine trials noted that adverse events occurred, but only four specified the nature of the events and reported how they were measured. An additional four trials indicated that no adverse events took place, but their authors did not include discussion of how such events were defined and monitored. The 15 remaining studies commented on deterioration alone, as opposed to adverse events in general. The majority of these measured deterioration using a RCI or a clinician-administered rating scale called the Clinical Global Impression of Improvement (CGI-I). These findings are consistent with that of the previously discussed study by Duggan and colleagues (2014), demonstrating under-reporting of negative effects in the clinical trials literature. Jonsson and colleagues further found that trials of treatments for PTSD were the most likely to include data on harm. One limitation of the study is that it only covered one year's worth of publications.

Although they are beyond the scope of this discussion, more specialized reviews relevant to deterioration exist as well. They address, for example, the realms of group therapy (Roback, 2000), children and youth (Nicholson, Foote, & Grigerick, 2009; Rhule, 2005), prevention and treatment of substance misuse (Moos, 2005; Moos, 2012; Werch & Owen, 2002), self-

administered treatments (Scogin et al., 1996), mobile applications (Naeem et al., 2016), and Internet-based treatments (Rozental et al., 2014; Rozental, Boettcher, Andersson, Schmidt, & Carlbring, 2015).

Two recent meta-analyses on client deterioration in Internet-based treatments exist as well. Internet-based treatments here refer to treatments in which therapeutic material (e.g. psychoeducation, skill-building activities, feedback) is administered through a web-based platform and may or may not include interaction with a therapist. They are distinguished from telehealth treatments, which may be delivered through an Internet-based service (e.g. Skype) but involve direct, real-time contact with a therapist as the primary means of intervention delivery. Ebert and colleagues (2016), through a systematic database search as well as contact with study authors in order to request raw data, collected 18 RCTs that provided information on the pre-treatment to post-treatment reliable deterioration of adult participants who were experiencing a depressive episode and who were enrolled in Internet-based guided self-help treatments. The majority of these treatments were cognitive-behavioral in nature. Reliable deterioration was defined as significant negative change on the CES-D (Center for Epidemiologic Studies Depression Scale) or the BDI, according to the RCI. The authors reported the average observed deterioration rate as relatively low (3.6%). They also identified that client level of education moderated deterioration effects such that those who did not continue education after high school were at higher risk for deterioration, though they determined that for every one participant with lower education who deteriorated, there were 9.38 participants with lower education who responded favorably to treatment. Overall, the risk of deterioration was significantly lower among control groups than among Internet-based guided self-help groups. However,

deterioration rates did not significantly differ between these groups when only participants who did not attend higher education were considered.

Rozental, Magnusson, Boettcher, Andersson, and Carlbring (2017) also conducted a meta-analysis on client deterioration in Internet-based treatments. Their project focused on 29 clinical trials of cognitive-behavioral therapy, both guided and unguided, administered via the Internet. The authors utilized the RCI as the metric to establish the occurrence of deterioration and applied it to the primary outcome measure for each trial. Participants represented multiple presenting concerns, including depressive disorders, anxiety disorders, sexual dysfunction, and PTSD. The authors reported that 5.8% of those enrolled in the identified treatment conditions experienced reliable deterioration, whereas 17.4% of controls deteriorated. They concluded that controls were more likely to deteriorate. They also observed (without presenting formal analyses) that deterioration rates tended to be higher among individuals presenting with problems that likely involved a significant other, such as relationship conflicts and erectile dysfunction. The authors speculated that such presentations entail a level of complexity that it may be challenging to address through a self-guided treatment and/or via the Internet. Rozental and colleagues' results also demonstrated that, among participants assigned to a treatment condition, the likelihood of deterioration was lower for those who were of older age, those who were in a relationship, and those who held a university degree. For both those assigned to control conditions and those assigned to treatment conditions, higher initial symptoms severity was associated with lower odds for deterioration. The authors noted this as a surprising finding, but speculated that those who enter clinical trials in high distress may experience regression to the mean over time.

Since the publication of Mohr's (1995) review, researchers have sought to expand his work by continuing to explore solutions for how to define, scientifically study, predict, understand the mechanisms of, and develop prevention and intervention efforts against deterioration. Despite progress in the study of deterioration over the years, the concerns with terminology, methods of measurement, limited causal explanations, and the file-drawer problem that surround the subject of deterioration still exist. In spite of the attention Mohr's review has garnered (364 citations as of the date of this writing, according to Google Scholar), deterioration remains a relatively under-studied area, especially in terms of systematic, quantitative research (Berk & Parker, 2009; Lilienfeld, 2007; Scott & Young, 2016).

Reviews published on the topic within the last 20 years tend to lack the broad scope undertaken by Mohr, at least when considering adult populations in individual psychotherapy. Instead, they examined specific treatments, diagnoses, or theoretical orientations. Additionally, some fail to distinguish deterioration from other negative occurrences (e.g. no response to treatment, temporary negative influence on the therapeutic relationship, temporary fluctuation in symptoms). A number are not systematic or comprehensively thorough in their searching and selection of literature.

A notable exception is a systematic review completed by Cahill, Barkham, and Stiles (2010), though its primary focus was not deterioration. These authors reviewed 31 studies published between 1990 and 2008, obtained through a combination of online database searching and hand-searching of selected journals. Their search centered on adult samples receiving psychotherapy in routine clinical settings. The authors sought to compare outcomes in practice-based research to benchmarks for treatment response previously reported in efficacy studies (e.g. controlled trials). They excluded studies in which self-help, computer-based, or manualized

treatments were utilized. Two sets of analyses, one examining effect sizes, and one examining reliable and clinically significant improvement (RCSI; using Jacobson and Truax's [1991] method) were performed. For each analysis, 14 of the 31 selected studies supplied information sufficient for the specified metric to be calculated (these were not all the same studies, though some of them overlapped). Rates of deterioration were presented within the RCSI summary. Notably, the efficacy benchmark (Westen & Morrison, 2001) that Cahill and colleagues used for comparison did not include a benchmark rate for deterioration. Thus, the authors offered little commentary on the topic of deterioration in their paper. However, the deterioration rates they found across studies ranged from 0 and 13%, with a mean of approximately four percent and a modal rate of one percent. The majority of the studies took place in the United Kingdom and utilized the CORE-OM as their outcome measure. Other outcome measures represented included the BDI, Beck Anxiety Inventory (BAI), Brief Symptom Inventory (BSI), and OQ-45.

The past two decades' worth of reviews that capture information about client deterioration provide valuable information, raise awareness of deterioration as a significant clinical concern, and help to guide future research. However, exhaustive investigation of deterioration in general, including what the body of psychotherapy outcome research conducted in recent years might reveal about this topic when examined as a whole, is still warranted. Therapy techniques may have grown in effectiveness in recent years, as may have outcome measures, training of clinicians, and quality of research methodology used to study deterioration (Lambert, 2013[a]). Relatedly, the frequency with which deterioration occurs in adult individual psychotherapy may have changed over time. A commonly estimated figure is an average of 5-10% (Lambert, 2013[a]; Lambert & Ogles, 2004). Cahill and colleagues' (2010) results roughly align with this estimate, albeit encompassing a wider range. The same can be said of another

notable study conducted by Hansen, Lambert, and Forman (2002). This was a large-scale review of psychotherapy outcomes in both clinical trials and routine practice, measured using the OQ-45. For routine practice settings, deterioration rates were reported across categories: 6.6% for an employee assistance program treating adjustment problems using a time-limited model, 9.7% for a university counseling center, 14.1% for members of a local health maintenance organization (HMO), 7.5% for members of a national HMO, 3.2% for a university-supported training clinic offering community mental health services, and 10.2% for a state-supported community mental health center.

Mohr's review revealed, on average, a wider range, approximately 1-20%, but the broader variety of study designs, sample sizes, and criteria used to indicate deterioration in his review must be taken into account. As discussed, variation is likely to occur depending on what definition of deterioration is being used, who is making the judgment, and what measures are being employed. Another example is that literature reviewed by Lilienfeld (2007) indicated that higher rates (10-15%) exist among clients receiving treatment for substance use disorders. Having updated information about rates of deterioration is useful not only when communicating this problem to members of the field, but also when discussing treatment options with clients. Knowledge of deterioration rates can also aid clinicians and administrators in determining how the deterioration rates in their practices compare to an average.

Despite the apparent value of using outcome measures to detect deterioration, many clinicians do not use them in practice. In fact, those who do are in the minority (37%), according to a national survey of licensed psychologists conducted by Hatfield and Ogles (2004). These authors found that respondents who did not assess outcome cited concerns about the procedure taking up too many resources (e.g. time, money), burdening both client and clinician with extra

paperwork, and ultimately not providing valuable information. As demonstrated by Duggan et al. (2014) and Jonsson et al. (2014), the scope of this issue is not limited to clinical practice, but exists in research as well, specifically in RCTs. Looking beyond RCTs to further explore how frequently the potential for deleterious outcomes goes unmonitored or unreported by psychotherapy researchers would also be beneficial. Such an investigation could help clarify the degree to which deterioration remains an under-examined topic versus an emerging subject of interest in current literature. Raising awareness of the extent of the problem and the factors that may maintain it are important steps in generating solutions for change.

Purpose of Study

The most recent, comprehensive review on the subject of deterioration in adult psychotherapy clients is nearly 20 years old (Mohr, 1995). More contemporary reviews exist, but are narrower in scope. Thus, a broad, updated synthesis of the empirical literature on deterioration is needed to reflect current knowledge and clinical practices, as well as to guide future research in this under-studied area. This is the goal of the current systematic review.

Topics addressed will include the frequency with which researchers assess and report deterioration in outcome studies, what definitions and measures of deterioration are currently used by researchers, and what proportion of client samples deteriorate. When such information is available, these deterioration rates will also be reported across categories, including client diagnoses, therapist level of education, setting in which treatment is conducted, clinician's theoretical orientation and whether or not outcome measures were administered in an ongoing manner throughout treatment. This is intended to assist in detecting potential markers for deterioration stemming from client, therapist, or intervention variables.

By reporting on the current state of the deterioration literature, it is expected that the current review will help fill a gap in therapy process and outcome research. For instance, it may

increase clinicians' awareness and detection of deterioration and suggest ways of tailoring interventions to target the problem. This in turn could result in a more successful and cost-effective course of treatment for the client, and a reduction in risk of harm. It may also encourage more frequent reporting of deterioration in outcome studies. Overall, an updated review is likely to facilitate future research on this complex issue by further clarifying the ways in which deterioration occurs, how it may be defined and measured, and which of its aspects are still not well understood.

CHAPTER 3

METHODS

Inclusion and Exclusion Criteria

Influenced by the methodology of Ackerman and Hillsenroth (2001), Jonsson and colleagues (2014), Lilienfeld (2007), and in part, Cahill and colleagues (2010), data collection was accomplished by several methods. Broadly, data on client deterioration and other negative effects were sought from quantitative studies of treatment outcome. For each phase of data collection, eligible studies involved adult (aged 18+) participants who received individual psychotherapy, who met a definition of psychological distress (e.g. DSM diagnostic criteria, reaching a given threshold on a symptom measure, seeking mental health services) that was defined by the study authors, and who provided data on at least one measure of psychological outcome.

In keeping with the approach of Mohr (1995) and considering that the primary indicated treatment is psychopharmacological in nature, studies in which the majority of participants were diagnosed with psychotic disorders were excluded. Studies in which individual therapy was combined with a non-individual modality (e.g. family, couples, or group therapy), and in which outcomes for each modality were not distinguished from each other, were not included. Studies in which participants were receiving concurrent individual therapy and psychopharmacological interventions were permitted, as this is a common occurrence in naturalistic studies of psychotherapy outcome. Treatments that involved guided self-help (e.g. therapist-facilitated bibliotherapy, web-based interventions coupled with personalized feedback from a therapist) were included, although treatments that were completely self-administered and involved no direct input from a trained facilitator were not. As mentioned above in the Review of Related

Literature section above, when it comes to factors that may contribute to client deterioration, the current review is focused primarily on factors that are present in a therapeutic encounter. This is because such factors are likely to be more readily identified and manipulated by clinicians and researchers in the service of reducing the incidence of deterioration.

If multiple research papers that utilized the same data set were found, papers that did not provide any new or additional information about deterioration beyond the original paper were excluded. Also excluded were studies in which the entire sample consisted of treatment responders. For instance, this would refer to studies in which a subsample of participants who had improved (rather than deteriorating or experiencing no change) during therapy continued to be administered outcome measures after therapy had concluded in order to assess long-term, post-treatment maintenance or loss of therapeutic gains. References for meta-analyses, reviews, qualitative studies, and studies that included only one participant (e.g. case studies) were recorded if they mentioned negative treatment effects or client deterioration. Data were not systematically collected or analyzed from these studies, with the exception of meta-analyses that reported rates of deterioration, using a definition of deterioration that aligned with the one selected for this review (i.e. a statistically significant or reliable worsening of a therapy client's condition compared to baseline, using an established metric). One meta-analysis (Vittengl et al., 2016) met this criterion. Its data will be commented upon separately in the Results and Discussion section.

For studies meeting inclusion criteria, one variable that was tracked was how the authors defined and measured negative effects or deterioration. Studies that provided information about rates of deterioration were distinguished from those that reported on related constructs such as temporary worsening of symptoms, treatment non-response, treatment drop-out, relapse, and

adverse events. According to the distinctions drawn by a number of authors (e.g. Lambert, 2011), relapse was considered a separate construct from deterioration in the current review, as was treatment dropout. Although premature, unilateral termination of therapy can be associated with client deterioration, there are a variety of other possible reasons for it (Lilienfeld, 2007; Ravitz et al., 2011). Additionally, the current review aimed to focus primarily on deterioration that occurred between pre-treatment and post-treatment.

Database Search

One method of data collection was a systematic search of psychotherapy outcome studies through the online database PsycINFO. For this phase of data collection, only studies that reported on deterioration and other negative effects occurring during psychotherapy were included. This database search was conducted using the keywords *psychotherapy* and *therapy*, each in combination with *harm*, deterior*, iatrogen*, worse*, negative*, impair*, adverse*, side-** (to capture side-effects), and *safe**. Additional search parameters were that studies needed to include samples of adult participants only and needed to be published in a peer-reviewed, English-language journal between January 2011 and September 2016. This initial search returned 8,191 results. To further refine these results, several limiters were placed. This involved removing a number of "Classifications" (Group and Family Therapy, Schizophrenia and Psychotic States, Transportation, Medical Treatment of Physical Illness, Clinical Psychopharmacology, Speech and Language Therapy, Psychopharmacology) as well as a number of "Subjects" or "Subject: Major Headings" (Schizophrenia, Psychosis, Side Effects (drug), Drug Therapy). This process then limited the results to 1,566 articles.

Titles and abstracts of these studies were then reviewed for inclusion and exclusion criteria. Studies that appeared to meet inclusion criteria then underwent a more thorough review

of their Methods and Results sections to confirm eligibility. In total, 1,529 studies were disqualified because they did not meet inclusion criteria. Of these 1,529 disqualified studies, 54 had met the initial criteria specified in the Inclusion and Exclusion Criteria section above. However, because they did not make mention of deterioration or other negative effects, they were not included for further analysis.

Thirty-seven studies that met inclusion criteria and mentioned some form of client deterioration or negative effects remained. They were tracked via a spreadsheet and their texts were reviewed to determine whether they provided information sufficient for the reader to determine rates of reliable deterioration (e.g. according to the Reliable Change Index [RCI; Jacobson & Truax, 1991]). Eighteen of these studies did provide such information. The remaining 19 studies did not include rates of reliable deterioration and instead reported upon related concepts or alternate definitions, such as general negative effects of treatment and adverse events.

Manual Search of Selected Journals

Beyond the above-mentioned keyword search and normal review of the literature, data collection involved manually searching 19 pre-selected journals for studies in which deterioration was reported but was not necessarily the focus of the paper (and thus might be missed through keyword searching). A manual search was also chosen in order to allow for investigation of what proportion of outcome studies monitor and report on deterioration compared to those that do not.

A list of the journals selected for manual review is found in Appendix A. They were selected for their focus on publishing psychotherapy process and outcome studies and to

represent a range of theoretical perspectives. In addition, many of the studies cited in the literature review for the current paper were published in journals included in the list.

This manual search covered the January 2011 to July or July/August 2016 issues of the selected journals, which encompassed 5,532 articles. Titles and abstracts of these studies were reviewed for inclusion and exclusion criteria. Studies that appeared to meet inclusion criteria then underwent a more thorough review of their Methods and Results sections to confirm eligibility. In total, 5,122 studies were disqualified because they did not meet inclusion criteria. The 410 remaining studies that met inclusion criteria were tracked via a spreadsheet and their texts were then reviewed to determine whether they addressed client deterioration, what measures or definitions of deterioration they used, and whether rates of reliable deterioration could be gleaned from them. Three of these 410 studies overlapped with the 37 eligible studies found during the online database keyword search phase of data collection. They were not counted twice during data analysis.

Two hundred and thirty-six of the 410 studies provided means or general findings only in their Results sections and did not mention client deterioration or negative effects. As shown in Table 1, these studies will be referred to as (Category A). Seventy-one of the 410 studies only provided results that combined the proportion of participants who worsened during treatment with those who experienced no change. These studies will be referred to as (Category B). In these cases, authors either specifically reported the aggregated percentage of participants who experienced deterioration or no change during treatment, or the authors reported the percentage of participants who improved during treatment (leaving the reader to infer that the remaining percentage of participants either experienced deterioration or no change). Thirty-four of the 410 studies only provided information on negative effects other than reliable deterioration (e.g.

relapse, adverse events, loss of "sudden gains"), reported general findings about an aspect of treatment being associated with negative results, or reported upon deterioration in a limited way (e.g. reported that a certain percentage of participants had withdrawn from the study due to worsening symptoms but did not provide rates of symptom worsening across the whole sample). These studies will be referred to as Category C. An additional 13 of the 410 studies provided information that fell into both Categories B and C. Finally, 56 studies provided complete rates of reliable deterioration. These rates were either reported directly or able to be inferred because data for each individual participant were presented in the paper or it was directly stated that all participants improved and/or experienced no reliable change. These studies will be referred to as (Category D).

Table 1
Studies Included in Manual Search, Separated by Category

Category	Number of Studies	Description
A	236	Reported on means or general findings only, did not mention deterioration or other negative effects
B	71	Combined the proportion of participants who worsened during treatment with those who experienced no change
C	34	Reported on negative effects other than reliable deterioration (e.g. adverse events)
B and C	13	Reported information that fell into both Categories B and C
D	56	Supplied sufficient information to determine rates of reliable deterioration, distinct from no reliable change
Total	410	

Studies that were in Category B and studies that were in Category C were reviewed to determine how many provided their treatment outcomes in terms of percentages or proportions, and how many utilized outcome measures that would make the calculation of reliable deterioration rates for the whole sample feasible (examples of when calculation of reliable deterioration for the whole sample would not be feasible would include when the outcome measure was the presence or absence of diagnostic criteria, or abstinence versus non-abstinence from a target behavior). 82 studies met both of these parameters. Their authors were then contacted via email to inquire whether they computed and/or would be willing to provide rates of reliable deterioration that occurred among their participants. Authors of two studies from Category D were also contacted. This was because their articles presented reliable deterioration rates that were collected at follow-up periods only. Emails sent to these authors inquired whether they would be willing to provide deterioration rates for post-treatment as well.

In total, 76 authors were contacted, as some of the 82 studies had the same author who served as the primary contact for questions about the study. Half of the contacted authors did not reply (n=38). The remaining authors who did reply either indicated that they did not have data on reliable deterioration available or that they did have these data available (i.e. in the form of raw data, already calculated in the form of unpublished data, or already calculated in another published study). Data on reliable deterioration were included for further analysis, with some exceptions. For instance, several authors stated they would calculate deterioration rates from their raw data, but never followed up to provide these data. Additionally, several authors suggested consulting other works that they or their colleagues had published. If authors provided references for additional published studies that did not meet inclusion criteria and/or did not include rates of deterioration, they were not included for further analysis. Raw data which the

authors provided but for which deterioration rates had not been calculated were excluded from further analysis as well. Also excluded were data that were associated with a study that the author intended to publish in the future but for which a full manuscript was not yet available. Overall, data from ten studies were selected for inclusion.

Unpublished Data

Data were also sought through sources other than published journals. One criticism of literature reviews is that they often fail to report the results of unpublished studies. This practice is warranted when a manuscript has been rejected due to poor methodology. However, as noted previously, methodologically sound papers may go unpublished because they fail to show the hypothesized treatment effects. Overlooking such studies can result in reporting stronger effects than actually exist. In order to avoid this bias and address the concerns raised by Lilienfeld (2007) about how it impacts the study of deterioration, unpublished data on deterioration were sought.

One source consulted was conference proceedings from the annual international meetings of the Society for Psychotherapy Research. Abstracts for brief paper presentations were searched in the PDF versions of the 2011-2016 conference programs using the keywords *negative effect*, *negative outcome*, *iatrogenic*, *adverse*, *worse*, *deterior**, and *harm**. Identified abstracts were then reviewed to determine whether they met inclusion criteria. Fifty-nine abstracts either did appear to meet criteria, or their eligibility was unclear from the abstract alone. For these cases, the title and authors were searched via Google Scholar and cross-referenced with the studies obtained during the online database keyword search and manual search phases of data collection. This step was taken in order to determine whether the results presented in the brief paper had later been published.

If the data from a brief paper had later been published, the published study met inclusion criteria for the review, and the published study mentioned client deterioration, then the study was included for further analysis. Data from the same study were not analyzed twice. If later publication of the brief paper results did not occur or if this was not able to be determined, the first author of the brief paper was contacted via email to request further information about the data and whether rates of deterioration were available. Through this process, eight studies that reported reliable deterioration rates were found. Five of them had not already been found through the manual or keyword search phases of data collection; three of them had already been found during the manual search. Seven of the eight studies were published in journals and one was a dissertation. Overall, five studies were obtained through this phase of data collection that represented new data (i.e. had not been found through other methods of data collection).

An additional means of seeking unpublished data was through contact with the editors of the clinical journals listed in Appendix A. The journal editors were emailed and asked whether they had recently rejected any manuscripts related to client deterioration. If so, further information on the study was requested. The majority of journal editors who replied indicated that they could not recall any such manuscripts and/or would not have ways of readily searching their records to determine whether any had been submitted. One journal editor did provide references for two published studies that met inclusion criteria and reported rates of reliable deterioration (Callahan et al. 2014; Rieck & Callahan 2013).

Studies Reporting Reliable Deterioration Rates

Across all methods of data collection, reliable deterioration rates were obtained from 88 studies. As shown in Table 2, 50 of these were found through manual search only. One study was found in both the manual and online database keyword searches. Eight studies had been

presented as brief papers at a meeting of the Society for Psychotherapy Research and were later published. Five of these had also been found during the manual search. One study was a dissertation whose findings had also been presented as a brief paper at a meeting of the Society for Psychotherapy Research. Sixteen studies were found through the database search only. Two studies were found through contact with journal editors. Ten studies fell into Categories B and/or C of the manual search and their authors, when contacted by email to inquire whether they computed rates of deterioration for their study, replied to provide such rates. From them, the following information was recorded: year of publication, publishing journal, the study's country of origin, study design (e.g. naturalistic, RCT), descriptive information for the sample (i.e. sample size, ages, gender, race/ethnicity, diagnoses), information about the treatment (i.e. treatment setting, therapist level of education, theoretical orientation or type of treatment utilized, when and how often outcome measures were administered), what measure or definition of deterioration was used, and the proportion of the sample that reliably deteriorated.

Table 2
Studies Reporting Reliable Deterioration Rates

Source	**Number of Studies**
Manual search only	50
Database search only	16
Both manual and database searches	1
SPR brief paper search (later published) only	4
Both SPR brief paper search and manual search	5
Contact with journal editors	2
Contact with study authors from categories B and C of manual search	10
Total	88

CHAPTER 4

RESULTS AND DISCUSSION

The current review included multiple research questions and approaches to data collection. As such, the results of data collection and related commentary will be grouped together by topic within this chapter rather than results and commentary being presented in separate chapters. By providing consolidated information about the implications of a particular set of data directly following the presentation of those data, the aim is to aid the reader in navigating and comprehending the material.

The current chapter will first focus on the results of the manual search method of data collection. Focus will then shift to the measurement and incidence of reliable deterioration in studies obtained across all methods of data collection. Topics will include the diverse outcome measures utilized in such studies, as well as the rates at which deterioration occurred in various groups. The remainder of the chapter will cover selected studies which highlight factors that potentially contribute to deterioration. Recommendations for future research and clinical practice will be included throughout.

