THE PROPER APPLICATION OF STATISTICAL MODELS IN COUNSELING PSYCHOLOGY: USING SURVIVAL ANALYSIS TO MODEL CLIENT OUTCOMES IN A TELEPSYCHOLOGY CLINIC

A Dissertation

by

TREY WILLIAM ARMSTRONG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Timothy R. Elliott
Committee Members,	Daniel F. Brossart
	Carly E. McCord
	James N. Burdine
Head of Department,	Shanna Hagan-Burke

August 2018

Major Subject: Counseling Psychology

ABSTRACT

There has been a "quiet methodological revolution" that has started to move psychology away from reliance on null hypothesis statistical significance testing, and the focus has shifted to the utilization of statistical/mathematical models to answer research questions. Survival data are a specific type of data that incorporate a discrete event and time. These events can vary and do not have to be literal survival, and in counseling psychology this more general conceptualization can be utilized in psychotherapy outcome research.

Outcome data was collected at the TCC on the PHQ-9 and CORE-B for clients who were residents of an underserved region of Texas. There is needed research looking at "treatment-as-usual" psychotherapy outcomes because it best represents the reality of providing clinical services, especially in a rural, underserved area. The clinically significant change paradigm was used as it provided a structure of measuring responses to therapy, and there was literature available for comparison. This type of data can demonstrate the strengths of survival analysis. This study presented specific research questions pertinent to clinics, practitioners, and researchers about client response to treatment, and these research questions were answered through a survival analytic framework.

Results from this study generally support other research showing that 11 to 14 sessions of psychotherapy lead to clinically significant change or reliable improvement on outcome measures, and additionally, there is much client growth in the first few sessions. The impact of client demographics remains unclear; however, insurance status appears to be a preliminary factor positively affecting clients in this region.

The use of survival analysis in the counseling psychology literature is minimal, with a few exceptions. However, many psychologists would not argue with a conceptualization of psychotherapy as a longitudinal process that occurs across a series of psychotherapy sessions. Moving forward, counseling psychologists are well suited to expand their clinical services through the use of telepsychology and to expand their thinking and research to involve statistical models that better represent the reality they are trying to capture and understand.

DEDICATION

This dissertation is dedicated to Mary Wacey who first taught me about psychology and inspired me to pursue a career in psychology. To Dr. Bruce Thompson and Dr. Victor Willson who provided a solid foundation in statistics and advanced training in longitudinal data analysis and have steered me down the career path I am on. Finally, to my parents who have always supported me and encouraged me to pursue my dreams.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Timothy Elliott, for all of his support and understanding throughout my doctoral program, and I would like to thank my committee for their guidance and support in preparing this dissertation. I would also like to thank Dr. Michael Longnecker for his participation on this committee on behalf of Dr. Brossart. His feedback and support were much appreciated.

Thanks also goes to the Telehealth Counseling Clinic staff and counselors, and particularly, Melissa Shaffer, whose data coding efforts at the Telehealth Counseling Clinic helped make this study possible.

Finally, I would also like to thank my loving parents who have supported me all throughout my education and while I was writing “my paper.”

NOMENCLATURE

CORE-SF/B	Clinical Outcomes in Routine Evaluation-Short Form/B
CSC	Clinically Significant Change
NC	No Change
PHQ	Patient Health Questionnaire
RD	Reliable Deterioration
RI	Reliable Improvement
TCC	Telehealth Counseling Clinic

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Timothy R. Elliott and Daniel F. Brossart of the Department of Educational Psychology and Professor Carly E. McCord and Professor James N. Burdine of the Department of Health Promotion and Community Health Sciences. Dr. Michael T. Longnecker of the Department of Statistics was present on behalf of Dr. Daniel F. Brossart. All work conducted for the dissertation was completed by the student independently.