Manual Search of Selected Journals

Results of the manual search address the question of what proportion of outcome studies of individual psychotherapy with adults report rates of deterioration. Fourteen percent of the 410 outcome studies reviewed by this method supplied information sufficient to determine rates of reliable deterioration (Category D). Seventeen percent aggregated the proportion of participants who worsened during treatment with those who experienced no change (Category B). Eight percent provided information on negative effects other than deterioration, reported general results related to deterioration, or reported upon deterioration in a limited way (Category C). A

separate 3% of studies reported information that fell into both Categories B and C. The remaining 58% of the studies did not supply information related to client deterioration (Category A).

Thus, although client deterioration has been acknowledged for decades, the majority of the outcome studies found in the manual search never addressed the topic. As previously discussed in the Review of Related Literature section, there are many possible reasons for this occurrence. It is not the norm for participants in psychotherapy outcome studies to experience deterioration, so it may be that researchers, journal editors, and the field as a whole place more value on typical and/or favorable results, in turn devoting more emphasis to them in publications (Barlow, 2010; Mohr, 1995; Vaughan et al., 2014). Another proposed explanation is that authors or journal editors may wish to shift focus away from deterioration out of concern that such information might negatively color how a particular treatment or group of clinicians is perceived (Mohr, 1995). On the other hand, it is possible that in a number of outcome studies, no participants (or a small number) experienced deterioration, and thus the authors did not perceive it as necessary to include these rates in the publication. Whatever the reasons, under-reporting of deterioration rates conceals potentially valuable information. Psychotherapy researchers have already made strides in understanding, detecting, and intervening with client deterioration. Each new psychotherapy outcome study that is conducted is an opportunity to contribute to this ongoing effort. However, the opportunity is missed when authors fail to consider the potential for client deterioration, track it, and report it in a precise way.

When authors report results and conclusions that are based on the average outcomes or change scores of a sample group, they do account for some of the impact that reliable deterioration has on overall treatment effects. Still, neglecting to present the full array of

treatment outcomes—neglecting to acknowledge both nomothetic and idiographic data— has costs (Barlow, 2010; Steenkamp & Litz, 2013). Such a practice limits research consumers' ability to determine how frequently a treatment might negatively affect or harm its participants (Benjamin, 2015). It also does not encourage continued research into individual differences among participants as potential predictors or mechanisms of deterioration, an area of study which researchers have highlighted as warranting further investigation (Castonguay et al., 2010; Parry, Crawford, & Duggan, 2016). Progress toward isolating these variables can be achieved through an increased focus on the rates of participants who deteriorate, as well as by investigating what factors may distinguish those who deteriorate from those who do not. An examination of the studies in Category B further illustrates these points.

Recall that the Category B studies aggregated rates of deterioration and no change into a single rate representing participants who did not respond to the treatment in the expected direction or at all. Among these studies, there were a variety of definitions of treatment response versus non-response (e.g. remission versus non-remission, reliable change versus no change, clinically significant change versus no change) and a variety of outcome measures used to track them. When looking at those studies that utilized the RCI as their method of detecting deterioration, combined rates of participants who either did not experience reliable change or experienced it in a negative direction ranged from approximately zero percent to 73%. This wide range may be attributed to the fact that diverse participant diagnoses, treatment types, and sample sizes are represented. Still, it serves to demonstrate that studies which report only the proportion of participants who respond to treatment in the expected direction—or those which aggregate the proportion of participants who either deteriorated or experienced no change—provide an indirect, incomplete, and potentially misleading measure of deterioration. The results of the

current study demonstrate that researchers are reporting these aggregated data more often than they are reporting separate rates of reliable deterioration and no reliable change, despite assertions in past literature that the two are separate constructs (Lambert, 2011).

It is challenging and risky to make inferences about the potential negative effects of treatments using such aggregated data. For example, a study may report that 30% of its participants either deteriorated or experienced no change. This is useful in demonstrating that the treatment itself is effective for the majority of its participants. Yet this form of incomplete reporting masks the frequency with which the treatment is ineffective for a portion of its participants and the frequency with which it may be harmful. Such a state of affairs is problematic in a field in which avoiding and minimizing harm is a core ethical tenet (APA, 2002). A treatment that is associated with worsening symptoms in five percent of its participants and no reliable change in 25% would likely be considered safer than a treatment for which those proportions are reversed.

It is important to be able to identify cases in which a particular treatment is associated with both high reward (e.g. 70% improvement rate) and high risk (e.g. 25% deterioration rate). Such cases raise ethical questions about whether the potential benefits of the treatment outweigh the potential harms. They also raise empirical questions about whether there are interaction effects between client factors (e.g. diagnosis) and elements of the treatment (e.g. therapist level of experience, therapeutic techniques utilized) that heighten the risk of deterioration for specific individuals or groups. This topic also has implications for informed consent. If the research literature is lacking in clear, distinct information about the rates of deterioration and improvement for a given treatment, clinicians may be limited in their ability to discuss the risks and benefits of the treatment with clients who are considering participating in it.

Efforts to further distinguish deterioration from related and potentially overlapping constructs may assist future researchers as they attempt to make cross-study comparisons of deterioration rates and to detect mechanisms of deterioration. One solution could involve journal editors requiring that published outcome studies include a defined method of measuring client deterioration (whether this is reliable change or some other metric) and that rates of deterioration be presented separately from rates of improvement and no change. Using methods such as the RCI could readily aid researchers in making these distinctions. Granted, not all outcome measures or research questions are compatible with these methods (Tarescavage & Ben-Porath, 2014). However, when they are able to be used, such calculations and reporting methods would likely not consume unreasonable amounts of researchers' time, nor space within their publications. If reporting deterioration rates becomes the norm, it may promote a culture of increased transparency, dialogue, and investigation related to the topic of psychotherapy effectiveness. The phenomenon of client deterioration is not always pleasant or convenient to acknowledge. Nevertheless, doing so is, in the long run, likely to improve the chances of delivering ethical and even more effective psychotherapies.

Studies Reporting Reliable Deterioration Rates (Multiple Search Modalities)

The 88 studies providing rates of reliable deterioration were drawn from 29 different journals (when published) and their authors represented 19 different countries throughout North America, South America, Europe, Oceania, and Asia. The United States was the country most commonly represented, followed by the United Kingdom (UK). These studies encompassed a range of study designs, among them randomized controlled trials, controlled trials, multiple baseline and single-case designs, case series designs, and naturalistic studies. The majority were

naturalistic studies. Sample sizes ranged from 3 to 17,520. A large majority of the participants identified themselves as White/Caucasian and female.

Deterioration was reported in diverse ways. Rates were presented for a number of different outcome measures that were administered at varying times (e.g. pre-treatment and post-treatment only, at a follow-up point, multiple times during treatment). Domains of client functioning assessed by the various outcome measures included disorder-specific symptoms (e.g. symptoms of eating disorders, obsessive-compulsive and related disorders, posttraumatic stress disorder (PTSD), depression, anxiety, psychosis, mania, grief), interpersonal/social functioning, self-compassion, sexual functioning, sleep, substance use, vocational functioning, violence, suicidality, quality of life, and cognitive functioning. The data reported by the authors may have come from an intention-to-treat sample, a completer sample, or both. In the vast majority of the studies, the RCI was the method used to calculate deterioration. Additional methods defined by authors included the percentage method (e.g. deterioration of 50% or more in baseline scores on an outcome measure; see Hiller, Schindler, & Lambert, 2012 and Ogles, 2013) and effect size. Rates of reliable deterioration discussed in the remainder of the current review will typically be rounded to the nearest whole number.

Use of Diverse Outcome Measures to Detect Deterioration

The range of outcome measures that the studies gathered for this review utilized to report deterioration rates merits further discussion. These measures included disorder- or symptom-specific measures as well as global measures. Some measures focused on a single dimension of functioning while others monitored change across multiple domains. As presented in the Review of Related Literature section, Ogles (2013) has emphasized that the diversity of extant psychotherapy outcome measures reflects recognition of the complex nature of treatment, client

change trajectories, and the many facets of psychological functioning. However, he asserted that this also creates confusion for those seeking to draw cross-study conclusions from the psychotherapy outcome literature.

For instance, if a variety of studies are being compared and they do not utilize a common instrument (or core battery of instruments) to measure change, it is unclear to what degree divergent findings between these studies may be attributed to inherent differences in the outcome measures, and thus obscure inferences about the treatment being studied. Indeed, Ogles'(2013) discussion cited multiple studies which demonstrated that outcome measures may significantly differ in the effect sizes they produce, despite purportedly measuring the same construct. Considering this, advantages of the use of a common outcome battery by psychotherapy researchers become apparent. One major benefit is that it would enable researchers to more readily conduct meta-analyses to explore whether comparable outcome measures have a differential impact on rates of deterioration.

At least in terms of studies reporting upon deterioration, the field does not seem to have fully embraced the notion of a core battery in practice yet. However, there were three groups of outcome measures that emerged as the most commonly administered in the studies collected for this review that provided information on reliable deterioration. These three groups included versions of the Beck Depression Inventory (BDI, BDI-II; used in 18 different studies), the Outcome Questionnaire-45 (OQ-45, OQ-45.2; used in 16 different studies), and the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; CORE-10; used in 12 different studies). Each of these measures are self-reports, can be routinely administered, and lend themselves to the calculation of a RCI. Yet while the BDI is focused on depressive symptomatology, the CORE and OQ instruments are broader in nature, assessing not only

symptoms but also risk of harm to self or others, functioning in interpersonal relationships, and functioning related to daily roles and responsibilities (e.g. work, school). Given the differences between these instruments, and the array of additional measures employed in the other studies gathered for this review, it is challenging to interpret and draw conclusions about the deterioration rates they detect. Forthcoming discussion will attend to specific examples of these challenges in greater depth.

General Rates of Deterioration at Post-Treatment

Among the studies gathered for the current review that contained deterioration rates, the majority of these rates referred to significant worsening that occurred between pre-treatment and post-treatment. Table 3 provides data for each of these studies. The data presented are drawn from samples who participated in treatment conditions that met the inclusion criteria defined for this review (e.g. consisted of individual therapy; were not completely self-guided interventions). Each sample size refers to the number of participants within that study whose pre- and post-treatment outcome data were analyzed by the author(s) in order to provide rates of deterioration. Thus, in some cases, the sample size listed in the table will represent only a portion of the total sample.

When viewing the studies in Table 3, one notable aspect is that the sample sizes vary widely. In an attempt to exercise caution in interpreting the results of studies with small sample sizes, studies with sample sizes > 15 (approximately three-fourths of the studies listed in Table 3) were examined further. Among these studies, reported deterioration rates ranged between 0 and 32%, with approximately 92% of the studies reporting rates between 0 and 10%. These ranges were not evenly distributed, as can be seen in Figure 1. Figure 1 depicts the frequencies of deterioration rates associated with the studies presented in Table 3 whose sample sizes exceeded

15. It should be noted that each rate does not necessarily correspond to a single study; some studies reported multiple rates (e.g. for different outcome measures, for different treatment conditions). As shown in the figure, the most commonly reported rate of deterioration was 0% and rates tended to cluster between zero and five percent . It was noted that studies with sample sizes smaller than 15 tended to produce deterioration rates representing either high (14-20%) or low (0%) extremes.

Table 3
Rates of Reliable Deterioration That Occurred Post-Treatment

Study	**Sample Size**	**Outcome Measure(s)**	**Deterioration Rate(s)**
Aafjes-van Doorn , Macdonald, Stein, Cooper, & Tucker (2014)	31	IIP-32, RS, SCQ-SF	IIP-32: 3% RS: 10% SCQ-SF: 10%
Alves et al. (2014)	6	BDI-II, ICG	0%, both measures
Ashworth et al. (2015)	21 (self-help condition excluded)	BDI-II, ISI	BDI-II: 0%, ISI: 6%
Aubochon-Endsley, Callahan, & Scott (2014)	65	BAI, BDI-II	BAI and/or BDI-II: 8%
Bambling & King (2013)	92	BDI	0%
Barr, Hodge, Levan, Bowen, & Knox (2012)	35	CORE-OM	9%
Beck, Burdett, & Lewis (2014)	103	CORE-OM GD score	3%
Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)	5	PGS, WOC (E-AS)	0%, both measures
Bernecker et al. (2016)*	57	BDI	2% using intake as baseline, 4% using session 1 as baseline

BAI (Beck Anxiety Inventory), BDI/BDI-II (Beck Depression Inventory), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), CORE-OM GD score (Clinical Outcomes in Routine Evaluation—Outcome Measure, Global Distress score), ICG (Inventory of Complicated Grief), (IIP-32 (Inventory of Interpersonal Problems-32), ISI (Insomnia Severity Index), PGS (Perinatal Grief Scale), RS (Remoralization Scale), SCQ-SF (Self-Compassion Questionnaire-Short Form), WOC (E-AS; Ways of Coping, Escape-Avoidance Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Bevan, Wittkowski, & Wells (2013)	6	EPDS, HADS	0%, both measures
Bhattacharya (2015)	3	BDI	0%
Boersma, Håkanson, Salomonsson, & Johansson (2015)	6	SCS, SIAS	0%, both measures
Bosley, Fisher, & Taylor (2016)	7	Visual analog survey	Negative affect (sad): 14%, Negative affect (irritability): 0%, Negative affect (fear): 0%, Worry: 0%
Boswell, McAleavey, Castonguay, Hayes, & Locke (2012)	864	CCAPS-62 Depression subscale	4%
Branson, Shafran, & Myles (2015)	1,247	GAD-7, PHQ-9 (RCI for both measures combined)	6% when student therapists were in CBT training, 3% six months after training had concluded
Briggie, Hilsenroth, Conway, Muran, & Jackson (2016)	243	GSI of BSI	5%
Bryan et al. (2012)	495	BHM-20 GMH scale	5%

BDI (Beck Depression Inventory), BHM-20 GMH scale (Behavioral Health Measure, Global Mental Health scale), CBT (cognitive-behavioral therapy), CCAPS-62 (Counseling Center Assessment of Psychological Symptoms), EPDS (Edinburgh Postnatal Depression Scale), GAD-7 (Generalized Anxiety Disorder Scale), GSI of BSI (Global Severity Index of Brief Symptom Inventory), HADS (Hospital Anxiety and Depression Scale), PHQ-9 (Patient Health Questionnaire), RCI (reliable change index), SCS (Self-Compassion Scale), SIAS (Social Interaction Anxiety Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Butollo, König, Karl, Henkel, & Rosner (2014)	21	IES-R, PDS	0%, both measures
Callahan et al. (2014)	216	OQ-45.2	Collected data from six clinics. Rate was 40% for one clinic, which was classified as an outlier and removed from further analysis. Among results for remaining clinics, range was 0-11% and mean was 7%.
Chow et al. (2015)	4,580 clients, 69 therapists	CORE-10	Therapists grouped into four categories from best - (1) to worst- performing (4). Respective mean rates were: 2%, 2%, 3%, 5%.
Dakwar & Levin (2013)	19	TLFB (measured drug use)	0%
de Jong (2012)**	1,494	OQ-45	9%
de Jong et al. (2014)	475 (232 in short-term [< 35 weeks] therapies, 243 in long-term [≥ 35weeks] therapies)	OQ-45	Full sample: 8% (NFb), 11% (FbT), 5% (FbTP). Short-term: 10% (NFb), 13% (FbT), 7% (FbTP). Long-term: 6% (NFb), 9% (FbT), 4% (FbTP).

FbT (feedback to therapist condition), FbTP (feedback to therapist and patient condition), IES-R (Impact of Event Scale-Revised), NFb (no feedback condition), OQ-45/OQ-45.2 (Outcome Questionnaire), PDS (Posttraumatic Diagnostic Scale), TLFB (Time Line Follow Back)

Rates have been rounded to the nearest whole number
**Dissertation

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Dennhag & Armelius (2012)	187 (full sample)	GSI of SCL-90-R, SASB affiliation dimension	GSI of SCL-90-R: 3% (full sample), SASB affiliation dimension (full sample): 0%
Dennhag, Ybrandt, & Armelius (2011)	235	GSI of SCL-90-R	3%
Farima, Dowlatabadi, Behzadi (2015)	3	CAQ, IUS, PSWQ, WDQ	0%, all measures
Feldner, Smith, Monson, & Zvolensky (2013)	4	CAPS, PDS, cigarettes smoked per day	0%, all measures
Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)	98	ORS (part of PCOMS)	6%
Frets, Kevenaar, & van der Heiden (2013)	6	BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures
Galovski, Blain, Mott, Elwood, & Houle (2012)	50	BDI-II, PDS	0%, both measures
Goldberg et al (2016)*	6,591	OQ-45	4%

BAI (Beck Anxiety Inventory), BDI-II (Beck Depression Inventory), BFNE (Brief Fear of Negative Evaluation scale), CAPS (Clinician-Administered PTSD Scale), CAQ (Cognitive Avoidance Questionnaire), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), IUS (Intolerance of Uncertainty Scale), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDS (Posttraumatic Diagnostic Scale), PSWQ (Penn State Worry Questionnaire), SASB (Structural Analysis of Social Behavior), SPAI-N (Social Phobia Anxiety Inventory), WDQ (Worry Domains Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Gonçalves et al. (2012)	6	BDI	0%
Goodman, Edwards, & Chung (2013)	3	GSI of SCL-90-R	0%
Gumz, Bauer, & Brähler (2012)	9	GSI of SCL-90-R	0%
Hardy, Tracey, Glidden-Tracey, Hess, & Rohlfing (2011)	210	OQ-45	0%
Hayes, Owen, & Bieschke (2015)	228 (148 White participants, 80 Racial/ Ethnic Minority participants)	OQ-45	5% of White participants, 3% of Racial/Ethnic Minority participants
Kellett, Bennett, Ryle, & Thake (2013)	17	CORE-OM	6%
Koszycki, Bilodeau, Raab-Mayo, & Bradwejn (2014)*	23	CGI-S	0%
Kramer, de Roten, Drapeau, & Despland (2013)	50	GSI of SCL-90-R	6%
Kramer, Pascual-Leone, Despland, & de Roten (2015)	32	BDI-II	6%
Kramer et al. (2011)	20	GSI of SCL-90-R	5%

BDI/BDI-II (Beck Depression Inventory), CGI-S (Clinical Global Impression-Severity scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), OQ-45 (Outcome Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Kraus, Castonguay, Boswell, Nordberg, & Hayes (2011)	6,960	TOP	Reported across TOP domains. Sexual functioning: 23%, work functioning: 23%, violence 16%, social functioning: 32%, panic/anxiety: 25%, substance abuse: 19%, psychosis: 23%, quality of life: 24%, sleep: 25%, suicidality: 15%, depression: 20%, mania: 8%.
Leibert & Dunne-Bryant (2015)*	81	OQ-45.2	1%
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5	CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.
Littleton, Grills, Kline, Schoemann, & Dodd (2016)	18 completed CES-D, 20 FDAS, 23 PSS-I (interactive program only)	CES-D, FDAS, PSS-I	CES-D: 11%, FDAS: 20%, PSS-I: 0%
Lopes et al. (2014)	40 (20 in NT, 20 in CBT)	BDI-II, OQ-45.2	OQ-45.2: 3% in NT, 10% in CBT. BDI-II: 0% in both groups.

BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), CES-D (Center for Epidemiologic Studies Depression Scale), FDAS (Four Dimensional Anxiety Scale), NT (narrative therapy), OQ-45.2 (Outcome Questionnaire), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), TOP (Treatment Outcome Package), VK-MFS (Veronen-Kilpatrick Modified Fear Survey)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Lutz et al. (2014)	326	PDSS-SR	1%
McElvaney & Timulak (2013)	11	CORE-OM	18%
McLay et al. (2011)	10 (VR-GET only)	CAPS	0%
Mechler & Holmqvist (2015)	1,157 (840 in primary care, 317 in psychiatric clinic)	CORE-OM	Primary care: 2%, Psychiatric clinic: 7%
Mellor-Clark, Twigg, Farrell, & Kinder (2013)	17,520	CORE-OM	2% average across six sites
Murphy, Rashleigh, & Timulak (2012)	110	ORS (part of PCOMS)	5% in feedback condition, 10% in no feedback condition
Murray, McKenzie, Murray, & Richelieu (2015)	305	CORE-OM	2%
Payne, Ciclitira, Starr, Marzano, & Brunswick (2015)	98	CORE-OM	2%
Polman, Bouman, van Geert, de Jong, & den Boer (2011)	7	OBQ, PI-R, Y-BOCS	OBQ: 14%, PI-R: 0%, Y-BOCS: 0%
Rieck & Callahan (2013)	133	OQ-45.2	8%

CAPS (Clinician-Administered PTSD Scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), OBQ (Obsessive Beliefs Questionnaire), OQ-45.2 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PI-R (Padua Inventory-Revised), VR-GET (virtual reality-graded exposure therapy), Y-BOCS (Yale-Brown Obsessive-Compulsive Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Rogers et al. (2014)*	41 (received HRT, not specified how many completed post-treatment assessment)	MGH-HPS	0%
Saxon & Barkham (2012)	10,786 clients, 119 therapists	CORE-OM	Therapists grouped into categories: above average, average, and below average. Respective mean rates were: 1%, 1%, 2%.
Saxon, Barkham, Foster, & Parry (2016)	6,405 participants, 85 therapists	CORE-OM	0-7% range across therapists. Therapists grouped into categories: above average, average, and below average. Respective mean rates were: 0%, 1%, 3%.
Simon, Lambert, Harris, Busath, & Vazquez (2012)	207 (98 in TAU, 109 in feedback + CST)	OQ-45	12% in TAU, 6% in feedback + CST
Strauss et al. (2013)*	383 (quality assurance intervention group, TK study only)	BSI	6%

BSI (Brief Symptom Inventory), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), CST (clinical support tools), HRT (habit reversal training), MGH-HPS (Massachusetts General Hospital Hairpulling Scale), OQ-45 (Outcome Questionnaire), TAU (treatment as usual), TK (Techniker Krankenkasse [German health insurance company])

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Taillon, O'Connor, Dupuis, & Lavoie (2013)	10	Y-BOCS (modified for BDD)	0%
Titov et al. (2015)	2,049	GAD-7, PHQ-9	GAD-7: 3%, PHQ-9: 2%
Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)	39 (patients only)	COWAT, Digit Span, RAVLT IR/DR, RCFT IR/DR, Stroop Test, TMA, TMB, ToL, WCST-64	RCFT DR: 5%, Stroop: 8%, TMA: 5%, TMB: 8%. 0%, all other measures.
Vaz, Conceição, & Machado (2013)	30	EDE-Q	EDE-Q food concern subscale: 3%, EDE-Q shape concern subscale: 3%
Vromans & Schweitzer (2011)	38	BDI-II, OQ-45.2 IR subscale	BDI-II: 5%, OQ-45.2 IR subscale: 3%
Weiss, Kivity, & Huppert (2014)	19	ASI-3, PDSS-SR	0%, both measures
Werbart, Levin, Andersson, & Sandell (2013)	177 completed SRH, 175 GSI, 176 QOLI (full sample)	GSI of SCL-90, QOLI, SRH	GSI: 3%, QOLI: 2%, SRH: 0%

ASI-3 (Anxiety Sensitivity Index), BDD (body dysmorphic disorder), BDI-II (Beck Depression Inventory), COWAT (Controlled Oral Word Association Task), EDE-Q (Eating Disorder Examination Questionnaire), GAD-7 (Generalized Anxiety Disorder Scale), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), OQ-45/OQ-45.2 IR subscale (Outcome Questionnaire, Interpersonal Relations subscale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PHQ-9 (Patient Health Questionnaire), QOLI (Quality of Life Inventory), RAVLT IR/DR (Rey Auditory and Verbal Learning Test, immediate recall/delayed recall), RCFT IR/DR (Rey Complex Figure Test, immediate recall/delayed recall), SRH (Self-Rated Health), TMA/TMB (Trail Making Test A/B), ToL (Tower of London), WCST-64 (Wisconsin Card Sorting Test), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Willson, Veale, & Freeston (2015)	6	BDI, Y-BOCS (modified for BDD)	0%, both measures
Yasky, King, & O'Brien (2015)	11	PHQ-15	0%

BDD (body dysmorphic disorder), BDI (Beck Depression Inventory), PHQ-15 (Patient Health Questionnaire), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

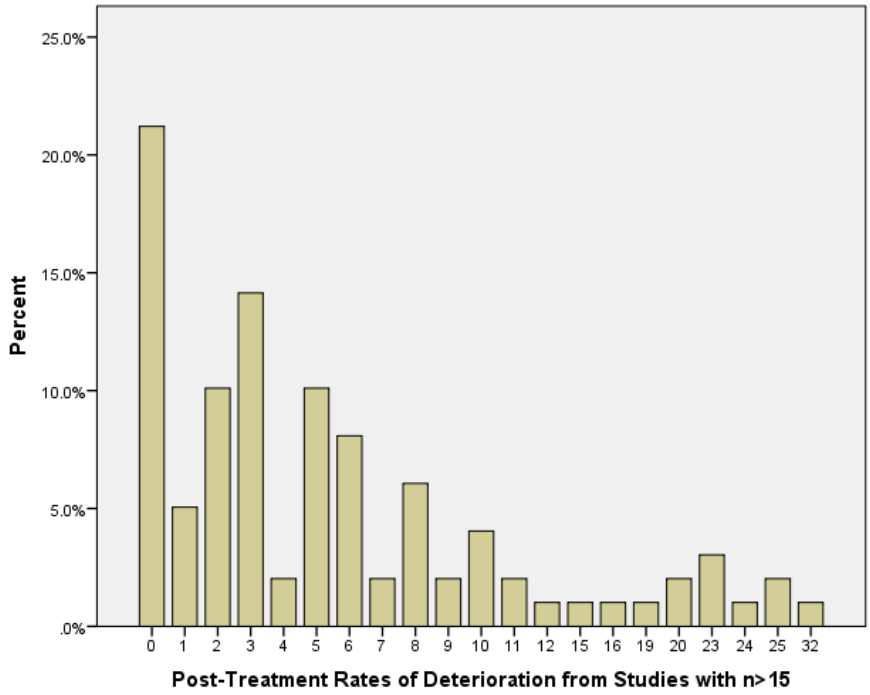

Figure 1. Distribution of post-treatment rates of deterioration from studies with sample sizes larger than 15.

Overall, these data suggest that rates of deterioration in psychotherapy outcome studies from 2011-2016 are somewhat lower than the estimated 5-10% average (Lambert, 2013[a]; Lambert & Ogles, 2004), as well as other rates reported in past research. For instance, in Shimokawa, Lambert, and Smart's (2010) meta-analytic review of a quality assurance system incorporating the OQ-45, rates of deterioration at treatment termination fell between 6 and 20% for the efficacy sample. Similarly, in Mohr's (1995) review, the vast majority of studies reported rates of client worsening between 0 and 20% (single-case studies and studies failing to report deterioration separately from no change excluded). A notable caveat when making comparisons to Mohr's work is that he broadly examined negative effects of treatment rather than limiting his scope to statistically reliable or significant worsening, as this review did. Additionally, not all of the studies he investigated utilized formal outcome measures. Thus, this may explain in part the

more frequent incidence of higher deterioration rates in his review. Somewhat similar to the current data, however, past benchmarking studies for the CORE-OM have produced rates of deterioration between 0 and 6% with an average of 2% in primary care counseling within the UK, and between 0 and 17% with an average of 1% in higher education counseling within the UK (CORE IMS, 2010[a,b]).

The prevalence of deterioration rates of zero percent in the current review is challenging to interpret. Perhaps this signifies that methods of treatment and assessment have become more effective in terms of early identification of, prevention of, or intervention for deterioration over time. An alternate explanation is that studies with low or no incidence of deterioration may be more readily published, and thus studies with higher deterioration rates are less prevalent in the literature due to the existence of a file-drawer problem. Again, it should also be noted that deterioration rates can vary depending upon methodological factors such as the outcome measure that is used (Ogles, 2013). Thus, perhaps a number of the outcome measures that were utilized within the studies collected for the current review demonstrate a floor or ceiling effect when it comes to detecting deterioration.

In a few studies represented within Table 3 and Figure 1, the reported rates of reliable deterioration were notably high. One study, conducted by Kraus, Castonguay, Nordberg, Boswell, and Hayes (2011), examined the naturalistic outpatient treatment of 6,960 individuals with various diagnoses reported rates of deterioration on the Treatment Outcome Package (TOP) in a variety of domains: sexual functioning (23%), work functioning (23%), violence (16%), social functioning (32%), panic/anxiety (25%), substance abuse (19%), psychosis (23%), quality of life (24%), sleep (25%), suicidality (15%), depression (20%), and mania (8%). This study was the only one included in the current review that both reported reliable deterioration rates and

used the TOP, a multidimensional self-report, as an outcome measure. As seen in the listing of domains above, the instrument assesses areas of daily life and functioning as well as symptoms relevant to particular psychological disorders. In a seminal research paper on the TOP (Kraus, Seligman, & Jordan, 2005), deterioration rates across domains ranged between 6 and 20%, with 67% of participants deteriorating in at least one domain despite the fact that 91% reliably improved in at least one domain. These studies highlights the fact that deterioration is not "all or nothing"; it is a dynamic, multi-faceted phenomenon.

Another example of a study reporting an outlying deterioration rate for one of its measures was a randomized controlled trial conducted by Littleton, Grills, Kline, Schoemann, and Dodd (2016). This trial investigated the effectiveness of an Internet-based, therapist-facilitated cognitive-behavioral treatment program for PTSD called From Survivor to Thriver. The sample consisted of women who had experienced rape and had been subsequently diagnosed with PTSD. None of the 23 participants who received Survivor to Thriver and completed a post-treatment measure of PTSD symptoms (PTSD Symptom Scale-Interview [PSS-I]) experienced deterioration on this measure relative to their pre-treatment scores. However, 11% deteriorated on a measure of depression and 20% on a measure of anxiety. The authors commented specifically on some these participants, stating that the two who deteriorated in terms of depression had each reported experiencing the recent death of a family member, which may have contributed to worsening depressive symptoms. Thus, it is important to remember that environmental factors in a client's life outside of psychotherapy can impact progress in treatment. The authors also observed that two of the four women who deteriorated in terms of anxiety still experienced significant reductions in PTSD symptoms, which were the primary target of the intervention.

As is apparent in Kraus and colleagues' (2011) and Littleton and colleagues'(2016) studies, as well as a number of studies throughout Table 3, rates of deterioration may vary notably depending upon what specific outcomes are being monitored and what outcome measures are utilized. In order to obtain a richer understanding of client deterioration and to be able to intervene in a more targeted way, clinicians and researchers alike may benefit from administering outcome measures at routine intervals throughout treatment. Although therapeutic change at the end of treatment is the anticipated goal, treatment may be more efficient if progress is being monitored consistently, rather than only at baseline and termination. Administering outcome measures regularly may also be advantageous in that progress in treatment is not always linear. Although a client may achieve a positive conclusion to therapy, they may also undergo worsening along the way that an outcome measure may assist in more rapidly and effectively detecting (Lambert, 2011).

Clinicians and outcome researchers may also consider the use of multiple outcome measures, or outcome measures that monitor multiple domains of functioning. There are cases in which a client's presenting concerns are quite circumscribed (e.g. a specific phobia) and the outcomes to which their treatment plan is directed are well-operationalized and directly measurable. Even in such circumstances, a mental health professional may do well to assess and approach a client's psychological concerns from a holistic perspective. This is not to say that clinicians should abandon focused treatment, attempting to thoroughly and directly attend to all aspects of a client's wellbeing at every meeting. Yet consider that when a physician is treating a patient's primary presenting complaint, the physician may perform additional relevant testing in order to identify comorbid factors that could complicate the primary treatment or contribute to the patient's distress. For example, a psychiatrist prescribing a drug to treat depression would

likely not only assess the patient's depressive symptoms at each visit but also assess for any adverse side effects, contraindicated use of alcohol, and so on. In a similar way, psychological concerns may be comorbid in nature (e.g. anxiety and depression, trauma and substance abuse, depression and suicidality) and difficulties in daily functioning, relationships, and other life stressors are often relevant to psychological concerns. Thus, it seems good practice for a clinician to judiciously select an outcome measure (or battery) that allow for assessment of multiple domains.

Further, it could be particularly useful to employ outcome measures that have been empirically researched such that outcome benchmarks for them can be determined. For example, this would allow researchers to identify whether deterioration rates in a new outcome study could be considered average relative to the rates that are typical for the outcome measure(s) they have chosen. It could also allow clinicians or administrators of treatment centers to determine whether the rates of deterioration for their particular caseload or setting are within expected limits given the outcome measure(s) they utilize. The discussion of variation in deterioration rates across outcome measures, as it applies to two specific measures (the CORE-OM and the OQ-45), will be expanded upon in the Therapist Effects and Effectiveness sub-section of General Findings Regarding Deterioration and Other Negative Effects below.

Rates of Deterioration Throughout Treatment

A small number of studies reported rates of reliable deterioration that occurred at any point during treatment rather than only comparing pre- and post-treatment scores. Authors of these studies were able to gather such information by administering outcome measures regularly as treatment was ongoing. These studies are listed in Table 4.

Table 4
Rates of Reliable Deterioration That Occurred Throughout Treatment

Study	**Sample Size**	**Outcome Measure(s)**	**Deterioration Rate(s)**
Amble, Gude, Stubdal, Andersen, & Wampold (2015)	259 (144 in feedback condition, 115 in no feedback condition)	OQ-45.2	Deteriorated between first session and any subsequent session: 6% in feedback condition, 9% in no feedback condition
de Jong, van Sluis, Nugter, Heiser, & Spinhoven (2012)	413	OQ-45	Deteriorated at any time during treatment: 16%
Erekson, Lambert, & Eggett (2015)	6,184 (3,092 in each group; randomly drawn from a larger sample and matched based upon age, gender, and initial symptom severity)	OQ-45	6% for participants attending sessions weekly, 9% for participants attending sessions biweekly
Murphy & Cramer (2014)	62	CORE-OM	Deteriorated between sessions 1 and 3: 3%
Richards, Timulak, & Hevey (2013)	25 in eCBT	BDI-II	0% at each administration
Tejas Haugen, Goldman, & Owen, 2015	36	OQ-45.2	"Sudden deteriorations" (worsening of symptoms that occurred between any session N and any session N+1): 27%

BDI-II (Beck Depression Inventory), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), eCBT (email-assisted cognitive-behavior therapy), OQ-45/OQ-45.2 (Outcome Questionnaire)

Rates have been rounded to the nearest whole number

One study (Tejas Haugen, Goldman, & Owen, 2015) utilized a unique definition of deterioration and reported a relatively high deterioration rate. The authors studied "sudden deteriorations" on the OQ-45.2 among 36 individuals diagnosed with PTSD who received integrated psychotherapy (combined cognitive-behavioral and psychodynamic approaches). Sudden deterioration was defined as clinically significant (according to the RCI) worsening of symptoms that occurred between any session N and any session N+1. This occurred among 27% of the sample, and the authors identified that sudden deterioration during the course of treatment predicted worse overall outcomes for participants who experienced them versus participants who did not. The authors discussed the fact that their deterioration rate was slightly higher than in other naturalistic studies, such as Lambert and Shimokawa's (2011). They speculated about potential explanations, including that many of their participants may not have received an adequate "dose" of therapy due to treatment dropout or early discontinuation of services.

Additionally, use of the sudden deteriorations construct could result in higher rates than examining deterioration between pre-treatment and post-treatment scores only. It is also worth noting that the treatment of trauma is a complex process, even when provided by a skilled practitioner. For some clients it may result, at least temporarily, in increased awareness of symptoms (e.g. uncomfortable bodily sensations) or the evocation of difficult memories as they are working to process these memories or learn skills to manage symptoms.

Further studies that examined reliable deterioration that occurred at any time during treatment did not report deterioration rates as high as those of Tejas Haugen and colleagues (2015). Rather, they were roughly within the ranges reported in past research such as Hansen, Lambert, and Forman's (2002) large-scale review of OQ-45 outcomes in both clinical trials and routine practice (which returned deterioration rates between 3 and 14% during treatment, with an

average of approximately eight percent). Indeed, the OQ-45 and OQ-45.2 are the most well-represented outcome measures among these studies. Richard, Timulak, and Hevey's (2013) study of email-assisted CBT stands in contrast to the rest of the studies in that it is the only one to report that no deterioration occurred. This may be due to the fact that their participants were administered the BDI-II every eight weeks, whereas participants of the other studies in Table 4 completed outcome measures more frequently and thus potentially provided more opportunities for deterioration to be detected. It is also possible that global outcome measures such as the OQ more readily detect deterioration because they assess more constructs than a unidimensional instrument like the BDI-II.

Another notable study was conducted by Murphy and Cramer (2014). It took place within a university-based mental health service and focused on 62 participants' first three sessions of psychotherapy. The authors identified that 11% of participants with various diagnoses who completed at least three sessions experienced reliable deterioration on the CORE-OM by the end of treatment. The proportion of participants who deteriorated between sessions 1 and 3 was three percent. Murphy and Cramer's findings in particular draw attention to the idea that not all clients deteriorate at a uniform time or rate. Rather, they may follow different trajectories (e.g. deteriorating consistently throughout treatment, deteriorating briefly before experiencing improvement; e.g. Nordberg, Castonguay, Fisher, Boswell, & Kraus, 2014; Owen et al., 2015) or present different markers of deterioration at different stages of treatment (Swift, Callahan, Heath, Herbert, and Levine, 2010).

It follows that researchers and clinicians alike could benefit from tracking the incidence of deterioration among therapy clients on a regular basis, rather than limiting the administration of outcome measures to pre- and post-treatment. This could alert treatment providers to

deterioration early enough that they might employ interventions to counteract it while treatment is still ongoing. Additionally, it would allow for closer study of the deterioration trajectories of particular clients and potentially more accurate prediction of when or how quickly deterioration may occur throughout the treatment process.

Rates of Deterioration in Intention-to-Treat Samples

Additional studies, shown in Table 5, reported deterioration rates for intention-to-treat (ITT) samples (as opposed to samples in which deterioration rates were only given for participants who completed treatment). In some of these cases, the last observation carried forward (LOCF) method was used to account for missing data on outcome measures. It should be considered that a proportion of individuals who do not complete treatment during outcome studies may do so because they are experiencing deterioration. It may be that higher rates of deterioration occur in intention-to-treat samples for this reason. Indeed, rates of deterioration above 10% were more common here than among those studies reporting post-treatment rates. A caveat is that considerably fewer studies reported deterioration rates for ITT samples than studies that reported rates for post-treatment completers, so a thorough comparison cannot readily be made.

Table 5
Rates of Reliable Deterioration in Intention-to-Treat Samples

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Blais et al. (2012)	822 (234 in psychotherapy only, 588 in combined treatment)	SOS-10	Psychotherapy only: 9%, Combined psychotherapy and pharmacotherapy: 12%
Blais et al. (2013)	861 (115 in psychotherapy only, 746 in combined treatment)	SOS-10	Psychotherapy only: 10%, Combined psychotherapy and pharmacotherapy: 12%
Connolly Gibbons et al. (2012)	32 (16 in SE, 16 in TAU)	BASIS-24, HAMD	BASIS-24: 0% in SE, 7% in TAU. HAM-D: 0%, both conditions.
Connolly Gibbons et al. (2015)	85 (39 in feedback condition, 46 in no feedback condition)	BASIS-24	0%, both conditions
Eddington, Silvia, Foxworth, Hoet, & Kwapil, (2015)*	49 (27 in CBT, 22 in SST)	BAI, BDI-II	BAI: 19% in CBT, 14% in SST. BDI: 19% in CBT, 18% in SST.

BAI (Beck Anxiety Inventory), BASIS-24 (Behavior and Symptom Identification Scale), BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), HAMD (Hamilton Depression Rating Scale), SE (Supportive-Expressive dynamic psychotherapy), SOS-10 (Schwartz Outcome Scale), SST (self-system therapy), TAU (treatment as usual)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Measure(s)	Deterioration Rate(s)
Hiller, Schindler, & Lambert (2012)	395	BDI	1% using Percentage method (deterioration = BDI score worsened by ≥ 50%), 3% using RCI
Lopes et al. (2014)	63 (34 in NT, 29 in CBT)	BDI-II, OQ-45.2	OQ-45.2: 3% in NT, 17% in CBT. BDI-II: 3% in NT, 0% in CBT
Schindler, Hiller, & Witthoft, (2011)	338 in naturalistic sample, 227 in subsample who did not meet exclusion criteria commonly used in RCTs	BDI	2%, both groups

BDI/BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), NT (narrative therapy), OQ-45.2 (Outcome Questionnaire), RCI (reliable change index), RCT (randomized controlled trial)

Rates have been rounded to the nearest whole number

Some of the highest rates among these ITT samples came from Eddington, Silvia, Foxworth, Hoet, & Kwapil's (2015) randomized trial comparing CBT and self-system therapy (SST; a structured, integrative treatment) as applied to the treatment of dysthymia and major depression. They found that 19% of those receiving CBT and 18% of those receiving SST reliably deteriorated on the BDI-II. Deterioration rates this high were not seen for the BDI or BDI-II among studies reporting post-treatment rates of deterioration. However, rates for these instruments among other ITT samples were not as high as Eddington and colleagues' (2015) either, and thus differences in sample size may also account for some of the discrepancy.

In a similar study, Lopes and colleagues (2014) conducted a controlled clinical trial comparing narrative therapy (NT) and CBT for the treatment of major depressive disorder. In their ITT sample, they found that three percent of the participants receiving NT reliably deteriorated on the BDI-II and the OQ-45.2. None of the participants receiving CBT reliably deteriorated on the BDI-II, although 17% reliably deteriorated on the OQ-45.2. Thus, those receiving CBT may have improved in terms of their depressive symptoms yet worsened in another area or areas of functioning that a more global outcome measure like the OQ-45.2 could detect.

By way of comparison, Lopes and colleagues also reported deterioration rates for their completer sample, revealing a lower proportion (10%) of individuals receiving CBT whose OQ-45.2 scores reliably deteriorated. Such results suggest that some of the individuals in the CBT condition who were deteriorating did not complete treatment. Indeed, another study collected for this review (Lutz et al., 2014) that examined deterioration both during and post- cognitive-behavioral treatment for panic disorder concluded that individuals who deteriorate within their first five sessions are more likely to discontinue treatment.

Schindler, Hiller, and Witthoft (2011)'s study is notable in that it explored how complex treatment presentations might influence deterioration. In their sample of 338 participants receiving naturalistic cognitive-behavioral therapy for depression, 2.4% deteriorated on the BDI. The authors also calculated the deterioration rate for a subsample of 227 participants who did not meet several exclusion criteria often used for RCTs (i.e. diagnosed with a substance use disorder, psychotic disorder, intellectual disability, organic brain disorder, or dysthymic disorder; presenting with risk of suicide). The rate for this subsample was 1.9%. Thus, the incidence of deterioration remained stable despite participants with potentially more severe presentations being excluded.

The study by Hiller, Schindler, and Lambert (2012) is additionally notable in that it investigated the impact of different methods of calculating deterioration rates. They applied two separate approaches, the RCI and the Percentage method (deterioration defined as at least a 50% worsening of an outcome measure score), to determine what proportion of their 395 participants deteriorated on the BDI while undergoing cognitive-behavioral treatment for depressive disorders. The rate was three percent when the RCI was employed and one percent when the percentage method was employed. The authors concluded that while the rates were similar, the RCI may be a "slightly more rigorous" (p. 6) way of measuring deterioration. Future attempts at replicating Hiller and colleagues' (2012) comparison between methods of calculating deterioration could further identify advantages and disadvantages of the various methods. In turn, this could possibly promote more uniformity in the way deterioration is calculated and reported in outcome studies, thus enhancing the ease of cross-study comparisons.

Given the above examples, it is recommended that future research report rates of deterioration for both intention-to-treat and completer samples. Studies in which completer rates

are the only ones reported run the risk of underrepresenting deterioration rates because those who did not complete treatment are unaccounted for. Being able to make direct comparisons between ITT sample deterioration rates and completer sample deterioration rates within the same study, as well as across multiple studies, would allow for more critical investigation of whether deterioration is relatively more prevalent during early stages of treatment or among those who discontinue treatment prematurely.

It would also be beneficial to focus increased attention on the reasons participants leave treatment prematurely. Examples of this were demonstrated in several studies within Categories B and C of the manual search, which identified that a proportion of their participants experienced worsening symptoms (not necessarily reliable deterioration) and left the study and/or received alternative treatments (e.g. Bastos, Guimaraes, & Trentini, 2015; Gloster et al., 2011; Schlögelhofer et al., 2013; Schmidt et al., 2015; Woody, Whittal, & McClean, 2011). Furthermore, Barnes and colleagues (2013), in a mixed methods study of 38 depressed participants who prematurely discontinued treatment in a RCT of CBT, found that 11% provided "it made me feel worse" as a reason they left treatment and 29% provided the reason "the therapy made me feel confused, annoyed, or upset." Although these responses do not capture reliable deterioration, they lend credence to the notion that a negative reaction to treatment can contribute to early dropout for a portion of clients. However, a counterpoint was found in a study by Saxon, Barkham, Foster, and Parry (2016), who reported no significant correlation between rates of dropout and rates of deterioration within the caseloads of individual therapists. Further efforts to understand the factors contributed to early discontinuation of treatment may lead to the development of techniques to improve treatment retention rates. This would not only benefit psychotherapy outcome researchers in terms of maximizing the amount of data they are able to

collect, it would also benefit clients in terms of maximizing their chances of receiving an adequate dose of treatment and attaining a positive therapeutic change.

Rates of Deterioration at Follow-Up

For five studies (Berger, Boettcher, & Caspar, 2014; Gullestad et al., 2012; Newman et al., 2011; Olthuis, Watt, Mackinnon, & Stewart, 2014; Wells & Colbear, 2012), only rates of deterioration that occurred between post-treatment and a follow-up point were provided. For these studies, deterioration rates fell between 0 and 9%, similar to rates of deterioration occurring during treatment that have been estimated and reported in past research. Additional studies included post-treatment to follow-up rates of deterioration as well as pre-treatment to post-treatment rates. The post-treatment to follow-up rates from these studies were not systematically collected for this review, as the focus was deterioration occurring during the treatment process or directly after treatment ended. Thus, the current review is not able to supply comprehensive information regarding the prevalence of deterioration during extended periods of time after treatment has concluded.

However, considering that some studies reporting deterioration are longitudinal in nature raises questions about whether or not post-treatment to follow-up rates of deterioration are similar to pre-treatment to post-treatment rates. Deterioration that occurs between post-treatment and follow-up may represent a delayed negative impact of treatment, deterioration that could have begun during treatment and persisted beyond treatment because it was never addressed, or deterioration due to the discontinuation of the treatment itself (e.g. a loss of a sense of support). Thus, perhaps current treatments could be improved in terms of adequately preparing clients for termination, helping clients build independent coping skills and sources of social support, or recommending "booster" sessions in the future. It is also possible that when researchers detect a

decline in a client's condition between the end of treatment and a follow-up point, they are capturing a concept akin to relapse or loss of the gains that were made in treatment. Recall that the current review considers relapse as potentially distinct from deterioration in that the former represents a recurrence of distress or symptoms, similar in nature and severity to those experienced at a previous time, after a period of improvement. However, not all researchers may make this distinction when reporting upon participant outcomes at follow-up.

Post-Treatment Deterioration Across Variables

In order to identify potential trends in deterioration across client, therapist, and treatment variables, deterioration rates were examined for several variables: diagnosis/presenting concerns, treatment settings, type of treatment/theoretical orientation, therapist level of education (student trainee or professional degree), and whether or not outcome measures were administered in an ongoing manner throughout treatment. The rates reported here represent deterioration at post-treatment. Studies that presented rates of deterioration which occurred between post-treatment and follow-up periods only, that represented participants who deteriorated at any point during therapy, or that represented ITT samples, were not utilized for this analysis. Additionally, when there was only one study that represented a particular category, it was not included for analysis within that particular variable. For example, if there was only one study in which all participants were diagnosed with an eating disorder, it was not included within the Diagnosis variable.

Tables 6, 8, 10, 12, and 14 provide deterioration rates and general information for each individual study included for analysis, presented separately for each variable. The data presented are drawn from individuals who participated in treatment conditions that met the inclusion criteria defined for this review (e.g. consisted of individual therapy; were not completely self-guided interventions). Each sample size refers to the number of participants within that study

whose pre- and post-treatment outcome data were analyzed by the author(s) in order to provide rates of reliable deterioration. Thus, in some cases, the sample size will represent only a portion of the total sample.

Tables 7, 9, 11, 13, and 15 summarize findings in terms ranges of deterioration rates for particular variables. Some categories within these variables have been created by collapsing across multiple sub-categories. For example, the Anxiety category in Table 7 includes diagnoses of generalized anxiety disorder, social phobia, and panic disorder.

Post-treatment deterioration across diagnoses/presenting problems. In an effort to observe whether deterioration tends to occur more frequently among groups of participants with particular diagnoses, rates of reliable deterioration are presented across categories of diagnoses/presenting problems. These can be seen in Tables 6 and 7. Studies in which participants' diagnoses/presenting concern were not specified or in which all participants did not all share the same primary diagnosis/presenting were excluded, except if the study provided information about rates of deterioration for each diagnosis/presenting concern separately. One study was counted twice because its participants were dually diagnosed with depression and a personality disorder. The reliable deterioration rate for this study was 0%.

Table 6
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Diagnoses/Presenting Problems

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		Grief		
Alves et al. (2014)	6	Complicated grief	BDI-II, ICG	0%, both measures
Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)	5	Grief following loss of a child	PGS, WOC (E-AS)	0%, both measures
Bhattacharya (2015)	3	Perinatal grief	BDI	0%
		Depression/Dysthymia		
Ashworth et al. (2015)	21 (self-help condition excluded)	Comorbid depression and insomnia	BDI-II, ISI	BDI-II: 0%, ISI: 6%
Bambling & King (2013)	92	MDD	BDI	0%
Bernecker et al. (2016)*	57	MDD	BDI	2% using intake as baseline, 4% using session 1 as baseline
Bevan, Wittkowski, & Wells (2013)	6	Post-partum depression	EPDS, HADS	0%, both measures

BDI/BDI-II(Beck Depression Inventory), EPDS (Edinburgh Postnatal Depression Scale), HADS (Hospital Anxiety and Depression Scale), ICG (Inventory of Complicated Grief), ISI (Insomnia Severity Index), MDD (major depressive disorder), PGS (Perinatal Grief Scale), WOC (E-AS; Ways of Coping, Escape-Avoidance Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		Depression/Dysthymia, cont.		
Gonçalves et al. (2012)	6	Major depression	BDI	0%
Gumz, Bauer, & Brähler (2012)	9	Comorbid depression and PD	GSI of SCL-90-R	0%
Kramer et al. (2011)	20	MDD (half of participants also had a comorbid PD)	GSI of SCL-90-R	5%
Lopes et al. (2014)	40 (20 in NT, 20 in CBT)	MDD	BDI-II, OQ-45.2	OQ-45.2: 3% in NT, 10% in CBT. BDI-II: 0% in both groups.
Vromans & Schweitzer (2011)	38	MDD, current depressive episode	BDI-II, OQ-45.2 IR subscale	BDI-II: 5%, OQ-45.2 IR subscale: 3%

BDI/BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), MDD (major depressive disorder), NT (narrative therapy), OQ-45.2/OQ-45.2 IR subscale (Outcome Questionnaire, Interpersonal Relations subscale), PD (personality disorder)

Rates have been rounded to the nearest whole number

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		Anxiety		
Boersma, Håkanson, Salomonsson, & Johansson (2015)	6	Social phobia	SCS, SIAS	0%, both measures
Bosley, Fisher, & Taylor (2016)	7	GAD	Visual analog survey	Negative affect (sad): 14%, Negative affect (irritability): 0%, Negative affect (fear): 0%, Worry: 0%
Farima, Dowlatabadi, Behzadi (2015)	3	GAD	CAQ, IUS, PSWQ, WDQ	0%, all measures
Frets, Kevenaar, & van der Heiden (2013)	6	Social phobia	BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures
Koszycki, Bilodeau, Raab-Mayo, & Bradwejn (2014)*	23	GAD	CGI-S	0%
Lutz et al. (2014)	326	Panic disorder	PDSS-SR	1%

BAI (Beck Anxiety Inventory), BFNE (Brief Fear of Negative Evaluation scale), CAQ (Cognitive Avoidance Questionnaire), CGI-S (Clinical Global Impression-Severity scale), GAD (generalized anxiety disorder), IUS (Intolerance of Uncertainty Scale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PSWQ (Penn State Worry Questionnaire), SCS (Self-Compassion Scale), SIAS (Social Interaction Anxiety Scale), SPAI-N (Social Phobia Anxiety Inventory), WDQ (Worry Domains Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		Anxiety, cont.		
Weiss, Kivity, & Huppert (2014)	19	Panic disorder	ASI-3, PDSS-SR	0%, both measures
		PTSD		
Butollo, König, Karl, Henkel, & Rosner (2014)	21	PTSD	IES-R, PDS	0%, both measures
Feldner, Smith, Monson, & Zvolensky (2013)	4	PTSD and smoking ≥ 20 cigarettes/day	CAPS, PDS, cigarettes smoked per day	0%, all measures
Galovski, Blain, Mott, Elwood, & Houle (2012)	50	PTSD	BDI-II, PDS	0%, both measures
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5	PTSD	CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.
Littleton, Grills, Kline, Schoemann, & Dodd (2016)	18 completed CES-D, 20 FDAS, 23 PSS-I (interactive program only)	PTSD	CES-D, FDAS, PSS-I	CES-D: 11%, FDAS: 20%, PSS-I: 0%

ASI-3 (Anxiety Sensitivity Index), BDI-II (Beck Depression Inventory), CAPS (Clinician-Administered PTSD Scale), CES-D (Center for Epidemiologic Studies Depression Scale), FDAS (Four Dimensional Anxiety Scale), IES-R (Impact of Event Scale-Revised), PDS (Posttraumatic Diagnostic Scale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), PTSD (posttraumatic stress disorder), VK-MFS (Veronen-Kilpatrick Modified Fear Survey)

Rates have been rounded to the nearest whole number

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		PTSD, cont.		
McLay et al. (2011)	10 (VR-GET only)	PTSD	CAPS	0%
		Obsessive-Compulsive and Related Disorders		
Polman, Bouman, van Geert, de Jong, & den Boer (2011)	7	OCD	OBQ, PI-R, Y-BOCS	OBQ: 14%, PI-R: 0%, Y-BOCS: 0%
Rogers et al. (2014)*	41 (received HRT, not specified how many completed post-treatment assessment)	Trichotillomania	MGH-HPS	0%
Taillon, O'Connor, Dupuis, & Lavoie (2013)	10	BDD	Y-BOCS (modified for BDD)	0%
Willson, Veale, & Freeston (2015)	6	BDD	BDI, Y-BOCS (modified for BDD)	0%, both measures

BDD (body dysmorphic disorder), BDI (Beck Depression Inventory) CAPS (Clinician-Administered PTSD Scale), HRT (habit reversal training), MGH-HPS (Massachusetts General Hospital Hairpulling Scale), OBQ (Obsessive Beliefs Questionnaire), OCD (obsessive-compulsive disorder), PI-R (Padua Inventory-Revised), VR-GET (virtual reality-graded exposure therapy), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Diagnosis/Presenting Problem	Outcome Measure(s)	Deterioration Rate(s)
		Obsessive-Compulsive and Related Disorders, cont.		
Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)	39 (patients only)	OCD	COWAT, Digit Span, RAVLT IR/DR, RCFT IR/DR, Stroop Test, TMA, TMB, ToL, WCST-64	RCFT DR: 5%, Stroop: 8%, TMA: 5%, TMB: 8%. 0%, all other measures.
		Personality Disorders.		
Goodman, Edwards, & Chung (2013)	3	BPD	GSI of SCL-90-R	0%
Gumz, Bauer, & Brähler (2012)	9	Comorbid depression and PD	GSI of SCL-90-R	0%
Kellett, Bennett, Ryle, & Thake (2013)	17	BPD	CORE-OM	6%

BPD (borderline personality disorder), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), COWAT (Controlled Oral Word Association Task), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), PD (personality disorder), RAVLT IR/DR (Rey Auditory and Verbal Learning Test, immediate recall/delayed recall), RCFT IR/DR (Rey Complex Figure Test, immediate recall/delayed recall), TMA/TMB (Trail Making Test A/B), ToL (Tower of London), WCST-64 (Wisconsin Card Sorting Test)

Rates have been rounded to the nearest whole number

Table 7
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Diagnoses/Presenting Problems

Client Primary Diagnosis/Presenting Problem	**Number of studies**	**Sample size range (n)**	**Deterioration Rate Range (%)**
Grief/Bereavement	3	3-6	0
Depression/Dysthymia	9	6-92	0-10
Anxiety	7	3-326	0-14
Posttraumatic Stress Disorder	6	4-50	0-20
Obsessive-Compulsive and Related Disorders	5	6- ≤ 41*	0-14
Personality Disorders	3	3-17	0-6

Some diagnoses have been collapsed across categories.
Rates have been rounded to the nearest whole number.
****Not specified how many completed post-treatment assessment***

There were few studies in which all participants were being treated for grief/bereavement. In each of these no participants were deemed as experiencing deterioration at post-treatment. However, this number cannot be considered conclusive due to the small number of studies represented here, and their notably small sample sizes.

In studies examining the treatment of depression and/or dysthymia, deterioration rates were 0-10%, with the vast majority being five percent or below. In studies examining the treatment of anxiety, posttraumatic stress, and obsessive-compulsive and related disorders, the range of deterioration rates was wider. Still, the vast majority of rates clustered around 0% for each category. Results here should be interpreted cautiously, as sample sizes were typically small among studies in each of these categories.

Results for studies examining the treatment of personality disorders should also be approached with caution due to small sample sizes (all below 20). Here, one study produced a deterioration rate of six percent whereas the other two produced rates of zero percent. This is somewhat surprising, given that previous reviews (Castonguay, 2010; Mohr, 1995) have

suggested higher rates of deterioration among individuals with personality disorders, particularly borderline personality disorder. However, the results presented here are not conclusive due to the small number of studies from which these data were drawn, and their relatively small number of participants. Borderline personality disorder in particular is not well represented in this review, despite it being empirically investigated more often than other personality disorders. A likely reason for the underrepresentation here is that studies in which participants had this diagnosis were often screened out because the treatment being studied was DBT and the authors did not present deterioration rates separately for the group therapy and individual therapy components.

Thus, a review of deterioration rates in treatment studies that include group therapy would be a valuable contribution in future research. Overall, more high-n studies that examine the treatment of specific disorders and report deterioration rates are needed, too. This would allow for further clarification of whether particular diagnoses or presenting concerns present greater risk for deterioration.

Post-treatment deterioration across treatment settings. Rates of deterioration were also examined across treatment settings. These can be seen in Tables 8 and 9. Studies that took place in multiple, disparate types of treatment settings or unspecified treatment settings were not included in this investigation, except if the deterioration rates for each type of site were presented separately.

Table 8
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Treatment Settings

Study	**Sample Size**	**Treatment Setting** *University-based community clinic/training clinic*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Briggie, Hilsenroth, Conway, Muran, & Jackson (2016)	243		GSI of BSI	5%
Butollo, König, Karl, Henkel, & Rosner (2014)	21		IES-R, PDS	0%, both measures
Callahan et al. (2014)	216		OQ-45.2	Collected data from six clinics. Rate was 40% for one clinic, which was classified as an outlier and removed from further analysis. Among results for remaining clinics, range was 0-11% and mean was 7%.
Dakwar & Levin (2013)	19		TLFB (measured drug use)	0%

GSI of BSI (Global Severity Index of Brief Symptom Inventory), IES-R (Impact of Event Scale-Revised), OQ-45.2 (Outcome Questionnaire), PDS (Posttraumatic Diagnostic Scale), TLFB (Time Line Follow Back)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Setting *University-based community clinic/training clinic, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Dennhag & Armelius (2012)	187 (full sample)		GSI of SCL-90-R, SASB affiliation dimension	GSI of SCL-90-R: 3% (full sample), SASB affiliation dimension (full sample): 0%
Dennhag, Ybrandt, & Armelius (2011)	235		GSI of SCL-90-R	3%
Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)	98		ORS (part of PCOMS)	6%
Hardy, Tracey, Glidden-Tracey, Hess, & Rohlfing (2011)	210		OQ-45	0%
Hayes, Owen, & Bieschke (2015)	228 (148 White participants, 80 Racial/ Ethnic Minority participants)		OQ-45	5% of White participants, 3% of Racial/Ethnic Minority participants

GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), SASB (Structural Analysis of Social Behavior)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Setting *University-based community clinic/training clinic, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Leibert & Dunne-Bryant (2015)*	81		OQ-45.2	1%
Lopes et al. (2014)	40 (20 in NT, 20 in CBT)		BDI-II, OQ-45.2	OQ-45.2: 3% in NT, 10% in CBT. BDI-II: 0% in both groups.
Rieck & Callahan (2013)	133		OQ-45.2	8%
Vromans & Schweitzer (2011)	38		BDI-II, OQ-45.2 IR subscale	BDI-II: 5%, OQ-45.2 IR subscale: 3%
Weiss, Kivity, & Huppert (2014)	19		ASI-3, PDSS-SR	0%, both measures
Yasky, King, & O'Brien (2015)	11	*University counseling center*	PHQ-15	0%
Boswell, McAleavey, Castonguay, Hayes, & Locke (2012)	864		CCAPS-62 Depression subscale	4%

ASI-3 (Anxiety Sensitivity Index), BDI/BDI-II (Beck Depression Inventory), CCAPS-62 (Counseling Center Assessment of Psychological Symptoms), CBT (cognitive-behavioral therapy), NT (narrative therapy), OQ-45.2/OQ-45.2 IR subscale (Outcome Questionnaire, Interpersonal Relations subscale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PHQ-15 (Patient Health Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Treatment Setting *University counseling center, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Goldberg et al (2016)*	6,591		OQ-45	4%
Kramer, Pascual-Leone, Despland, & de Roten (2015)	32		BDI-II	6%
Murphy, Rashleigh, & Timulak (2012)	110		ORS (part of PCOMS)	5% in feedback condition, 10% in no feedback condition
Murray, McKenzie, Murray, & Richelieu (2015)	305		CORE-OM	2%
		Internet-based		
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5	Therapist support and feedback via written online communication	CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.

BDI-II (Beck Depression Inventory), CES-D (Center for Epidemiologic Studies Depression Scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), FDAS (Four Dimensional Anxiety Scale), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), VK-MFS (Veronen-Kilpatrick Modified Fear Survey)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Treatment Setting	Outcome Measure(s)	Deterioration Rate(s)
		Internet-based, cont.		
Littleton, Grills, Kline, Schoemann, & Dodd (2016)	18 completed CES-D, 20 FDAS, 23 PSS-I (interactive program only)	Therapist support and feedback via written and video-recorded online communication	CES-D, FDAS, PSS-I	CES-D: 11%, FDAS: 20%, PSS-I: 0%
Titov et al. (2015)	2,049	Therapist support and feedback via phone or email	GAD-7, PHQ-9	GAD-7: 3%, PHQ-9: 2%
		Community mental health		
Aubochon-Endsley, Callahan, & Scott (2014)	65		BAI, BDI-II	BAI and/or BDI-II: 8%
Beck, Burdett, & Lewis (2014)	103		CORE-OM GD score	3%
Frets, Kevenaar, & van der Heiden (2013)	6		BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures

BAI (Beck Anxiety Inventory), BDI-II (Beck Depression Inventory), BFNE (Brief Fear of Negative Evaluation scale), CES-D (Center for Epidemiologic Studies Depression Scale), CORE-OM GD score (Clinical Outcomes in Routine Evaluation—Outcome Measure, Global Distress score), FDAS (Four Dimensional Anxiety Scale), OQ-45 (Outcome Questionnaire), GAD-7 (Generalized Anxiety Disorder Scale), PHQ-9 (Patient Health Questionnaire), PSS-I (PTSD Symptom Scale-Interview), SPAI-N (Social Phobia Anxiety Inventory),

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Setting *Hospital (outpatient) or medical center*	Outcome Measure(s)	Deterioration Rate(s)
Bernecker et al. (2016)*	57		BDI	2% using intake as baseline, 4% using session 1 as baseline
Bryan et al. (2012)	495		BHM-20 GMH scale	5%
McLay et al. (2011)	10 (VR-GET only)		CAPS	0%
Simon, Lambert, Harris, Busath, & Vazquez (2012)	207 (98 in TAU, 109 in feedback + CST)		OQ-45	12% in TAU, 6% in feedback + CST
Taillon, O'Connor, Dupuis, & Lavoie (2013)	10	*Inpatient*	Y-BOCS (modified for BDD)	0%
Gumz, Bauer, & Brähler (2012)	9		GSI of SCL-90-R	0%
Goodman, Edwards, & Chung (2013)	3		GSI of SCL-90-R	0%

BDD (Body dysmorphic disorder), BDI (Beck Depression Inventory), BHM-20 GMH scale (Behavioral Health Measure, Global Mental Health scale), CAPS (Clinician-Administered PTSD Scale), CST (Clinical Support Tools), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), OQ-45 (Outcome Questionnaire), TAU (treatment as usual), VR-GET (virtual reality graded exposure therapy), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Table 9
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Treatment Settings

Treatment Setting	Number of studies	Sample size range (n)	Deterioration Rate Range (%)
University-based community clinic/training clinic	15	19-243	0-11
University counseling center	5	32-6,591	2-10
Internet-based	3	5-2,049	0-20
Community mental health	3	6-103	0-8
Hospital (outpatient) or medical center	5	10-495	0-12
Inpatient	2	3-9	0

Some treatment settings have been collapsed across categories
Rates have been rounded to the nearest whole number.

The most well-represented category of treatment setting was university-based community and/or training clinics. Here, deterioration rates were 0-10%, with the majority being five percent or below. This is comparable to Hansen, Lambert, and Forman's (2002) reported 3.2% deterioration rate on the OQ-45 for a university-supported training clinic. Variability in rates existed within some studies, according to site (Callahan et al., 2014) and outcome measure (Lopes et al., 2014). For instance, in Lopes and colleagues' (2014) study, the OQ-45.2 detected differing rates of deterioration between two treatment conditions, while the BDI-II did not.

For the remainder of treatment setting categories, there were relatively few studies representing each category. The majority of studies that took place in university counseling centers reported deterioration rates between two and six percent. This range is lower than Hansen, Lambert and Forman's (2002) 9.7% deterioration rate on the OQ-45 for a university counseling center. It is slightly higher than the average benchmark rate of one percent on the CORE-OM found across multiple higher education counseling centers (CORE IMS, 2010[b]). As Table 8 demonstrates, rates for treatments that took place via an online platform, in a hospital or

medical center, or in a community health center were more variable. This is perhaps explained by diversity in sample sizes and in outcome measures utilized.

The two studies that took place in inpatient settings reported 0% deterioration rates, but had notably small sample sizes. The dearth of studies representing inpatient treatment may be in part attributed to the fact that many inpatient treatment programs are centered on group therapy, studies of which were not included for this review. Additionally, because patients are often consistently monitored face-to-face in inpatient settings, the administration of routine outcome measures may not be as common. Overall, further research focused on deterioration in particular treatment settings may assist in developing setting-specific benchmarks and allow for clearer comparisons across sites.

Post-treatment deterioration across treatment type/theoretical orientation. Tables 10 and 11 present deterioration rates for different types of psychotherapy or theoretical approaches to treatment. Studies in which the type of psychotherapy employed was not specified or in which the study clinicians' work represented a variety of disparate theoretical approaches, were not included. An exception was that if a study reported rates of deterioration for each type of treatment separately, it was included. Some studies were included in multiple categories due to the fact that they involved multiple treatment conditions. The integrative/eclectic treatments presented here typically involved syntheses of 2-3 treatment approaches or theoretical orientations, including cognitive-behavioral, humanistic, interpersonal, gestalt, psychodynamic/psychoanalytic, mindfulness-based, DBT, and ACT. Imagery rescripting is an imagery-based approach originally developed to treat trauma. In the two studies presented here, it was run as a stand-alone treatment and adapted for other presenting concerns, namely body dysmorphic disorder and social phobia. Interpersonal therapy and humanistic/client-centered

therapy were not well-represented, though they were sometimes integrated with other approaches.

Table 10
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Treatment Type/Theoretical Orientation

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral		
Ashworth et al. (2015)	21 (self-help condition excluded)	CBT-I	BDI-II, ISI	BDI-II: 0%, ISI: 6%
Aubochon-Endsley, Callahan, & Scott (2014)	65	"primarily cognitive-behavioral interventions"	BAI, BDI	BAI and/or BDI-II: 8%
Bambling & King (2013)	92	EPST (cognitive-behavioral)	BDI	0%
Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)	5	Cognitive-behavioral	PGS, WOC (E-AS)	0%, both measures
Bhattacharya (2015)	3	Cognitive-behavioral	BDI	0%
Bosley, Fisher, & Taylor (2016)	7	CBT	Visual analog survey	Negative affect (sad): 14%, Negative affect (irritability): 0%, Negative affect (fear): 0%, Worry: 0%

BAI (Beck Anxiety Inventory), BDI/BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), CBT-I (cognitive-behavioral therapy for insomnia), EPST (Extended Problem-Solving Treatment), ISI (Insomnia Severity Index), PGS (Perinatal Grief Scale), WOC (E-AS; Ways of Coping, Escape-Avoidance Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral, cont.		
Branson, Shafran, & Myles (2015)	1,247	CBT	GAD-7, PHQ-9 (RCI for both measures combined)	6% when student therapists were in CBT training, 3% six months after training had concluded
Bryan et al. (2012)	495	"cognitive-behavioral approaches adapted from empirically supported treatments"	BHM-20 GMH scale	5%
Dennhag & Armelius (2012)	88 (CT subsample; 30 received CT1, 58 received CT2)	CT	GSI of SCL-90-R, SASB affiliation dimension	GSI of SCL-90-R: 10% (CT1), 0% (CT2). SASB affiliation dimension: 0% (CT1), 0% (CT2).
Feldner, Smith, Monson, & Zvolensky (2013)	4	Cognitive-behavioral combined with pharmacological treatment	CAPS, PDS, cigarettes smoked per day	0%, all measures

BHM-20 GMH scale (Behavioral Health Measure, Global Mental Health scale), CAPS (Clinician-Administered PTSD Scale), CBT (cognitive-behavioral therapy), CT (cognitive therapy), CT1 (one semester of cognitive therapy), CT2 (2 semesters of cognitive therapy), GAD-7 (Generalized Anxiety Disorder Scale), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), PDS (Posttraumatic Diagnostic Scale), PHQ-9 (Patient Health Questionnaire), RCI (reliable change index), SASB (Structural Analysis of Social Behavior)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral, cont.		
Galovski, Blain, Mott, Elwood, & Houle (2012)	50	MPCT (cognitive)	BDI-II, PDS	0%, both measures
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5	From Survivor to Thriver (CBT; delivered online, therapist-facilitated)	CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.
Littleton, Grills, Kline, Schoemann, & Dodd (2016)	18 completed CES-D, 20 FDAS, 23 PSS-I (interactive program only)	From Survivor to Thriver (CBT; delivered online, therapist-facilitated)	CES-D, FDAS, PSS-I	CES-D: 11%, FDAS: 20%, PSS-I: 0%
Lopes et al. (2014)	20 (CBT condition only)	CBT	BDI-II, OQ-45.2	BDI-II: 0%, OQ-45.2: 10%
Lutz et al. (2014)	326	CBT	PDSS-SR	1%
McLay et al. (2011)	10 (VR-GET only)	VR-GET (behavioral)	CAPS	0%

BDI-II (Beck Depression Inventory), CAPS (Clinician-Administered PTSD Scale), CBT (cognitive-behavioral therapy), CES-D (Center for Epidemiologic Studies Depression Scale), FDAS (Four Dimensional Anxiety Scale), MPCT (modified cognitive processing therapy), OQ-45.2 (Outcome Questionnaire), PDS (Posttraumatic Diagnostic Scale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), VK-MFS (Veronen-Kilpatrick Modified Fear Survey), VR-GET (virtual-reality graded exposure therapy)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral, cont.		
Polman, Bouman, van Geert, de Jong, & den Boer (2011)	7	CT	OBQ, PI-R, Y-BOCS	OBQ: 14%, PI-R: 0%, Y-BOCS: 0%
Rogers et al. (2014)*	41 (received HRT, not specified how many completed post-treatment assessment)	HRT (behavioral)	MGH-HPS	0%
Taillon, O'Connor, Dupuis, & Lavoie (2013)	10	IBT (cognitive)	Y-BOCS (modified for BDD)	0%
Titov et al. (2015)	2,049	MindSpot Clinic (cognitive-behavioral; delivered online, therapist-facilitated)	GAD-7, PHQ-9	GAD-7: 3%, PHQ-9: 2%

BDI-II (Beck Depression Inventory), BDD (body dysmorphic disorder), CT (cognitive therapy), GAD-7 (Generalized Anxiety Disorder Scale), HRT (habit reversal training, IBT (inference-based therapy), MGH-HPS (Massachusetts General Hospital Hairpulling Scale), OBQ (Obsessive Beliefs Questionnaire), PHQ-9 (Patient Health Questionnaire), PI-R (Padua Inventory-Revised), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral, cont.		
Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)	39 (patients only)	CBT	COWAT, Digit Span, RAVLT IR/DR, RCFT IR/DR, Stroop Test, TMA, TMB, ToL, WCST-64	RCFT DR: 5%, Stroop: 8%, TMA: 5%, TMB: 8%. 0%, all other measures.
Vaz, Conceição, & Machado (2013)	30	Therapist-guided self-help CBT	EDE-Q	EDE-Q food concern subscale: 3%, EDE-Q shape concern subscale: 3%
Werbart, Levin, Andersson, & Sandell (2013)	29 completed SRH, 30 GSI, 30 QOLI (CBT condition only)	CBT	GSI of SCL-90, QOLI, SRH	GSI: 0%, QOLI: 0%, SRH: 0%

CBT (cognitive-behavioral therapy), COWAT (Controlled Oral Word Association Task), EDE-Q (Eating Disorder Examination Questionnaire), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), QOLI (Quality of Life Inventory), RAVLT IR/DR (Rey Auditory and Verbal Learning Test, immediate recall/delayed recall), RCFT IR/DR (Rey Complex Figure Test, immediate recall/delayed recall), SRH (Self-Rated Health),TMA/TMB (Trail Making Test A/B), ToL (Tower of London), WCST-64 (Wisconsin Card Sorting Test),

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Cognitive/behavioral/cognitive-behavioral, cont.		
Weiss, Kivity, & Huppert (2014)	19	CBT	ASI-3, PDSS-SR	0%, both measures
		Psychodynamic		
Aafjes-van Doorn , Macdonald, Stein, Cooper, & Tucker (2014)	31	EDT (short-term)	IIP-32, RS, SCQ-SF	IIP-32: 3% RS: 10% SCQ-SF: 10%
Dennhag & Armelius (2012)	99 (PDT subsample; 59 received PDT1, 40 received PDT2)	PDT	GSI of SCL-90-R, SASB affiliation dimension	GSI of SCL-90-R: 2% (PDT1), 5% (PDT2). SASB affiliation dimension: 0% (PDT1), 0% (PDT2).
Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)	98	Psychodynamic	ORS (part of PCOMS)	6%
Goodman, Edwards, & Chung (2013)	3	Psychodynamic	GSI of SCL-90-R	0%

ASI-3 (Anxiety Sensitivity Index), CBT (cognitive-behavioral therapy), EDT (Experiential Dynamic Therapy), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90), IIP-32 (Inventory of Interpersonal Problems-32), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PDT (psychodynamic therapy), PDT1 (one semester of psychodynamic therapy), PDT2 (2 semesters of psychodynamic therapy), RS (Remoralization Scale), SASB (Structural Analysis of Social Behavior), SCQ-SF (Self-Compassion Questionnaire-Short Form),

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Psychodynamic, cont.		
Gumz, Bauer, & Brähler (2012)	9	Psychodynamic	GSI of SCL-90-R	0%
Kramer, de Roten, Drapeau, & Despland (2013)	50	BPI	GSI of SCL-90-R	6%
Kramer, Pascual-Leone, Despland, & de Roten (2015)	32	STDP	BDI-II	6%
Kramer et al. (2011)	20	BPI	GSI of SCL-90-R	5%
Werbart, Levin, Andersson, & Sandell (2013)	116 completed SRH, 115 GSI, 115 QOLI (psychodynamic therapy condition only)	Psychodynamic	GSI of SCL-90, QOLI, SRH	GSI: 4%, QOLI: 2%, SRH: 0%
Yasky, King, & O'Brien (2015)	11	Psychodynamic	PHQ-15	0%

BDI-II (Beck Depression Inventory), BPI (brief psychodynamic intervention), GSI of SCL-90/SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), PHQ-15 (Patient Health Questionnaire), QOLI (Quality of Life Inventory), SRH (Self-Rated Health), STDP (Short-Term Dynamic Psychotherapy)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Integrative/eclectic		
Barr, Hodge, Levan, Bowen, & Knox (2012)	35	"Pluralistic," primarily humanistic	CORE-OM	9%
Bevan, Wittkowski, & Wells (2013)	6	Metacognitive therapy (includes cognitive, behavioral, and mindfulness elements)	EPDS, HADS	0%, both measures
Boersma, Håkanson, Salomonsson, & Johansson (2015)	6	CFT (adapted portions from ACT, CBT, DBT)	SCS, SIAS	0%, both measures
Butollo, König, Karl, Henkel, & Rosner (2014)	21	DET (includes interpersonal, cognitive-behavioral, and gestalt elements)	IES-R, PDS	0%, both measures
Kellett, Bennett, Ryle, & Thake (2013)	17	CAT	CORE-OM	6%
McElvaney & Timulak (2013)	11	Combined CBT and person-centered therapy	CORE-OM	18%

ACT (Acceptance and Commitment Therapy), CAT (cognitive-analytic therapy), CBT (cognitive-behavioral therapy), CFT (compassion focused therapy), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), DBT (Dialectical Behavior Therapy), DET (Dialogical Expsoure in Trauma Therapy), EPDS (Edinburgh Postnatal Depression Scale), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), HADS (Hospital Anxiety and Depression Scale), IES-R (Impact of Event Scale-Revised), PDS (Posttraumatic Diagnostic Scale), SCS (Self-Compassion Scale), SIAS (Social Interaction Anxiety Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Treatment Type/Theoretical Orientation	Outcome Measure(s)	Deterioration Rate(s)
		Integrative/Eclectic, cont.		
Werbart, Levin, Andersson, & Sandell (2013)	31 (integrative/eclectic therapy condition only)	Integrative/eclectic	GSI of SCL-90, QOLI, SRH	GSI: 0%, QOLI: 3%, SRH: 0%
		Narrative		
Lopes et al. (2014)	20 (NT condition only)	Narrative therapy	BDI-II, OQ-45.2	BDI-II: 0%, OQ-45.2: 3%
Vromans & Schweitzer (2011)	38	Narrative therapy	BDI-II, OQ-45.2 IR subscale	BDI-II: 5%, OQ-45.2 IR subscale: 3%
		Imagery rescripting		
Frets, Kevenaar, & van der Heiden (2013)	6	Imagery rescripting	BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures
Willson, Veale, & Freeston (2015)	6	Imagery rescripting	BDI, Y-BOCS (modified for BDD)	0%, both measures

BAI (Beck Anxiety Inventory), BDD (body dysmorphic disorder), BDI/BDI-II (Beck Depression Inventory), BFNE (Brief Fear of Negative Evaluation scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), NT (narrative therapy), OQ-45.2/OQ-45.2 IR subscale (Outcome Questionnaire, Interpersonal Relations subscale), QOLI (Quality of Life Inventory), SPAI-N (Social Phobia Anxiety Inventory), SRH (Self-Rated Health), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

Table 11
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Treatment Type/Theoretical Orientation

Treatment	**Number of studies**	**Sample size range (n)**	**Deterioration Rate Range (%)**
Cognitive/behavioral/cognitive-behavioral	24	3-2,049	0-20
Psychodynamic	10	3-118	0-10
Integrative/eclectic	7	6-35	0-18
Narrative	2	20-38	3-5
Imagery rescripting	2	6	0

Some treatments have been collapsed across categories.
Rates have been rounded to the nearest whole number.

The vast majority of studies employing cognitive, behavioral, or cognitive-behavioral treatment reported deterioration rates between zero and eight percent, though higher rates (including one incidence of 11%, two incidences of 10%, two incidences of 14%, and two incidences of 20%) did occur. Notably, though, none of these higher rates was the only deterioration rate reported within its respective study. For each study that included these higher deterioration rates, there was also reported at least one zero percent deterioration rate for another outcome measure that was administered. Thus, here again, variation in rates occurred within studies according to outcome measure. However, there was not a clear trend for a particular type of outcome measure (e.g. global vs. specific) producing higher rates.

For psychodynamic therapy, rates were fairly evenly spread between 0 and six percent, with one study reporting up to 10% on two of its outcome measures. For integrative/eclectic therapy, four of the studies reported rates between zero and three percent, while the remaining studies reported rates of six percent, nine percent, and 18%, respectively. Sample sizes in this category were notably small, which may contribute to the variability. Small sample sizes were also apparent among the narrative therapy and imagery rescripting categories. In the latter, both

studies reported no incidence of deterioration. In the former, encompassing two studies that reported deterioration rates for two outcome measures each, rates were zero percent and three percent in one study and three percent and five percent in the other.

Overall, these data do not lend clear evidence to support the superiority or inferiority of a particular theoretical approach to treatment in terms of limiting the occurrence of deterioration. Though firmer conclusions might be drawn if more studies containing larger sample sizes, representing particular schools of therapeutic practice and reporting deterioration were available, it may be that just as most bona-fide therapies are comparable in their general effectiveness, their rates of deterioration may be similar as well (APA, 2013). In seeking answers regarding how to select a treatment approach that poses the least risk to a client, it may be more profitable to examine interactions between the treatment approach and other variables, such as the client's demographic factors and presenting concerns or the therapist's areas of competence (Mohr, 1995).

Post-treatment deterioration across therapist levels of education. Tables 12 and 13 broadly separates rates of reliable deterioration by therapist level of education, either graduate student trainee or licensed professional. Studies in which the therapists represented a combination of both professionals and student trainees were not included unless deterioration rates were provided for each category separately.

Table 12
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Therapist Levels of Education

Study	**Sample Size**	**Therapist Level of Education** *Professional degree*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Barr, Hodge, Levan, Bowen, & Knox (2012)	35		CORE-OM	9%
Chow et al. (2015)	4,580 clients, 69 therapists		CORE-10	Therapists grouped into four categories from best - (1) to worst-performing (4). Respective mean rates were: 2%, 2%, 3%, 5%.
Dakwar & Levin (2013)	19		TLFB (measured drug use)	0%
de Jong et al. (2014)	475 (232 in short-term [< 35 weeks] therapies, 243 in long-term [≥ 35weeks] therapies)		OQ-45	Full sample: 8% (NFb), 11% (FbT), 5% (FbTP). Short-term: 10% (NFb), 13% (FbT), 7% (FbTP). Long-term: 6% (NFb), 9% (FbT), 4% (FbTP).

CORE-OM/CORE-10 (Clinical Outcomes in Routine Evaluation—Outcome Measure/10-item version), FbT (feedback to therapist condition), FbTP (feedback to therapist and patient condition), NFb (no feedback condition), OQ-45 (Outcome Questionnaire), TLFB (Time Line Follow Back)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Professional degree, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Frets, Kevenaar, & van der Heiden (2013)	6		BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures
Galovski, Blain, Mott, Elwood, & Houle (2012)	50		BDI-II, PDS	0%, both measures
Goodman, Edwards, & Chung (2013)	3		GSI of SCL-90-R	0%
Gumz, Bauer, & Brähler (2012)	9		GSI of SCL-90-R	0%
Kellett, Bennett, Ryle, & Thake (2013)	17		CORE-OM	6%
Koszycki, Bilodeau, Raab-Mayo, & Bradwejn (2014)*	23		CGI-S	0%
Kramer, de Roten, Drapeau, & Despland (2013)	50		GSI of SCL-90-R	6%

BAI (Beck Anxiety Inventory), BDI-II (Beck Depression Inventory), BFNE (Brief Fear of Negative Evaluation scale), CGI-S (Clinical Global Impression-Severity scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), PDS (Posttraumatic Diagnostic Scale), SPAI-N (Social Phobia Anxiety Inventory)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Therapist Level of Education *Professional degree, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Kramer, Pascual-Leone, Despland, & de Roten (2015)	32		BDI-II	6%
Kramer et al. (2011)	20		GSI of SCL-90-R	5%
Kraus, Castonguay, Boswell, Nordberg, & Hayes (2011)	6,960		TOP	Reported across TOP domains. Sexual functioning: 23%, work functioning: 23%, violence 16%, social functioning: 32%, panic/anxiety: 25%, substance abuse: 19%, psychosis: 23%, quality of life: 24%, sleep: 25%, suicidality: 15%, depression: 20%, mania: 8%.
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5		CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.

BDI-II (Beck Depression Inventory), CES-D (Center for Epidemiologic Studies Depression Scale), FDAS (Four Dimensional Anxiety Scale), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), TOP (Treatment Outcome Package), VK-MFS (Veronen-Kilpatrick Modified Fear Survey)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Professional degree, cont.*	Outcome Measure(s)	Deterioration Rate(s)
McLay et al. (2011)	10 (VR-GET only)		CAPS	0%
Mechler & Holmqvist (2015)	1,157 (840 in primary care, 317 in psychiatric clinic)		CORE-OM	Primary care: 2%, Psychiatric clinic: 7%
Murphy, Rashleigh, & Timulak (2012)	110		ORS (part of PCOMS)	5% in feedback condition, 10% in no feedback condition
Simon, Lambert, Harris, Busath, & Vazquez (2012)	207 (98 in TAU, 109 in feedback + CST)		OQ-45	12% in TAU, 6% in feedback + CST

BDI-II (Beck Depression Inventory), CAPS (Clinician-Administered PTSD Scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), CST (Clinical Support Tools), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), TAU (treatment as usual), VR-GET (virtual reality-graded exposure therapy)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Professional degree, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Taillon, O'Connor, Dupuis, & Lavoie (2013)	10		Y-BOCS (modified for BDD)	0%
Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)	39 (patients only)		COWAT, Digit Span, RAVLT IR/DR, RCFT IR/DR, Stroop Test, TMA, TMB, ToL, WCST-64	RCFT DR: 5%, Stroop: 8%, TMA: 5%, TMB: 8%. 0%, all other measures.
Vaz, Conceição, & Machado (2013)	30		EDE-Q	EDE-Q food concern subscale: 3%, EDE-Q shape concern subscale: 3%
Werbart, Levin, Andersson, & Sandell (2013)	177 completed SRH, 175 GSI, 176 QOLI (full sample)		GSI of SCL-90, QOLI, SRH	GSI: 3%, QOLI: 2%, SRH: 0%

BDD (body dysmorphic disorder), COWAT (Controlled Oral Word Association Task), EDE-Q (Eating Disorder Examination Questionnaire), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), QOLI (Quality of Life Inventory), RAVLT IR/DR (Rey Auditory and Verbal Learning Test, immediate recall/delayed recall), RCFT IR/DR (Rey Complex Figure Test, immediate recall/delayed recall), SRH (Self-Rated Health),TMA/TMB (Trail Making Test A/B), ToL (Tower of London), WCST-64 (Wisconsin Card Sorting Test), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Professional degree, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Yasky, King, & O'Brien (2015)	11	*Graduate student*	PHQ-15	0%
Alves et al. (2014)	6		BDI-II, ICG	0%, both measures
Ashworth et al. (2015)	21 (self-help condition excluded)		BDI-II, ISI	BDI-II: 0%, ISI: 6%
Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)	5		PGS, WOC (E-AS)	0%, both measures
Bevan, Wittkowski, & Wells (2013)	6		EPDS, HADS	0%, both measures
Boersma, Håkanson, Salomonsson, & Johansson (2015)	6		SCS, SIAS	0%, both measures

BDI-II (Beck Depression Inventory), EPDS (Edinburgh Postnatal Depression Scale), HADS (Hospital Anxiety and Depression Scale), ICG (Inventory of Complicated Grief), ISI (Insomnia Severity Index), PGS (Perinatal Grief Scale), PHQ-15 (Patient Health Questionnaire), SCS (Self-Compassion Scale), SIAS (Social Interaction Anxiety Scale), WOC (E-AS; Ways of Coping, Escape-Avoidance Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Graduate student, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Branson, Shafran, & Myles (2015)	1,247		GAD-7, PHQ-9 (RCI for both measures combined)	6% when student therapists were in CBT training, 3% six months after training had concluded
Butollo, König, Karl, Henkel, & Rosner (2014)	21		IES-R, PDS	0%, both measures
Callahan et al. (2014)	216		OQ-45.2	Collected data from six clinics. Rate was 40% for one clinic, which was classified as an outlier and removed from further analysis. Among results for remaining clinics, range was 0-11% and mean was 7%.

CBT (cognitive-behavioral therapy), GAD-7 (Generalized Anxiety Disorder Scale), IES-R (Impact of Event Scale-Revised), OQ-45.2 (Outcome Questionnaire), PDS (Posttraumatic Diagnostic Scale), PHQ-9 (Patient Health Questionnaire), RCI (reliable change index)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Graduate student, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Dennhag & Armelius (2012)	187 (full sample)		GSI of SCL-90-R, SASB affiliation dimension	GSI of SCL-90-R: 3% (full sample), SASB affiliation dimension (full sample): 0%
Feldner, Smith, Monson, & Zvolensky (2013)	4		CAPS, PDS, cigarettes smoked per day	0%, all measures
Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)	98		ORS (part of PCOMS)	6%
Hardy, Tracey, Glidden-Tracey, Hess, & Rohlfing (2011)	210		OQ-45	0%
Hayes, Owen, & Bieschke (2015)	228 (148 White participants, 80 Racial/ Ethnic Minority participants)		OQ-45	5% of White participants, 3% of Racial/Ethnic Minority participants

CAPS (Clinician-Administered PTSD Scale), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDS (Posttraumatic Diagnostic Scale), SASB (Structural Analysis of Social Behavior)

Rates have been rounded to the nearest whole number

Study	Sample Size	Therapist Level of Education *Graduate student, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Leibert & Dunne-Bryant (2015)*	81		OQ-45.2	1%
Polman, Bouman, van Geert, de Jong, & den Boer (2011)	7		OBQ, PI-R, Y-BOCS	OBQ: 14%, PI-R: 0%, Y-BOCS: 0%
Rieck & Callahan (2013)	133		OQ-45.2	8%
Rogers et al. (2014)*	41 (received HRT, not specified how many completed post-treatment assessment)		MGH-HPS	0%
Weiss, Kivity, & Huppert (2014)	19		ASI-3, PDSS-SR	0%, both measures

ASI-3 (Anxiety Sensitivity Index), HRT (habit reversal training), MGH-HPS (Massachusetts General Hospital Hairpulling Scale), OBQ (Obsessive Beliefs Questionnaire), OQ-45.2 (Outcome Questionnaire), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PI-R (Padua Inventory-Revised), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Table 13
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Across Therapist Levels of Education

Therapist Level of Education	Number of studies	Sample size range (n)	Deterioration Rate Range (%)
Professional degree	24	5-6,960	0-32
Graduate student	18	4-1,247	0-14

Rates have been rounded to the nearest whole number.

The vast majority of rates for both categories fell between zero and six percent and tended to cluster around 0%. It may come as a surprise that there is not more of a discrepancy between categories. One might imagine that individuals with professional degrees would see comparatively lower rates of deterioration among their clients because their training and experience afford them enhanced skill in detecting and intervening with deterioration. One study included in this review directly compared rates of deterioration across two different levels of training (Branson, Shafran, & Myles, 2015). In the study, an average of 6% of clients treated by graduate student therapists who were undergoing a training course in CBT experienced reliable deterioration. However, this average rate of deterioration dropped significantly to 3% in the 6 months following the students' completion of their course.

All in all, the current findings regarding deterioration and therapist education should be interpreted conservatively. This is because the domains do not fully account for the variety in breadth and depth (e.g in terms of specialized training received or years of practice) of experience that members of each might possess. These findings suggest that in the future, researchers may do well to continue examining how interactions between therapist level of education/experience and client presenting concerns may differentially impact deterioration rates.

Post-treatment deterioration between studies in which outcomes were and were not routinely tracked. Tables 14 and 15 broadly separates rates of reliable deterioration according to whether their studies administered outcome measures on a routine basis (i.e. multiple times during the course of treatment), or only at pre-treatment and post-treatment. Notably, all of these studies only reported rates of deterioration at post-treatment. However, the two categories are worth examining comparatively because the routine administration of outcome measures are hypothesized to provide clinicians with richer information about how their clients are progressing in treatment, which in turn might lead to more accurate interventions and better outcomes. Studies in which this information was not specified were not included here. One study was counted twice because it provided separate rates of deterioration for a condition in which outcomes were administered regularly and a condition in which they were only administered at pre- and post-treatment.

Of the studies included here, the majority did routinely track outcomes. They varied in how frequently they administered outcome measures. One study utilized time series data in which outcomes were collected from participants daily. Frequency for the administration of outcome measures for the remaining studies within this category ranged from every therapy session (which was often weekly; 26 studies did so) to every 13 weeks.

Table 14
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Between Studies in Which Outcomes Were and Were not Routinely Tracked

Study	**Sample Size**	**Outcome Tracking** *Tracked outcomes throughout treatment*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Alves et al. (2014)	6		BDI-II, ICG	0%, both measures
Ashworth et al. (2015)	21 (self-help condition excluded)		BDI-II, ISI	BDI-II: 0%, ISI: 6%
Aubochon-Endsley, Callahan, & Scott (2014)	65		BAI, BDI-II	BAI and/or BDI-II: 8%
Beck, Burdett, & Lewis (2014)	103		CORE-OM GD score	3%
Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)	5		PGS, WOC (E-AS)	0%, both measures
Bernecker et al. (2016)*	57		BDI	2% using intake as baseline, 4% using session 1 as baseline
Bhattacharya (2015)	3		BDI	0%

BAI (Beck Anxiety Inventory), BDI/BDI-II (Beck Depression Inventory), CORE-OM GD score (Clinical Outcomes in Routine Evaluation—Outcome Measure, Global Distress score), ICG (Inventory of Complicated Grief), ISI (Insomnia Severity Index), PGS (Perinatal Grief Scale), WOC (E-AS; Ways of Coping, Escape-Avoidance Scale)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Tracking *Tracked outcomes throughout treatment, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Bosley, Fisher, & Taylor (2016)	7		Visual analog survey	Negative affect (sad): 14%, Negative affect (irritability): 0%, Negative affect (fear): 0%, Worry: 0%
Branson, Shafran, & Myles (2015)	1,247		GAD-7, PHQ-9 (RCI for both measures combined)	6% when student therapists were in CBT training, 3% six months after training had concluded
Bryan et al. (2012)	495		BHM-20 GMH scale	5%

BHM-20 GMH scale (Behavioral Health Measure, Global Mental Health scale), CBT (cognitive-behavioral therapy), GAD-7 (Generalized Anxiety Disorder Scale), OQ-45.2 (Outcome Questionnaire), PHQ-9 (Patient Health Questionnaire), RCI (reliable change index)

Rates have been rounded to the nearest whole number

Study	**Sample Size**	**Outcome Tracking** *Tracked outcomes throughout treatment, cont.*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Callahan et al. (2014)	216		OQ-45.2	Collected data from six clinics. Rate was 40% for one clinic, which was classified as an outlier and removed from further analysis. Among results for remaining clinics, range was 0-11% and mean was 7%.
Dakwar & Levin (2013)	19		TLFB (measured drug use)	0%
de Jong (2012)**	1,494		OQ-45	9%
de Jong et al. (2014)	475 (232 in short-term [< 35 weeks] therapies, 243 in long-term [≥ 35weeks] therapies)		OQ-45	Full sample: 8% (NFb), 11% (FbT), 5% (FbTP). Short-term: 10% (NFb), 13% (FbT), 7% (FbTP). Long-term: 6% (NFb), 9% (FbT), 4% (FbTP).

FbT (feedback to therapist condition), FbTP (feedback to therapist and patient condition), NFb (no feedback condition), OQ-45/OQ-45.2 (Outcome Questionnaire), TLFB (Time Line Follow Back)

Rates have been rounded to the nearest whole number
**Dissertation

Study	Sample Size	Outcome Tracking *Tracked outcomes throughout treatment, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Farima, Dowlatabadi, Behzadi (2015)	3		CAQ, IUS, PSWQ, WDQ	0%, all measures
Feldner, Smith, Monson, & Zvolensky (2013)	4		CAPS, PDS, cigarettes smoked per day	0%, all measures
Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)	98		ORS (part of PCOMS)	6%
Frets, Kevenaar, & van der Heiden (2013)	6		BAI, BFNE, SPAI-N Social Phobia subscale	0%, all measures
Galovski, Blain, Mott, Elwood, & Houle (2012)	50		BDI-II, PDS	0%, both measures
Goldberg et al (2016)*	6,591		OQ-45	4%

BAI (Beck Anxiety Inventory), BDI-II (Beck Depression Inventory), BFNE (Brief Fear of Negative Evaluation scale), CAPS (Clinician-Administered PTSD Scale), CAQ (Cognitive Avoidance Questionnaire), IUS (Intolerance of Uncertainty Scale), OQ-45 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDS (Posttraumatic Diagnostic Scale), PSWQ (Penn State Worry Questionnaire), SPAI-N (Social Phobia Anxiety Inventory), WDQ (Worry Domains Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Tracking *Tracked outcomes throughout treatment, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Goodman, Edwards, & Chung (2013)	3		GSI of SCL-90-R	0%
Hardy, Tracey, Glidden-Tracey, Hess, & Rohlfing (2011)	210		OQ-45	0%
Hayes, Owen, & Bieschke (2015)	228 (148 White participants, 80 Racial/ Ethnic Minority participants)		OQ-45	5% of White participants, 3% of Racial/Ethnic Minority participants
Kellett, Bennett, Ryle, & Thake (2013)	17		CORE-OM	6%
Leibert & Dunne-Bryant (2015)*	81		OQ-45.2	1%
Lopes et al. (2014)	40 (20 in NT, 20 in CBT)		BDI-II, OQ-45.2	OQ-45.2: 3% in NT, 10% in CBT. BDI-II: 0% in both groups.

BDI-II (Beck Depression Inventory), CBT (cognitive-behavioral therapy), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), GSI of SCL-90-R (Global Severity Index of Symptom Check List 90-Revised), NT (narrative therapy), OQ-45/OQ-45.2 (Outcome Questionnaire)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	**Sample Size**	**Outcome Tracking** *Tracked outcomes throughout treatment, cont.*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Lutz et al. (2014)	326		PDSS-SR	1%
Mechler & Holmqvist (2015)	1,157 (840 in primary care, 317 in psychiatric clinic)		CORE-OM	Primary care: 2%, Psychiatric clinic: 7%
Murphy, Rashleigh, & Timulak (2012)	110		ORS (part of PCOMS)	5% in feedback condition, 10% in no feedback condition
Polman, Bouman, van Geert, de Jong, & den Boer (2011)	7		OBQ, PI-R, Y-BOCS	OBQ: 14%, PI-R: 0%, Y-BOCS: 0%
Rieck & Callahan (2013)	133		OQ-45.2	8%

BDI-II (Beck Depression Inventory), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), OBQ (Obsessive Beliefs Questionnaire), OQ-45.2 (Outcome Questionnaire), ORS (Outcome Rating Scale), PCOMS (Partners for Change Outcome Management System), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PI-R (Padua Inventory-Revised), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Outcome Tracking *Tracked outcomes throughout treatment, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Simon, Lambert, Harris, Busath, & Vazquez (2012)	207 (98 in TAU, 109 in feedback + CST)		OQ-45	12% in TAU, 6% in feedback + CST
Strauss et al. (2013)*	383 (quality assurance intervention group, TK study only)		BSI	6%
Titov et al. (2015)	2,049		GAD-7, PHQ-9	GAD-7: 3%, PHQ-9: 2%
Vromans & Schweitzer (2011)	38		BDI-II, OQ-45.2 IR subscale	BDI-II: 5%, OQ-45.2 IR subscale: 3%
Weiss, Kivity, & Huppert (2014)	19		ASI-3, PDSS-SR	0%, both measures
Yasky, King, & O'Brien (2015)	11		PHQ-15	0%

ASI-3 (Anxiety Sensitivity Index), BDI-II (Beck Depression Inventory), BSI (Brief Symptom Inventory), CST (clinical support tools), GAD-7 (Generalized Anxiety Disorder Scale), OQ-45/OQ-45.2 IR subscale (Outcome Questionnaire, Interpersonal Relations subscale), PDSS-SR (Panic Disorder Severity Scale-Self-Report), PHQ-9/PHQ-15 (Patient Health Questionnaire), TAU (treatment as usual), TK (Techniker Krankenkasse [German health insurance company])

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Tracking *Tracked outcomes pre- and post-treatment only*	Outcome Measure(s)	Deterioration Rate(s)
Aafjes-van Doorn , Macdonald, Stein, Cooper, & Tucker (2014)	31		IIP-32, RS, SCQ-SF	IIP-32: 3% RS: 10% SCQ-SF: 10%
Bambling & King (2013)	92		BDI	0%
Bevan, Wittkowski, & Wells (2013)	6		EPDS, HADS	0%, both measures
Boersma, Håkanson, Salomonsson, & Johansson (2015)	6		SCS, SIAS	0%, both measures
Briggie, Hilsenroth, Conway, Muran, & Jackson (2016)	243		GSI of BSI	5%
Butollo, König, Karl, Henkel, & Rosner (2014)	21		IES-R, PDS	0%, both measures
Gonçalves et al. (2012)	6		BDI	0%

BDI (Beck Depression Inventory), EPDS (Edinburgh Postnatal Depression Scale), GSI of BSI (Global Severity Index of Brief Symptom Inventory), HADS (Hospital Anxiety and Depression Scale), IES-R (Impact of Event Scale-Revised), IIP-32 (Inventory of Interpersonal Problems-32), PDS (Posttraumatic Diagnostic Scale), RS (Remoralization Scale), SCS (Self-Compassion Scale), SCQ-SF (Self-Compassion Questionnaire-Short Form), SIAS (Social Interaction Anxiety Scale)

Rates have been rounded to the nearest whole number

Study	Sample Size	Outcome Tracking *Tracked outcomes pre- and post-treatment only, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Littleton, Buck, Rosman, & Grills-Taquechel (2012)	5		CES-D, FDAS, PSS-I, PTCI, VK-MFS	CES-D: 20%. 0%, all other measures.
Littleton, Grills, Kline, Schoemann, & Dodd (2016)	18 completed CES-D, 20 FDAS, 23 PSS-I (interactive program only)		CES-D, FDAS, PSS-I	CES-D: 11%, FDAS: 20%, PSS-I: 0%
McLay et al. (2011)	10 (VR-GET only)		CAPS	0%
Payne, Ciclitira, Starr, Marzano, & Brunswick (2015)	98		CORE-OM	2%
Rogers et al. (2014)*	41 (received HRT, not specified how many completed post-treatment assessment)		MGH-HPS	0%

CAPS (Clinician-Administered PTSD Scale), CES-D (Center for Epidemiologic Studies Depression Scale), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), FDAS (Four Dimensional Anxiety Scale), HRT (habit reversal training), MGH-HPS (Massachusetts General Hospital Hairpulling Scale), PSS-I (PTSD Symptom Scale-Interview), PTCI (Posttraumatic Cognitions Inventory), VK-MFS (Veronen-Kilpatrick Modified Fear Survey), VR-GET (virtual reality-graded exposure therapy)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	Sample Size	Outcome Tracking *Tracked outcomes pre- and post-treatment only, cont.*	Outcome Measure(s)	Deterioration Rate(s)
Saxon, Barkham, Foster, & Parry (2016)	6,405 participants, 85 therapists		CORE-OM	0-7% range across therapists. Therapists grouped into categories: above average, average, and below average. Respective mean rates were: 0%, 1%, 3%.
Strauss et al. (2013)*	214 (control group, TK study only)		BSI	5%
Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)	39 (patients only)		COWAT, Digit Span, RAVLT IR/DR, RCFT IR/DR, Stroop Test, TMA, TMB, ToL, WCST-64	RCFT DR: 5%, Stroop: 8%, TMA: 5%, TMB: 8%. 0%, all other measures.

BSI (Brief Symptom Inventory), CORE-OM (Clinical Outcomes in Routine Evaluation—Outcome Measure), COWAT (Controlled Oral Word Association Task), RAVLT IR/DR (Rey Auditory and Verbal Learning Test, immediate recall/delayed recall), RCFT IR/DR (Rey Complex Figure Test, immediate recall/delayed recall), TK (Techniker Krankenkasse [German health insurance company]), TMA/TMB (Trail Making Test A/B), ToL (Tower of London), WCST-64 (Wisconsin Card Sorting Test)

Rates have been rounded to the nearest whole number
*Deterioration rates not published in the original article; obtained through email communication with the author(s)

Study	**Sample Size**	**Outcome Tracking** *Tracked outcomes pre- and post-treatment only, cont.*	**Outcome Measure(s)**	**Deterioration Rate(s)**
Vaz, Conceição, & Machado (2013)	30		EDE-Q	. EDE-Q food concern subscale: 3%, EDE-Q shape concern subscale: 3%
Werbart, Levin, Andersson, & Sandell (2013)	177 completed SRH, 175 GSI, 176 QOLI (full sample)		GSI of SCL-90, QOLI, SRH	GSI: 3%, QOLI: 2%, SRH: 0%
Willson, Veale, & Freeston (2015)	6		BDI, Y-BOCS (modified for BDD)	0%, both measures

BDD (body dysmorphic disorder), BDI (Beck Depression Inventory), EDE-Q (Eating Disorder Examination Questionnaire), GSI of SCL-90 (Global Severity Index of Symptom Check List 90), QOLI (Quality of Life Inventory), SRH (Self-Rated Health), Y-BOCS (Yale-Brown Obsessive Compulsive Scale)

Rates have been rounded to the nearest whole number

Table 15
Rates of Reliable Deterioration That Occurred Post-Treatment, Reported Between Studies in Which Outcomes Were and Were not Routinely Tracked

Outcome Tracking	Number of studies	Sample size range (n)	Deterioration Rate Range (%)
Tracked outcomes throughout treatment	37	3-6,591	0-14
Tracked outcomes pre- and post-treatment only	18	5-6,405	0-20

Rates have been rounded to the nearest whole number.

For both categories, the majority of studies that tracked outcomes throughout treatment reported rates between zero and six percent. As can be seen in Table 14, the studies that contain deterioration rates that represent the upper range (i.e. 14% and 20%) for their respective categories could be considered outliers. However, caution is recommended in comparing these two categories, as there were not as many studies that tracked outcomes at pre- and post-treatment only represented relative to those that tracked outcomes throughout treatment. Another limitation in attempting to compare these two categories is that when outcome measures are administered routinely, there is often not a way of confirming that the treating clinicians carefully viewed or utilized this information such that it changed the way they intervened with clients. Despite an apparent similarity in deterioration rates between these two categories, there is evidence from past research to suggest that when clinicians receive formal feedback along with the results of their routinely-administered outcome measures, this has salutary effects in terms of reducing deterioration. This topic will be addressed in more depth in the General Findings Regarding Deterioration and Other Negative Effects section.

Summary of deterioration rates across categories. All in all, there were not clear distinctions in the magnitude of deterioration rates between domains in the categories selected. This suggests a need for more studies that represent particular categories (e.g. specific diagnoses)

and in which researchers include deterioration rates. It also suggests the need for continued investigation of interaction effects between variables, or mechanisms by which deterioration might occur. One example is found in a study by Ellison, Levy, Cain, Ansell, and Pincus (2013). These authors did not report deterioration rates in their paper. However, they observed that their participants who scored high on a measure of narcissistic grandiosity and low on a measure of narcissistic vulnerability experienced worsening of sleep-related symptoms during the early stages of psychotherapy. Continued discussion will center on additional variables that may play a role in the incidence of deterioration.

General Findings Regarding Deterioration and Other Negative Effects

Additional variables that may contribute to deterioration were also examined, although the studies that represent them were fewer in number compared to those that were referenced in Tables 6, 8, 10, 12, and 14 and discussed in the previous section. Not all studies reviewed in the current section included rates of reliable deterioration. However, those that did not include such rates still reported general results pertaining to deterioration, or reported on other negative effects that occurred during treatment.

Therapist Effects and Effectiveness

Among the studies included in this review, several examined deterioration through the lens of therapist effectiveness, investigating whether the clients of particular therapists may exhibit higher rates of deterioration than others. This topic has previously been researched, for example by Okiishi and colleagues (2006), who studied OQ-45 outcomes of 6,499 clients seeking services at a university counseling center. These clients were treated by 71 therapists, whom the authors rank-ordered in terms of their effectiveness. Those therapists who fell in the bottom ten percent saw an average of 11% of their clients deteriorate, whereas those in the top

ten percent saw 5% of their clients deteriorate. The authors observed that significantly fewer of the top-ranked therapists' clients deteriorated relative to the clients of bottom-ranked therapists. Additionally, this 5-11% range is similar to the oft-cited average proportion of clients (approximately 5-10%) who deteriorate during psychotherapy outcome studies (Lambert, 2013[a]; Lambert & Ogles, 2004). Thus, therapist variables appear to influence client deterioration, although Okiishi and colleagues (2006) found no evidence that the variables of therapist sex, level and type of training, or theoretical orientation had a differential impact.

Studies included in the current review may shed additional light on the role that individual differences among therapists play in client deterioration. Similar to Okiishi and colleagues (2006), Branson and colleagues (2015) found evidence that clinical competence (which is not necessarily the same as level of training) may be a major therapist variable that influences the incidence of deterioration, at least among graduate trainees practicing CBT. Their results indicated that clients (n=1,247) were more likely to deteriorate on the Generalized Anxiety Disorder Scale (GAD-7) and Patient Health Questionnaire (PHQ-9) if they were being treated by a therapist whose scores on the Cognitive Therapy Scale-Revised (an instrument intended to measure therapist competence) fell within the bottom ten percent of scores of the 43 graduate student acting as therapists in the study. Banham and Schweitzer (2016) have published similar findings. Their sample consisted of 58 therapists-in-training and 611 participants who completed the OQ-45.2 at every session. Using pre-post OQ-45.2 change scores, the authors rank-ordered the therapists according to the average degree of therapeutic progress their clients achieved. While they did not report what proportion of clients deteriorated, they did report that the least effective therapist saw his or her clients deteriorate by an average of .30 points on the OQ-45.2 at each session. Thus, Branson and colleagues (2015) and Banham and Schweitzer

(2016) lend credence to the idea that negative client outcomes are the norm for a portion of trainee therapists.

Saxon and Barkham (2012), however, reported less striking differences in deterioration rates across therapists. They found a two percent rate of deterioration among therapists they classified as "below average" in terms of their clients' recovery rates, while the "average" and "above average" therapists generally had a one percent rate of deterioration among their clients. Therapists were assigned to these categories based upon residuals generated through multilevel modeling. The study sample comprised 10,786 clients and 119 therapists. A later study by Saxon and colleagues (2016), examining 6,405 participants and 85 therapists, resulted in comparable findings. These included a 0-7% range of reliable deterioration across therapists, with "average" therapists producing a mean deterioration rate of approximately one percent among their clients, "above average" therapists (of which there was only one in this study) producing a mean of zero percent, and "below average" therapists (of which there were only four in this study) producing a mean of approximately three percent. Therapists were placed into categories using a similar method to that of Saxon and Barkham (2012). Additionally, Chow and colleagues (2015) reported similar results, with "best performing" therapists averaging a two percent rate of reliable deterioration and "worst performing" therapists averaging a five percent rate. These numbers were generated from a sample of 69 therapists representing 45 different organizations and treating 4,580 clients.

The results of the above studies are not as dramatic as those of Okiishi and colleagues (2006) and do not capture as wide a range as some of the rates of deterioration reported earlier in this Results and Discussion section. A notable point, however, is that Saxon and Barkham, Saxon and colleagues, and Chow and colleagues all utilized the CORE-OM, an instrument whose

average deterioration rates are generally consistent with the rates reported by these authors. The CORE-OM's deterioration rates also tend to be lower than the average deterioration rates for the OQ-45 (the outcome measure used by Okiishi and colleagues; see, e.g., Cahill, Barkham, & Stiles, 2010; CORE IMS, 2010[a,b]; Hansen, Lambert, & Forman, 2002; Shimokawa, Lambert, & Smart, 2010). Therefore, it is important to consider that variability in deterioration rates may stem not only from the therapist, but also from the nature of the outcome measure that is being used. Relatedly, questions are raised about whether the difference between the one percent (for below average therapists) and three percent (for above average therapists) rates of deterioration on the CORE-OM found by Saxon and colleagues (2016) may be comparable to the difference between the five percent and 11% deterioration on the OQ-45 found by Okiishi and colleagues (2006).

Just as past research has identified potentially harmful treatments, future research should continue to attend to how to classify and detect potentially harmful therapists. Kraus and colleagues (2011) have already broken ground in this area, operationalizing a "harmful therapist" as one whose clients, on average, experience reliable deterioration. Out of a sample of 3,222 therapists, the authors also reported the percentage that could be classified as harmful across multiple outcome domains on the TOP. These included sexual functioning (12% of therapists), work functioning (7%), violence (16%), social functioning (14%), panic/anxiety (10%), substance abuse (16%), psychosis (9%), quality of life (5%), sleep (9%), suicidality (7%), depression (3%), and mania (0.3%). The authors also reported that there were not strong correlations in therapist effectiveness across domains. Therefore, just as clients may deteriorate in one domain while improving or experiencing no change in another, the results of Kraus and colleagues' study suggest that therapists who bring harm to their clients do not always do so in a

global sense. Rather, some therapists seem to be effective in intervening with specific concerns and not others. This may be accounted for by the presence or absence of training in specialized areas, such as assessing and treating sexual concerns, substance abuse, and violent behavior.

The topic of domains in which therapists may tend to harm clients is salient for training programs. For example, training clinic supervisors should assess new trainees' areas of strength and weakness in order to determine which clients are appropriate for these therapists-in-training to begin treating, as well as to plan what further training is needed to remediate a trainee's lack of skill. Next, supervisors may attend to screening clients who are assigned to student therapists in order to ensure that the client's presenting concerns are ones that the student is competent to treat. Training programs may serve as a first line of defense against deterioration, at least in terms of ameliorating potential therapist effects that contribute, raising student awareness of the occurrence of deterioration, and encouraging use of steps to prevent or intervene, such as regular use of outcome measures and/or feedback systems (Banham & Schweitzer, 2015; Castonguay et al., 2010).

Saxon and colleagues (2016) found that therapist effects accounted for 10.1% of the variance in deterioration in their sample (n=6,405). Thus, therapist-specific factors do seem to contribute to deterioration. Still, this leaves 90% of the remaining variance to be accounted for, pointing to the need for future research to address additional factors that may increase/inflate or decrease/deflate deterioration rates. These may include client (e.g. demographics), methodological (e.g. outcome measures used), treatment (e.g. frequency of sessions), and environmental (e.g. client life stressors) variables.

Client Socio-Cultural Factors

One seemingly under-studied topic is whether deterioration rates differ depending upon client race or ethnicity. Relevant to this question, a 2014 dissertation by DeGeorge examined outcomes for depression symptoms and panic symptoms in a study of naturalistic psychotherapy. The author reported rates of participants whose symptoms worsened to any degree (as opposed to strictly reliable deterioration). Results indicated that seven percent of the 1,187 Caucasian participants experienced worsening depression during treatment whereas none of the 294 Hispanic participants and none of the 91 African American participants experienced such worsening. However, 22% of Hispanic participants experienced worsening panic symptoms, while 13% in both the African American and Caucasian samples experienced this.

In another naturalistic study, one that did report rates of reliable deterioration, Hayes, Owen, and Bieschke (2015) found no statistically significant difference in deterioration rates on the OQ-45 between White participants (n=148) and Racial/Ethnic Minority participants (n=80). Similarly, Saxon and colleagues (2016) found that ethnicity was not a significant predictor of deterioration of at least one point (with a change of five or more points constituting reliable change) on the CORE-OM within their sample of 6,405. Conversely, in an earlier study that was not found during the data collection phase of this review but is discussed here for comparison, Moos, Moos, and Finney (2001) concluded that among 2,616 individuals seeking treatment for substance use disorders, African American participants were at higher risk of experiencing deterioration during treatment.

These three studies provide mixed results regarding the interplay between client racial and ethnic identity and deterioration. However, they do not provide an exhaustive representation of the literature addressing race/ethnicity and therapeutic outcome and thus caution should be

exercised when interpreting them. Further investigation is needed to clarify the relationship between client racial/ethnic identity and the incidence of deterioration. If there is indeed a significant connection, such research could facilitate prevention and intervention efforts being more readily implemented for clients who present demographics that may be risk factors for deterioration. As an example, Wendt, Gone, and Nagata (2015) have previously called for research to afford greater attention to the intersection of various multicultural issues and the potential for harm to psychotherapy clients. One step in the effort to achieve this goal would be attempts at establishing greater representation of racial and ethnic minorities as participants in future studies. Additionally, if a connection between racial/ethnic identity and deterioration is more thoroughly established, examining mechanisms (e.g. mismatch in client and therapist preferences and expectations for therapy due to cultural differences, racially/ethnically-based stressors in a client's life) by which this relationship occurs is a recommended area of inquiry.

Additional sources of cultural diversity that warrant more research attention in terms of how they may impact deterioration include gender, socio-economic status, education, and age. Information about client socio-economic status was not systematically collected for this review. Nevertheless, its potential to elevate risk for client deterioration merits more study, as clients with limited socio-economic resources may have difficulty accessing quality care and attending sessions regularly. Additionally, they may experience significant life stressors that can exacerbate symptoms and interfere with progress in treatment. An example of this is found in Saxon and colleagues (2016), who identified unemployment to be a significant predictor for deterioration of at least one point on the CORE-OM. This finding highlights a particular population that could benefit from targeted interventions to address vocational distress, financial concerns, and barriers to attending treatment. A recommended area for further research is the

effectiveness of treatments that are alternatives or adjuncts to traditional therapy (e.g. teletherapy, guided self-help, case management) for clients whose basic needs (e.g. food, shelter) are not being met or who find it challenging to attend therapy regularly due to constraints related to finances, transportation, childcare, and so on.

Information on client levels of education was also not systematically collected for this review. However, given Ebert and colleagues' (2016) and Rozental and colleagues' (2017) recent findings suggesting higher risk of reliable deterioration for clients who have lower levels of education and who are receiving Internet-based treatments, this is a worthwhile area of further study. In the future, researchers may wish to investigate whether a similar effect is found among clients receiving other types of therapies (e.g. those that occur face-to-face with, or entail more structure and input from, the therapist) or whether alternate modalities of treatment delivery more effectively address the discrepancy in outcomes between clients who attended higher education and those who did not. Additionally, exploration of possible risk factors external to the treatment itself that may exist concurrently with lower education (e.g. lower socio-economic status, difficulties with advanced reading comprehension) is recommended (Ebert et al., 2016).

Because this review focused on adult participants and the majority of the studies analyzed included participants with a wide range of ages, minimal conclusions can be drawn in the current discussion about how age may influence the incidence of deterioration in psychotherapy. However, Saxon and colleagues (2016), whose sample consisted of individuals aged 18 and older, identified older client age as a significant predictor of deterioration of at least one point on the CORE-OM. This stands in contrast to the findings of Rozental and colleagues (2017), who identified being older as associated with reduced likelihood of deterioration, at least among those receiving Internet-based cognitive-behavioral therapies. As the life expectancy of

adults rises over time, the number of older adults seeking treatment may increase. Thus, it will be important to expand upon Saxon and colleagues' (2016) and Rozental and colleagues' (2017) work in order to better understand any unique protective or risk factors that may exist among this population, and in turn more effectively meet their treatment needs.

Conclusions that can be drawn about the impact of gender on client deterioration are similarly limited in the context of the current review because none of the studies explicitly separated rates of deterioration based upon gender. Eleven out of 88 studies reporting rates of reliable deterioration did contain samples in which 100% of the participants identified as women. Deterioration rates in these studies ranged between 0 and 20%, though in the majority of studies rates fell between zero and two percent and sample sizes consisted of 10 participants or less. No studies reporting rates of reliable deterioration that were examined in this review contained samples in which all participants identified as men. Thus, continued investigation on the potential impact of gender on deterioration, including more attention to issues that may be unique to transgender individuals, is suggested.

Treatment Dose and Duration

Previous research has suggested that the duration of a client's course of treatment or their treatment "dose" can also impact therapy outcomes. Several studies included in the current review addressed this topic as it applies to deterioration. For instance, Erekson, Lambert, and Eggett (2015) concluded that participants who attended therapy sessions biweekly were 1.4 times more likely to experienced reliable deterioration on the OQ-45 than those who attended weekly sessions; the rate of deterioration for weekly attenders was six percent whereas the rate for biweekly attenders was nine percent. Their sample consisted of 6,184 participants (3,092 participants from the weekly group and 3,092 from the biweekly group) who were matched

based upon age, gender, and initial symptom severity. Their findings offer a point to which clinicians as well as researchers should attend. Challenges faced by organizations that provide psychotherapy may include a high demand for services, being understaffed, and serving clients who experience barriers (e.g. financial constraints, unreliable transportation, lack of childcare, busy schedules) to attending weekly sessions. One consequence of agencies trying to adapt to these challenges may be that clients are scheduled to attend sessions less often than once a week. Given this, future research would do well to investigate what client and treatment factors may make scheduling therapy appointments less often than weekly contraindicated due to increased risk of deterioration.

An additional study (Dennhag and Armelius, 2012) compared the outcomes of 89 participants who attended one semester of psychotherapy (either cognitive therapy or psychodynamic therapy) at a university-based training clinic with the outcomes of 98 participants who attended for two semesters. Using the Global Severity Index (GSI) of the Symptom Checklist-90 (SCL-90), these authors found that those who engaged in one semester of cognitive therapy had the highest rates of reliable deterioration (10%), followed by those who engaged in two semesters of psychodynamic therapy (5%), those who engaged in one semester of psychodynamic therapy (2%), and those who engaged in 2 semesters of cognitive therapy (0%). The average rate of deterioration for the whole sample was three percent. This study raises questions about how dose or length of therapy and theoretical approach to treatment may interact to affect deterioration. Perhaps diverse theoretical approaches differ in the expectations they set for the clients at the beginning of treatment (e.g. that clients should expect to begin observing changes in their condition quickly versus that clients should expect to spend an extended amount of time in therapy before they will notice significant change), the type of clients they attract, or

how quickly clients adapt to the treatment model or theory-specific techniques (e.g. transference interpretations, being assigned out-of-treatment homework).

Somewhat conversely, de Jong and colleagues (2014) did not find evidence that there was a significant difference in deterioration rates on the OQ-45 between 232 clients who engaged in "short-term" (defined as less than 35 weeks in length) therapies and 243 clients who engaged in "long-term" (defined as at least 35 weeks in length) therapies. However, it should be pointed out that what constitutes short-term or long-term treatment is relative. For instance, it is possible that a treatment fitting within de Jong and colleagues' (2014) definition of short-term could exceed the length of the most extensive treatment (i.e. two semesters) offered to Dennhag and Armelius's (2012) participants.

Formal Feedback on Client Progress

The existing research on deterioration suggests that it is a multi-faceted construct and that clinicians are often not successful at predicting or documenting its occurrence (Hannan et al., 2005; Hatfield et al., 2010). Routine administration of outcome measures, including those that provide explicit feedback on client progress, may provide a guide for clinicians beyond what their clients spontaneously verbalize in session and what clinicians assess face-to-face at each session. Past research has produced evidence suggesting that when clinicians receive feedback about their clients' progress, this is associated with reduced proportions of clients deteriorating post-treatment compared to clients whose clinicians do not receive such feedback (e.g. Lambert, 2013[a]; Lambert et al., 2001; Shimokawa et al., 2010). This effect appears robust, although it has not been replicated in all studies (Lambert, 2007).

Several studies included in the current review addressed the relationship between feedback and deterioration, with mixed conclusions. For example, Amble, Gude, Stubdal, Andersen, and Wampold (2015) studied the effect of feedback about client progress provided via

the OQ-Analyst software (Lambert, 2012) on a naturalistic sample of 259 participants across multiple treatment settings. They found a significant main effect such that feedback increased the likelihood of improvement and decreased the likelihood of both deterioration and no reliable change) on the OQ-45.2. The rate of deterioration for the feedback condition was six percent and the rate for the no-feedback condition was nine percent.

Also using the OQ-Analyst, Simon, Lambert, Harris, Busath, and Vazquez (2012) performed a naturalistic study of 207 individuals attending therapy at a hospital-based outpatient clinic. Participants were randomly assigned to one of two conditions. In one condition, treating clinicians would receive feedback and use of Clinical Support Tools (CSTs) provided via the OQ-Analyst software (Lambert, 2012; Whipple et al., 2003). They were also instructed to share OQ feedback with their clients at each session. In the other condition, clinicians would not have access to feedback or CSTs. All participants completed the OQ-45. The percentage of participants who deteriorated in the no-feedback condition (12%) was double that of participants in the feedback condition (6%).

A similar spread in reliable deterioration rates was reported by Murphy, Rashleigh, and Timulak (2012). Their RCT studied 110 individuals attending therapy in a university counseling center. Reliable deterioration rates, using the Outcome Rating Scale of the Partners for Change Outcome Management System (PCOMS), were 5% for participants whose clinicians received feedback and 10% for participants whose clinicians did not. However, these authors clarified that there was not a not a significant difference in mean post-treatment rates of deterioration between participants in the feedback versus no-feedback groups. In a similar vein to Murphy and colleagues, de Jong and colleagues (2014) found that in a randomized controlled trial of 475 individuals attending outpatient therapy, feedback did not make a significant difference in the

proportion of clients who deteriorated on the OQ-45 across three different conditions: feedback to therapist (11% deterioration), feedback to therapist and patient (5% deterioration), and no feedback (8% deterioration).

These mixed results suggest that continued research may be needed in order to clarify under what conditions the receipt of feedback is associated with significantly reduced rates of deterioration. An important point to note is that studies utilizing feedback cannot always ensure that their therapists are actually viewing the feedback they receive or tailoring their approach to treatment based upon that feedback. While de Jong and colleagues (2014) did not find evidence suggesting that the provision of feedback to both therapists and clients is superior in effectiveness to other methods, the question of how sharing results of feedback with a client might influence deterioration warrants further exploration. Making a joint review of feedback between clinician and client a regular practice may hold a clinician more accountable to acknowledging deterioration when it occurs and providing targeted interventions to address it. It may also foster a sense of trust, transparency, and collaboration between the two parties. The presence of these qualities would likely strengthen the working therapeutic alliance, which has been associated with improvement of symptoms in past research (e.g. DeRubeis, Brotman, & Gibbons, 2005; Falkenström, Ekeblad, & Holmquvist, 2016).

McAleavey, Nordberg, Kraus, & Castonguay (2012), for instance, have recommended discussion with clients regarding results on outcome measures, particularly those that require the client to provide the ratings. Their rationale included the fact that items on outcome measures may be interpreted differently by different clients. Additionally, clients may make accidental errors as they are completing the measure. The authors also posited that, beyond reaching a mutual clarification of outcome data, a discussion of outcome measure results between client and

clinician can strengthen the therapeutic alliance. It is notable that, in order to make such a discussion possible and optimally useful, outcome measures need to be administered more often than pre- and post-treatment.

Studies Comparing Psychopharmacological and Psychotherapeutic Treatments

Several studies compared outcomes for individual psychotherapy, pharmacotherapy, and combined treatment. For instance, Blais and colleagues (2012) conducted a naturalistic study of 1,209 individuals with various diagnoses who were receiving treatment as usual (TAU) at an academic medical center. Using the SOS-10 as an outcome measure, they identified that nine percent of those receiving psychotherapy alone experienced reliable deterioration, whereas the rate was 12% for those receiving a combination of psychotropic medication and psychotherapy. Utilizing a similar design and setting, Blais and colleagues (2013) examined outcomes for 1,322 outpatients undergoing treatment for depression. As in the prior study, the proportion of participants who deteriorated on the SOS-10 was ten percent for those receiving TAU psychotherapy only, 12% for a combination of TAU psychotherapy and psychotropic medication, and 13% for psychotropic medication alone.

In another study examining the treatment of depression, Vittengl and colleagues (2016) conducted a meta-analysis of 16 RCTs that compared the effectiveness of pharmacotherapy and cognitive-behavioral psychotherapy. Their results indicated that for either treatment, approximately one percent of participants experienced reliable deterioration of depressive symptoms. Rates were similar for both the Hamilton Depression Rating Scale (HAM-D) and the BDI.

Though the proportion of participants who deteriorated in these studies varied notably depending upon whether a global outcome scale or a symptom-specific outcome scale was used, rates of deterioration in each study were similar between pharmacological treatment,

psychotherapy, and combined treatment. Such a finding points to the idea that differential rates of deterioration may occur due to factors beyond treatment modality. These factors could include methodological effects (e.g. outcome measures), therapist effects (e.g. competence), client effects (e.g. motivation for treatment), the therapeutic alliance, and stressors in the client's life external to treatment.

Initial Symptom Severity

Because reliable deterioration is often defined as a client's negative change relative to a baseline score, it is important to consider how the incidence of deterioration may be impacted by the symptom severity clients are experiencing when they first present for treatment. Two studies collected for the current review addressed this topic directly. Vittengl and colleagues' (2016) meta-analysis on RCTs comparing pharmacotherapy and cognitive-behavioral therapy for depression found that those with lower pre-treatment BDI scores were more likely to experience reliable deterioration. The authors speculated about reasons this effect might occur. They highlighted that all of the participants were experiencing unipolar depression, which is often episodic or cyclical in nature. Thus, those who reliably deteriorated may have sought treatment when their symptoms were less severe, and then during the course of treatment experienced the onset or worsening of a major depressive episode that represented the typical trajectory of their illness rather than harm directly caused by the treatment.

Comparably, Saxon and colleagues (2016) also found that clients with lower initial symptom severity on the CORE-OM were more likely to deteriorate. Their sample comprised 6,405 participants representing a variety of diagnoses. The authors offered a proposed explanation for their findings: a ceiling effect for those with high initial symptom severity such that their symptoms may be at a peak when treatment commences, leaving little potential for them to reliably worsen further. These conclusions intersect with those drawn by Rozental and

colleagues (2017) in their meta-analysis of deterioration in participants undergoing Internet-based cognitive-behavioral therapy, in which they found that deterioration was less likely for individuals presenting with higher initial symptom severity.

The findings in these studies are notable because a clinician who concludes, through a baseline assessment, that a client is not experiencing severe distress might be less likely to attend to potential signs of worsening in the client. In such a scenario, routine administration of outcome measures or use of a feedback system could be useful in terms of alerting the clinician to when the client may be on a negative trajectory. Receiving such information can prompt the clinician to reassess what interventions are appropriate for the client at any given time and tailor the treatment approach accordingly.

The findings in these studies also align, in part, with conclusions drawn by Mohr (1995). His review suggested higher rates of negative outcome among clients with lower initial symptom severity, but only when the outcome measure was based on the client's self-report. According to Mohr, the effect is reversed (i.e. higher levels of initial symptom severity are associated with increased incidence of negative outcomes) when the outcome measure is based on the clinician's ratings of the client. A possible interpretation is that clinicians who are rating clients' initial symptoms as more severe may be more alert to and able to readily detect deterioration when it occurs. Researchers would do well to further explore how the incidence of deterioration is impacted based upon who is providing the ratings for the outcome measures—whether clinician, client, or another treatment stakeholder. Additionally, Mohr (1995) and other authors (e.g. McAleavey, et al., 2012) have recommended that clinicians administer a combination of both client-rated and clinician-rated instruments, a recommendation echoed here.

Other Negative Effects

Separate from reliable deterioration, several constructs related to negative treatment effects emerged in studies collected for the current review. Due to the fact that the current review centered primarily on deterioration, this discussion of other negative effects is not comprehensive. These other negative effects included adverse events (e.g. suicide attempts, self-harm, hospitalizations), non-response to treatment, general premature termination of treatment or withdrawing from the study specifically due to worsening symptoms (e.g. Bastos, Guimaraes, & Trentini, 2015; Gloster et al., 2011; Schlögelhofer et al., 2013; Schmidt et al., 2015; Woody, Whittal, & McClean, 2011), onset of mood episodes (e.g. Reynolds III et al., 2014), symptom substitution (e.g. Peterson et al. 2016), meeting criteria for a disorder or problem that one had not met criteria for previously, (e.g. Holmqvist, Philips, & Mellor-Clark, 2015; Monti et al., 2014; van't Veer-Tazelaar et al., 2011), relapse/loss of gains after treatment or after a period of improvement during treatment (e.g. Adler, Harmeling, & Walder-Biesanz, 2013; Greenfield, Gunthert, & Haaga, 2011; Harvey et al., 2014; Lutz et al., 2013), and a variety of other negative effects (e.g. "minimal," "minor," or "transient" worsening of symptoms; participants self-reporting at the end of treatment that they generally felt "worse" but there was no baseline measure with which to compare this pre-treatment information; e.g. Keller, Feeny, & Zoellner, 2014; Kurzweil, 2012).

It is encouraging to see such a variety of negative effects measured and reported upon. However, they were not always clearly defined. For instance, some authors indicated that adverse events occurred during the study, but did not provide clarification about what these adverse events entailed. Further specification in such cases is recommended in order to increase transparent communication between authors and consumers of research, as well as to facilitate

future research on the incidence, nature, and prevention of negative effects that occur within psychotherapy.

Moreover, more specific and consistent definitions of the variety of negative effects would greatly enhance the ability of researchers and clinicians alike to communicate precisely about the topic and to make cross-study comparisons. Influenced by Linden (2012), Linden and Schermuly-Haupt (2014), and Parry, Crawford, and Duggan (2016), it is proposed that general terms such as "adverse effects" or "negative effects" of psychotherapy serve as the broadest category. Under this umbrella, "adverse events" may be considered discrete occurrences that take place during treatment or directly after treatment. In contrast, it is suggested that reliable deterioration is a more persistent worsening of symptoms that is measured via a specific statistical criterion or metric. Distinguishing whether rates of reliable deterioration have been measured during treatment, at post-treatment, or after treatment has ended is also recommended.

Unpublished Data

Relatively few unpublished studies were identified in the current review. As a result, there is likely not adequate information to support or refute past literature (e.g. Lilienfeld, 2007; Mohr, 1995) that has proposed that information about client deterioration may fall victim to the "file-drawer problem," specifically that bias on the part of researchers and publishers could lead to under-publication of psychotherapy outcome studies whose results show higher rates of deterioration than what is commonly reported in the current literature. Several authors who conducted studies found in the current review were willing to provide their deterioration rates upon request via email, although these authors did not publish the deterioration rates in their original papers. Examination of these deterioration rates did not offer strong evidence that the authors had excluded the rates from their publications because the rates were higher than

average. Nevertheless, publication bias may be present when it comes to deterioration and this topic warrants continued research attention. One way to conduct further investigation would be for journal editors to put out calls for unpublished literature that includes information on deterioration.

Limitations

There were several limitations to this review. Despite efforts to provide a comprehensive synthesis, the scope of the project was narrowed by a focus on adult participants and individual psychotherapy only. Furthermore, the search methods likely did not capture every psychotherapy outcome study reporting deterioration rates between 2011 and 2016, though a considerable sample of journals and outcome research studies were represented. In addition, due to the definitions of deterioration and psychotherapy that were used in this project, certain diagnoses were likely not examined in their full depth. These include substance abuse (e.g. due to relapse being a common means of defining deterioration in studies of this presenting concern), psychosis (e.g. due to psychopharmacological interventions often being the treatment of choice), and eating disorders and borderline personality disorder (e.g. due to combined group and individual therapy often being the treatment of choice). Thus, caution in generalizing these findings to populations with these diagnoses is advised.

CHAPTER 5

CONCLUSIONS

One major purpose of the current review was determine how often and in what forms client deterioration is acknowledged in contemporary psychotherapy outcome studies. Results indicated that reliable deterioration rates are not clearly and distinctly reported in the majority of such studies. Remediating this problem would improve the quality and transparency of psychotherapy outcome research in general, and enhance future research on the topic of deterioration in particular. Thus, it is recommended that psychotherapy outcome researchers make a habit of considering how to monitor deterioration and other negative effects among their participants, directly state in their publications how these concepts were defined and measured, and present the rate at which they occur for defined samples (e.g. intention-to-treat, completer). An important area of future research would be the comparison of deterioration rates in completer samples versus intention-to-treat samples, as clients who deteriorate may drop out before treatment is complete. Additionally, treating participants who reliably deteriorate as separate from participants who experience no reliable change is recommended in future research in order to more precisely report on the effectiveness of psychotherapies.

Among the deterioration rates collected in this review, the majority referred to deterioration that occurred when comparing pre- to post-treatment scores on outcome measures. These post-treatment rates encompassed a wide range. However, the vast majority fell between 0 and 10% and tended to cluster between zero and five percent, with zero percent deterioration being reported prevalently. Overall, such rates are lower than those reported in some previous large-scale studies of deterioration. This may represent a shift toward greater effectiveness in the treatment or prevention of deterioration in recent years, or it may also be a product of publication

bias, small sample sizes, or a floor or ceiling effect among some of the outcome measures administered.

This investigation also explored several factors that may be associated with differential rates of deterioration and/or require further research to explicate a potential relationship. Deterioration is a multi-faceted concept and how to define and measure it has been a long-debated subject. Thus, it is not surprising that one major factor that appears to impact deterioration rates is the outcome measure that is used to detect it. Greater consistency in the outcome measures that are employed in psychotherapy research, and in turn, cross-study comparisons of deterioration rates for various instruments, are a recommended next step of investigation. Utilization of a core battery of outcome measures, including both symptom-specific and global measures, is one way this could be accomplished. Furthermore, continued study is recommended in order to examine how client socio-cultural factors, client initial symptom severity, therapist effects, treatment dose and duration, and the use of routine outcome monitoring and progress feedback may directly impact, or interact to impact, deterioration rates. Several studies included in the current review have already investigated such questions, but their results would benefit from replication.

While client deterioration appears to be a growing topic of interest within psychotherapy research, there is much about it that remains to come to light. Acknowledging its occurrence and uniformly including it in published research is essential to fostering greater understanding of the nature of deterioration, both empirically and theoretically. Such knowledge can then guide efforts to prevent and intervene within clinical work, increasing the likelihood of effective and ethical treatment.

References

Ackerman, S. J., & Hilsenroth, M. J. (2001). A review of therapist characteristics and techniques negatively impacting the therapeutic alliance. *Psychotherapy: Theory, Research, Practice, Training, 38*(2), 171-185. https://doi.org/10.1037//0033-3204.38.2.171

Adler, J. M., Harmeling, L. H., & Walder-Biesanz, I. (2013). Narrative meaning making is associated with sudden gains in psychotherapy clients' mental health under routine clinical conditions. *Journal of Consulting and Clinical Psychology, 81*(5), 839-845. https://doi.org/10.1037/a0033774

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*(12), 1060–1073. https://doi.org/10.1037//0003-066x.57.12.1060

American Psychological Association (2013). Recognition of psychotherapy effectiveness. *Psychotherapy, 50*(1), 102-109. https://doi.org/10.1037/a0030276

Atkins, D. C., Bedics, J. D., McGlinchey, J. B., & Beauchaine, T. P. (2005). Assessing clinical significance: Does it matter which method we use?. *Journal of Consulting and Clinical Psychology*, *73*(5), 982-989. https://doi.org/10.1037/0022-006x.73.5.982

Banham, J. A., & Schweitzer, R. D. (2015). Trainee-therapists are not all equal: Examination of therapeutic efficiency, effectiveness and early client dropout after 12 months of clinical training. *Psychology and Psychotherapy: Theory, Research and Practice*, *89*, 148-162. https://doi.org/10.1111/papt.12071

Barlow, D. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist, 65*(1), 13-20. https://doi.org/10.1037/a0015643

Barnes, M., Sherlock, S., Thomas, L., Kessler, D., Kuyken, W., Owen-Smith, A., ... & Turner, K. (2013). No pain, no gain: Depressed clients' experiences of cognitive behavioural

therapy. *British Journal of Clinical Psychology*, *52*(4), 347-364. https://doi.org/10.1111/bjc.12021

Bastos, A. G., Guimaraes, L. S. P., & Trentini, C. M. (2015). The efficacy of long-term psychodynamic psychotherapy, fluoxetine and their combination in the outpatient treatment of depression. *Psychotherapy Research*, *25*(5), 612-624. https://doi.org/10.1080/10503307.2014.935519

Benjamin, L. S. (2015). The arts, crafts, and sciences of psychotherapy. *Journal of Clinical Psychology, 71*(11), 1070-1082. https://doi.org/10.1002/jclp.22217

Bergin, A. E. (1966). Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology*, *71*(4), 235-246.

Berk, M., & Parker, G. (2009). The elephant on the couch: Side-effects of psychotherapy. *Australian and New Zealand Journal of Psychiatry*, *43*(9), 787-794. https://doi.org/10.1080/00048670903107559

Boisvert, C. M., & Faust, D. (2002). Iatrogenic symptoms in psychotherapy. *American Journal of Psychotherapy, 56*(2), 244-259.

Boisvert, C. M., & Faust, D. (2006). Practicing psychologists' knowledge of general psychotherapy research findings: Implications for science-practice relations. *Professional Psychology: Research and Practice*, *37*(6), 708-716. https://doi.org/10.1037/0735-7028.37.6.708

Bootzin, R. R., & Bailey, E. T. (2005). Understanding placebo, nocebo, and iatrogenic treatment effects. *Journal of Clinical Psychology*, *61*(7), 871-880. https://doi.org/10.1002/jclp.20131

Bystedt, S., Rozental, A., Andersson, G., Boettcher, J., & Carlbring, P. (2014). Clinicians' perspectives on negative effects of psychological treatments. *Cognitive Behaviour Therapy*, *43*(4), 319-331. https://doi.org/10.1080/16506073.2014.939593

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49*(4), 421-453. https://doi.org/10.1348/014466509x470789

Callahan, J. L., Gustafson, S. A., Misner, J. B., Paprocki, C. M., Sauer, E. M., Saules, K. K., ... & Wise, E. H. (2014). Introducing the association of psychology training clinics' collaborative research network: A study on client expectancies. *Training and Education in Professional Psychology, 8*(2), 95-103. https://doi.org/10.1037/tep0000047

Campbell, T. C. (2005). An introduction to clinical significance: An alternative index of intervention effect for group experimental designs. *Journal of Early Intervention, 27*(3), 210-227. https://doi.org/10.1177/105381510502700307

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*(1), 34-49. https://doi.org/10.1037/a0017330

[a.] CORE IMS (2010). Recovery and improvement rates. In *Benchmarks for primary care counselling services*. Retrieved from http://www.coreims.co.uk/Support_User_ Benchmarking.html.

[b.] CORE IMS (2010). Recovery and improvement rates. In *Benchmarks for higher education counselling services*. Retrieved from http://www.coreims.co.uk/Support_User_ Benchmarking.html.

Crawford, M. J., Thana, L., Farquharson, L., Palmer, L., Hancock, E., Bassett, P., ... & Parry, G. D. (2016). Patient experience of negative effects of psychological treatment: Results of a national survey. *The British Journal of Psychiatry, 208*(3), 260-265. https://doi.org/10.1192/bjp.bp.114.162628

DeGeorge, J. L. (2014). *Individual differences in psychotherapy change among ethnic minority patients* (unpublished doctoral dissertation). University of Massachusetts-Amherst, Amherst, MA.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice, 12*(2), 174-183. https://doi.org/10.1093/clipsy/bpi022

Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *American Psychologist, 65*(1), 21-33. https://doi.org/10.1037/a0017299

Duggan, C., Parry, G., McMurran, M., Davidson, K., & Dennis, J. (2014). The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. *Trials, 15*(335), 1-9. https://doi.org/10.1186/1745-6215-15-335

Ebert, D. D., Donkin, L., Andersson, G., Andrews, G., Berger, T., Carlbring, P., ... & Kleiboer, A. (2016). Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. *Psychological Medicine, 46*(13), 2679-2693. https://doi.org/10.1017/s0033291716001562

Edgerton, J. E., & Campbell III, R. J. (1994). *American psychiatric glossary* (7th ed.). Washington, DC: American Psychiatric Press, Inc.

Ellison, W. D., Levy, K. N., Cain, N. M., Ansell, E. B., & Pincus, A. L. (2013). The impact of pathological narcissism on psychotherapy utilization, initial symptom severity, and early-treatment symptom change: A naturalistic investigation. *Journal of Personality Assessment, 95*(3), 291-300. https://doi.org/10.1080/00223891.2012.742904

Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of Consulting and Clinical Psychology, 84*(8), 738-751. https://doi.org/10.1037/ccp0000119

Gloster, A. T., Wittchen, H. U., Einsle, F., Lang, T., Helbig-Lang, S., Fydrich, T., ... & Gerlach, A. L. (2011). Psychological treatment for panic disorder with agoraphobia: a randomized controlled trial to examine the role of therapist-guided exposure in situ in CBT. *Journal of Consulting and Clinical Psychology, 79*(3), 406-420. https://doi.org/10.1037/a0023584

Gold, J., & Stricker, G. (2011). Failures in psychodynamic psychotherapy. *Journal of Clinical Psychology, 67*(11), 1096-1105. https://doi.org/10.1002/jclp.20847

Greenfield, M. F., Gunthert, K. C., & Haaga, D. A. (2011). Sudden gains versus gradual gains in a psychotherapy training clinic. *Journal of Clinical Psychology, 67*(1), 17-30. https://doi.org/10.1002/jclp.20748

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163. https://doi.org/10.1002/jclp.20108

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*(3), 329-343. https://doi.org/10.1093/clipsy/9.3.329

Harvey, A. G., Bélanger, L., Talbot, L., Eidelman, P., Beaulieu-Bonneau, S., Fortier-Brochu, É., ... & Mérette, C. (2014). Comparative efficacy of behavior therapy, cognitive therapy, and cognitive behavior therapy for chronic insomnia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 82*(4), 670-683. https://doi.org/10.1037/a0036606

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17*(1), 25-32. https://doi.org/10.1002/cpp.656

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35*(5), 485-491. https://doi.org/10.1037/0735-7028.35.5.485

Hiller, W., Schindler, A. C., & Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research, 22*(1), 1-11. https://doi.org/10.1080/10503307.2011.616237

Høglend, P., Hersoug, A. G., Bøgwald, K. P., Amlo, S., Marble, A., Sørbye, Ø., ... & Crits-Christoph, P. (2011). Effects of transference work in the context of therapeutic alliance and quality of object relations. *Journal of Consulting and Clinical Psychology, 79*(5), 697-706. https://doi.org/10.1037/a0024863

Holmqvist, R., Philips, B., & Mellor-Clark, J. (2015). Client and therapist agreement about the client's problems—Associations with treatment alliance and outcome. *Psychotherapy Research, 26*(4), 399-409. https://doi.org/10.1080/10503307.2015.1013160

Hopko, D. R., Magidson, J. F., & Lejuez, C. W. (2011). Treatment failure in behavior therapy: Focus on behavioral activation for depression. *Journal of Clinical Psychology*, *67*(11), 1106-1116. https://doi.org/10.1002/jclp.20840

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*(4), 678-685. https://doi.org/10.1037//0022-006x.61.4.678

Ioannidis, J. P., Evans, S. J., Gotzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine*, *141*(10), 781–788. https://doi.org/10.7326/0003-4819-141-10-200411160-00009

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*. *15*(4), 336-352. https://doi.org/10.1016/s0005-7894(84)80002-7

Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*(3), 300-307. https://doi.org/10.1037//0022-006x.67.3.300

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*(1), 12-19. https://doi.org/10.1037/10109-042

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), 1-8. https://doi.org/10.1016/j.cct.2014.02.005

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*(3), 332-339. https://doi.org/10.1037//0022-006x.67.3.332

Keller, S. M., Feeny, N. C., & Zoellner, L. A. (2014). Depression sudden gains and transient depression spikes during treatment for PTSD. *Journal of Consulting and Clinical Psychology, 82(*1), 102-111. https://doi.org/10.1037/a0035286

Kendall, P. C., Comer, J. S., & Chow, C. (2013). The randomized controlled trial: Basics and beyond. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 40-61). New York, NY: Oxford University Press.

Kim, S. H., Beretvas, S. N., & Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modeling. *Measurement and Evaluation in Counseling and Development, 42*(4), 275-295. https://doi.org/10.1177/0748175609354616

Kraus, D. R., Seligman, D., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61,* 285–314. https://doi.org/10.1002/jclp.20084

Kurzweil, S. (2012). Psychodynamic therapy for depression in women with infants and young children. *American Journal of Psychotherapy, 66*(2), 181-199.

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), 1-14. https://doi.org/10.1080/10503300601032506

Lambert, M. J. (2011). What have we learned about treatment failure in empirically supported treatments? Some suggestions for practice. *Cognitive and Behavioral Practice, 18*(3), 413-420. https://doi.org/10.1016/j.cbpra.2011.02.002

Lambert, M. J. (2012). Helping clinicians to use and learn from research-based systems: The OQ-analyst. *Psychotherapy*, *49*(2), 109-114. https://doi.org/10.1037/a0027110

[a.] Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M.J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.) (pp. 169-218). Hoboken, NJ: John Wiley & Sons, Inc.

[b.] Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50*(1), 42-51. https://doi.org/10.1037/a0030682

Lambert, M. J., Bergin, A. E., & Collins, J. L. (1977). Therapist-induced deterioration in psychotherapy. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research* (pp. 452-481). Pergamon Press Inc.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1999). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy*, *3*(4), 249-258.

[a.] Lambert, M. J., Gregersen, A. T., & Burlingame, G. M. (2004). The Outcome Questionnaire—45. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (3rd ed., Vol. 3) (pp. 191-234). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[b.] Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., … Burlingame, G. M. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)*. Salt Lake City, UT: OQ Measures, LLC.

Lambert, M. J. & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139 –193). New York, NY: Wiley.

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*(1), 72-79. https://doi.org/10.1037/a0022238

Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced?. *Psychotherapy Research*, *11*(1), 49-68. https://doi.org/10.1093/ptr/11.1.49

Laurenceau, J. P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, *27*(6), 682-695. https://doi.org/10.1016/j.cpr.2007.01.007

Leventhal, L., & Huynh, C. (1996). Directional decisions for two-tailed tests: Power, error rates, and sample size. *Psychological Methods*, *1*(3), 278–292. https://doi.org/10.1037//1082-989x.1.3.278

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*(1), 53-70. https://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*(4), 286-296. https://doi.org/10.1002/cpp.1765

Linden, M., & Schermuly-Haupt, M. L. (2014). Definition, assessment and rate of psychotherapy side effects. *World Psychiatry*, *13*(3), 306-309. https://doi.org/10.1002/wps.20153

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.A., Jorasz, C., ... & Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14-24. https://doi.org/10.1080/10503307.2012.693837

Marble, A., Høglend, P., & Ulberg, R. (2011). Change in self-protection and symptoms after dynamic psychotherapy: The influence of pretreatment motivation. *Journal of Clinical Psychology, 67*(4), 355-367. https://doi.org/10.1002/jclp.20771

Mash, E. J., & Hunsley, J. (1993). Assessment considerations in the identification of failing psychotherapy: Bringing the negatives out of the darkroom. *Psychological Assessment, 5*(3), 292-301. https://doi.org/10.1037//1040-3590.5.3.292

Mays, D. T., & Franks, C. M. (1985). *Negative outcome in psychotherapy and what to do about it*. New York, NY: Springer Publishing Company.

McAleavey, A. A., Nordberg, S. S., Kraus, D., & Castonguay, L. G. (2012). Errors in treatment outcome monitoring: Implications for real-world psychotherapy. *Canadian Psychology, 53*(2), 105-114. https://doi.org/10.1037/a0027833

Meister, R., von Wolff, A., Mohr, H., Nestoriuc, Y., Härter, M., Hölzel, L., & Kriston, L. (2016). Adverse event methods were heterogeneous and insufficiently reported in randomized trials on persistent depressive disorder. *Journal of Clinical Epidemiology, 71*, 97-108. https://doi.org/10.1016/j.jclinepi.2015.10.007

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice, 2*(1), 1-27. https://doi.org/10.1111/j.1468-2850.1995.tb00022.x

Monti, P. M., Colby, S. M., Mastroleo, N. R., Barnett, N. P., Gwaltney, C. J., Apodaca, T. R., ... & Biffl, W. L. (2014). Individual versus significant-other-enhanced brief motivational

intervention for alcohol in emergency care. *Journal of Consulting and Clinical Psychology, 82*(6), 936-948. https://doi.org/10.1037/a0037658

Moos, R. H. (2005). Iatrogenic effects of psychosocial interventions for substance use disorders: Prevalence, predictors, prevention. *Addiction, 100,* 595–604. https://doi.org/10.1111/j.1360-0443.2005.01073.x

Moos, R. H. (2012). Iatrogenic effects of psychosocial interventions: Treatment, life context, and personal risk factors. *Substance Use & Misuse*, *47*(13-14), 1592-1598. https://doi.org/10.3109/10826084.2012.705710

Moos, R. H., Moos, B. S., & Finney, J. W. (2001). Predictors of deterioration among patients with substance-use disorders. *Journal of Clinical Psychology*, *57*(12), 1403-1419. https://doi.org/10.1002/jclp.1105

Naeem, F., Gire, N., Xiang, S., Yang, M., Syed, Y., Shokraneh, F., ... & Farooq, S. (2016). Reporting and understanding the safety and adverse effect profile of mobile apps for psychosocial interventions: An update. *World Journal of Psychiatry*, *6*(2), 187-191. https://doi.org/10.5498/wjp.v6.i2.187

Nicholson, H., Foote, C., & Grigerick, S. (2009). Deleterious effects of psychotherapy and counseling in the schools. *Psychology in the Schools*, *46*(3), 232-237. https://doi.org/10.1002/pits.20367

Nolan, S. A., Strassle, C. G., Roback, H. B., & Binder, J. L. (2004). Negative treatment effects in dyadic psychotherapy: A focus on prevention and intervention strategies. *Journal of Contemporary Psychotherapy*, *34*(4), 311-330. https://doi.org/10.1007/s10879-004-2526-6

Nordberg, S. S., Castonguay, L. G., Fisher, A. J., Boswell, J. F., & Kraus, D. (2014). Validating the rapid responder construct within a practice research network. *Journal of Clinical Psychology, 70*(9), 886-903. https://doi.org/10.1002/jclp.22077

Nutt, D. J., & Sharpe, M. (2008). Uncritical positive regard? Issues in the efficacy and safety of psychotherapy. *Journal of Psychopharmacology*, *22*(1), 3-6. https://doi.org/10.1177/0269881107086283

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 134-166). Hoboken, NJ: John Wiley & Sons, Inc.

Ogles, B. M., Lunnen, K. M., & Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*(3), 421-446. https://doi.org/10.1016/s0272-7358(99)00058-6

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology*, *62*(9), 1157-1172. https://doi.org/10.1002/jclp.20272

OQ Measures (2014). Measures. Retrieved from http://oqmeasures.wpengine.com/measures/.

Owen, J., Adelson, J., Budge, S., Wampold, B., Kopta, M., Minami, T., & Miller, S. (2015). Trajectories of change in psychotherapy. *Journal of Clinical Psychology*, *71*(9), 817-827. https://doi.org/10.1002/jclp.22191

Overington, L., & Ionita, G. (2012). Progress monitoring measures: A brief guide. *Canadian Psychology*, *53*(2), 82-92. https://doi.org/10.1037/a0028017

Parry, G. D., Crawford, M. J., & Duggan, C. (2016). Iatrogenic harm from psychological therapies–time to move on. *The British Journal of Psychiatry*, *208*(3), 210-212. https://doi.org/10.1192/bjp.bp.115.163618

Peterson, A. L., McGuire, J. F., Wilhelm, S., Piacentini, J., Woods, D. W., Walkup, J. T., ... & Scahill, L. (2016). An empirical examination of symptom substitution associated with behavior therapy for Tourette's Disorder. *Behavior Therapy, 47*(1), 29-41. https://doi.org/10.1016/j.beth.2015.09.001

Ravitz, P., McBride, C., & Maunder, R. (2011). Failures in interpersonal psychotherapy (IPT): Factors related to treatment resistances. *Journal of Clinical Psychology*, *67*(11), 1129-1139. https://doi.org/10.1002/jclp.20850

Reynolds III, C. F., Thomas, S. B., Morse, J. Q., Anderson, S. J., Albert, S., Dew, M. A., ... & Stack, J. A. (2014). Early intervention to preempt major depression among older black and white adults. *Psychiatric Services, 65*(6), 765-773. https://doi.org/10.1176/appi.ps.201300216

Rhule, D. M. (2005). Take care to do no harm: Harmful interventions for youth problem behavior. *Professional Psychology: Research and Practice, 36*(6), 618-625. https://doi.org/10.1037/0735-7028.36.6.618

Rieck, T., & Callahan, J. L. (2013). Emotional intelligence and psychotherapy outcomes in the training clinic. *Training and Education in Professional Psychology, 7*(1), 42-52. https://doi.org/10.1037/a0031659

Roback, H. B. (2000). Adverse outcomes in group psychotherapy: Risk factors, prevention, and research directions. *Journal of Psychotherapy Practice and Research*, *9*(3), 113-122.

Ronk, F. R., Hooke, G. R., & Page, A. C. (2012). How consistent are clinical significance classifications when calculation methods and outcome measures differ?. *Clinical Psychology: Science and Practice*, *19*(2), 167-179. https://doi.org/10.1111/j.1468-2850.2012.01281.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. https://doi.org/10.1037//0033-2909.86.3.638

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*(1), 12-19. https://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*(3), 223-236. https://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*(6), 1-22. https://doi.org/10.1371/journal.pone.0157503

Rozental, A., Magnusson, K., Boettcher, J., Andersson, G., & Carlbring, P. (2017). For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy. *Journal of Consulting and Clinical Psychology*, *85*(2), 160-177. https://doi.org/10.1037/ccp0000158

Schindler, A. C., Hiller, W., & Witthöft, M. (2011). Benchmarking of cognitive-behavioral therapy for depression in efficacy and effectiveness studies—How do exclusion criteria affect treatment outcome?. *Psychotherapy Research, 21*(6), 644-657. https://doi.org/10.1080/10503307.2011.602750

Schlögelhofer, M., Willinger, U., Wiesegger, G., Eder, H., Priesch, M., Itzlinger, U., ... & Aschauer, H. (2014). Clinical study results from a randomized controlled trial of cognitive behavioural guided self-help in patients with partially remitted depressive disorder. *Psychology and Psychotherapy: Theory, Research and Practice, 87*(2), 178-190. https://doi.org/10.1111/papt.12008

Schmidt, U., Magill, N., Renwick, B., Keyes, A., Kenyon, M., Dejong, H., ... & Watson, C. (2015). The Maudsley Outpatient Study of Treatments for Anorexia Nervosa and Related Conditions (MOSAIC): Comparison of the Maudsley Model of Anorexia Nervosa Treatment for Adults (MANTRA) with Specialist Supportive Clinical Management (SSCM) in outpatients with broadly defined anorexia nervosa: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 83*(4), 796-807. https://doi.org/10.1037/ccp0000019

Schnurr, P. P., Friedman, M. J., Engel, C. C., Foa, E. B., Shea, M. T., Chow, B. K., ... & Bernardy, N. (2007). Cognitive behavioral therapy for posttraumatic stress disorder in women: A randomized controlled trial. *The Journal of the American Medical Association, 297*(8), 820-830. https://doi.org/10.1001/jama.297.8.820

Scogin, F., Floyd, M., Jamison, C., Ackerson, J., Landreville, P., & Bissonnette, L. (1996). Negative outcomes: What is the evidence on self-administered treatments?. *Journal of*

Consulting and Clinical Psychology, 64(5), 1086-1089. https://doi.org/10.1037//0022-006x.64.5.1086

Scott, J., & Young, A. H. (2016). Psychotherapies should be assessed for both benefit and harm. *British Journal of Psychiatry, 208*(3), 208-209. https://doi.org/10.1192/bjp.bp.115.169060

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298-311. https://doi.org/10.1037/a0019247

Steenkamp, M. M., & Litz, B. T. (2013). Psychotherapy for military-related posttraumatic stress disorder: Review of the evidence. *Clinical Psychology Review, 33*(1), 45-53. https://doi.org/10.1016/j.cpr.2012.10.002

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*(3), 187-196. https://doi.org/10.1037//0003-066x.32.3.187

Svartberg, M., Stiles, T. C., & Seltzer, M. H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *American Journal of Psychiatry, 161*(5), 810-817. https://doi.org/10.1176/appi.ajp.161.5.810

Swift, J. K., Callahan, J. L., Heath, C. J., Herbert, G. L., & Levine, J. C. (2010). Applications of the psychotherapy phase model to clinically significant deterioration. *Psychotherapy: Theory, Research, Practice, Training, 47*(2), 235-248. https://doi.org/10.1037/a0019787

Tarescavage, A. M., & Ben-Porath, Y. S. (2014). Psychotherapeutic outcomes measures: A critical review for practitioners. *Journal of Clinical Psychology, 70*(9), 808-830. https://doi.org/10.1002/jclp.22080

van't Veer-Tazelaar, P. J., van Marwijk, H. W., van Oppen, P., van der Horst, H. E., Smit, F., Cuijpers, P., & Beekman, A. T. (2011). Prevention of late-life anxiety and depression has sustained effects over 24 months: A pragmatic randomized trial. *The American Journal of Geriatric Psychiatry, 19*(3), 230-239. https://doi.org/10.1097/jgp.0b013e3181faee4d

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*(4), 849-855. https://doi.org/10.1016/j.comppsych.2014.01.001

Vittengl, J. R., Jarrett, R. B., Weitz, E., Hollon, S. D., Twisk, J., Cristea, I., ... & Faramarzi, M. (2016). Divergent outcomes in cognitive-behavioral therapy and pharmacotherapy for adult depression. *American Journal of Psychiatry*, *173*(5), 481-490. https://doi.org/10.1176/appi.ajp.2015.15040492

Vøllestad, J., Sivertsen, B., & Nielsen, G. H. (2011). Mindfulness-based stress reduction for patients with anxiety disorders: Evaluation in a randomized controlled trial. *Behaviour Research and Therapy*, *49*(4), 281-288. https://doi.org/10.1016/j.brat.2011.01.007

Watson, J. C. (2011). Treatment failure in humanistic and experiential psychotherapy. *Journal of Clinical Psychology*, *67*(11), 1117-1128. https://doi.org/10.1002/jclp.20849

Wendt, D. C., Gone, J. P., & Nagata, D. K. (2015). Potentially harmful therapy and multicultural counseling: Extending the conversation. *The Counseling Psychologist, 43*, 393–403. https://doi.org/10.1177/0011000015576801

Werch, C. E., & Owen, D. M. (2002). Iatrogenic effects of alcohol and drug prevention programs. *Journal of Studies on Alcohol, 63*(5), 581-590. https://doi.org/10.15288/jsa.2002.63.581

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*, 875–899. https://doi.org/10.1037//0022-006x.69.6.875

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 and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59-68. https://doi.org/10.1037//0022-0167.50.1.59

Woody, S. R., Whittal, M. L., & McLean, P. D. (2011). Mechanisms of symptom reduction in treatment for obsessions. *Journal of Consulting and Clinical Psychology, 79*(5), 653-664. https://doi.org/10.1037/a0024827

Yalom, I. D., & Lieberman, M. A. (1971). A study of encounter group casualties. *Archives of General Psychiatry, 25* (1), 16-30. https://doi.org/10.1001/archpsyc.1971.01750130018002

Appendix A

List of Journals Selected for Manual Review

1. American Journal of Psychotherapy
2. Clinical Psychology and Psychotherapy
3. Clinical Psychology Review
4. Clinical Psychology: Science and Practice
5. Counselling and Psychotherapy Research
6. Journal of Cognitive Psychotherapy
7. Journal of Clinical Psychology
8. Journal of Consulting and Clinical Psychology
9. Journal of Contemporary Psychotherapy
10. Journal of Counseling and Development
11. Journal of Counseling Psychology
12. Journal of Humanistic Counseling
13. Journal of Psychotherapy Integration
14. Professional Psychology: Research and Practice
15. Psychodynamic Practice (formerly known as Psychodynamic Counselling)
16. Psychological Services
17. Psychology and Psychotherapy: Theory, Research, and Practice
18. Psychotherapy
19. Psychotherapy Research

Appendix B

List of Studies Reporting Reliable Deterioration Rates

1. Aafjes-van Doorn , Macdonald, Stein, Cooper, & Tucker (2014)
2. Alves et al. (2014)
3. Amble, Gude, Stubdal, Andersen, & Wampold (2015)
4. Ashworth et al. (2015)
5. Aubochon-Endsley, Callahan, & Scott (2014)
6. Bambling & King (2013)
7. Barr, Hodge, Levan, Bowen, & Knox (2012)
8. Beck, Burdett, & Lewis (2014)
9. Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow (2012)
10. Berger, Boettcher, & Caspar (2016)
11. Bernecker et al. (2016)*
12. Bevan, Wittkowski, & Wells (2013)
13. Bhattacharya (2015)
14. Blais et al. (2012)
15. Blais et al. (2013)
16. Boersma, Håkanson, Salomonsson, & Johansson (2015)
17. Bosley, Fisher, & Taylor (2016)
18. Boswell, McAleavey, Castonguay, Hayes, & Locke (2012)
19. Branson, Shafran, & Myles (2015)
20. Briggie, Hilsenroth, Conway, Muran, & Jackson (2016)
21. Bryan et al. (2012)
22. Butollo, König, Karl, Henkel, & Rosner (2014)
23. Callahan et al. (2014)
24. Chow et al. (2015)
25. Connolly Gibbons et al. (2012)
26. Connolly Gibbons et al. (2015)
27. Dakwar & Levin (2013)
28. de Jong (2012)**
29. de Jong, van Sluis, Nugter, Heiser, & Spinhoven (2012)
30. de Jong et al. (2014)
31. Dennhag & Armelius (2012)
32. Dennhag, Ybrandt, & Armelius (2011)
33. Eddington, Silvia, Foxworth, Hoet, & Kwapil, (2015)*
34. Erekson, Lambert, & Eggett (2015)
35. Farima, Dowlatabadi, Behzadi (2015)
36. Feldner, Smith, Monson, & Zvolensky (2013)

37. Fisher, Atzil-Slonim, Bar-Kalifa, Rafaeli, & Peri (2016)
38. Frets, Kevenaar, & van der Heiden (2013)
39. Galovski, Blain, Mott, Elwood, & Houle (2012)
40. Goldberg et al (2016)*
41. Gonçalves et al. (2012)
42. Goodman, Edwards, & Chung (2013)
43. Gullestad et al. (2012)
44. Gumz, Bauer, & Brähler (2012)
45. Hardy, Tracey, Glidden-Tracey, Hess, & Rohlfing (2011)
46. Hayes, Owen, & Bieschke (2015)
47. Hiller, Schindler, & Lambert (2012)
48. Kellett, Bennett, Ryle, & Thake (2013)
49. Koszycki, Bilodeau, Raab-Mayo, & Bradwejn (2014)*
50. Kramer, de Roten, Drapeau, & Despland (2013)
51. Kramer, Pascual-Leone, Despland, & de Roten (2015)
52. Kramer et al. (2011)
53. Kraus, Castonguay, Boswell, Nordberg, & Hayes (2011)
54. Leibert & Dunne-Bryant (2015)*
55. Littleton, Buck, Rosman, & Grills-Taquechel (2012)
56. Littleton, Grills, Kline, Schoemann, & Dodd (2016)
57. Lopes et al. (2014)
58. Lutz et al. (2014)
59. McElvaney & Timulak (2013)
60. McLay et al. (2011)
61. Mechler & Holmqvist (2015)
62. Mellor-Clark, Twigg, Farrell, & Kinder (2013)
63. Murphy & Cramer (2014)
64. Murphy, Rashleigh, & Timulak (2012)
65. Murray, McKenzie, Murray, & Richelieu (2015)
66. Newman et al. (2011)*
67. Olthuis, Watt, Mackinnon, & Stewart (2014)
68. Payne, Ciclitira, Starr, Marzano, & Brunswick (2015)
69. Polman, Bouman, van Geert, de Jong, & den Boer (2011)
70. Richards, Timulak, & Hevey (2013)
71. Rieck & Callahan (2013)
72. Rogers et al. (2014)*
73. Saxon & Barkham (2012)
74. Saxon, Barkham, Foster, & Parry (2016)
75. Schindler, Hiller, & Witthoft, (2011)
76. Simon, Lambert, Harris, Busath, & Vazquez (2012)

77. Strauss et al. (2013)*
78. Taillon, O'Connor, Dupuis, & Lavoie (2013)
79. Tejas Haugen, Goldman, & Owen, 2015
80. Titov et al. (2015)
81. Vandborg, Hatmann, Bennedsen, Pedersen, & Thomsen (2015)
82. Vaz, Conceição, & Machado (2013)
83. Vromans & Schweitzer (2011)
84. Weiss, Kivity, & Huppert (2014)
85. Wells & Colbear (2012)
86. Werbart, Levin, Andersson, & Sandell (2013)
87. Willson, Veale, & Freeston (2015)
88. Yasky, King, & O'Brien (2015)

*Deterioration rates not published in the original article; obtained through email communication with the author(s)

**Dissertation