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Counseling psychologists are in the business of change. Oftentimes, this change stems from therapy, and psychologists want to know how they can positively impact their clients. Change can also take the academic interest of psychologists who want to discover how clients may change or respond when receiving therapy. Counseling psychology, and all other fields of psychology, utilize the scientific method as a means to acquire new information. They pose research questions and develop experiments and studies to find answers to these questions. As with all scientific fields, psychologists rely heavily on statistical tools (broadly defined) to help them answer their research questions.

Inside the Psychologist's Statistical Toolbox

Psychologists' choice and rationale behind using specific tools to answer specific questions must always be evaluated to ensure the novel knowledge acquired *meaningfully* contributes to the literature—acknowledging what advantages it possesses and a critical analysis of the shortcomings associated with its nature and inclusion. In the way that a carpenter would not choose a screwdriver to hammer a nail, certain statistical tools should be chosen over others because they lend themselves to certain situations, and thus, providing a strong rationale for their use in certain situations over others should not be difficult. However, as a carpenter may hammer a nail into a board with a screwdriver, so too can psychologists use statistical tools to answer questions that the tools are not best suited for. They may accomplish the job but also only approximate the answer at best and mislead their user at worst. Psychologists' use of statistical tools and mathematical models will be further explored before turning to the outline and

discussion of survival data—a specific type of data, research questions one could ask with such data, and a presentation of the favored statistical tools (survival analysis) to answer them.

Psychology's statistical pedagogy centers on the teaching of null hypothesis significance testing (NHST), and it is generally considered "*the* method for evaluating scientific hypotheses" (Haller & Krauss, 2002, p. 2). Psychology undergraduate students often acquire competency/mastery in the calculation of a statistical significance test (the "how"), and often, they are not aware of the reasoning or meaning behind what they are calculating (the "why"; Haller & Krauss, 2002). After taking one statistics course, the literature shows that the average student "cannot describe the underlying idea of NHST" (Falk & Greenbaum, 1995; Gigerenzer & Krauss, 2001). If given the question of how to test mean differences between a sample of men and women on some dependent variable, students would likely accurately identify the use of an independent-samples *t*-test to test these differences statistically. The insight into how this works or the interrelatedness of these tests within the general linear model (GLM; Zientek & Thompson, 2009) is unknown to them. Given this, to gain statistical competency in the field of psychology, students and researchers must be able to explain the "why" behind their chosen methodology to ensure it is appropriate for the research question it aims to answer.

Since its inception, people have raised concerns regarding the use of NHST as a scientific tool, and this debate continues today (Nickerson, 2000; Savalei & Dunn, 2015). In recent history, Cohen (1994) sparked a revolution in psychology and education, among other fields, with his article, "The Earth is Round ($p < .05$)" and brought the debate to a crisis point (Rodgers, 2010). While the debate continues, other methodologies continue forward with a "quiet revolution" supplying alternatives to NHST (Rodgers, 2010). One such "methodological revolution" argues for a departure from the NHST framework or paradigm, originally proposed by Fisher and

Neyman-Pearson, to a framework or paradigm centered on the building, comparing, and evaluating of statistical/mathematical models.

What is Survival Data and Why is it Interesting?

One such statistical model of interest in psychology is modeling survival data using nonparametric, semiparametric, or parametric survival analytic approaches. Traditionally, survival data was, quite literally, data that included when someone died after being observed for a certain period of time. In this case, the event of interest is death; more specifically, the time until the event occurs—time until death. Patient survival after a particular treatment, occurrence of certain side effects, or time until an additional infection are all examples of survival data utilized in clinical medical research. Literal survival however, does not have to be the only focus of such data, and the event under study can be adapted to different situations that are more useful. Landau (2002) discusses how survival analysis can be utilized in the field of psychology and provides a good example of analyzing psychological data. In psychology, survival data can be created by following two key points: 1) the event needs to be discrete and 2) certainty regarding its presence or absence must be known. In other words, the event must either happen or not happen, and the researcher must know whether it did or did not occur at a specific time point (e.g., after a specific measurement, after a psychological experimental condition, after a therapy session, etc.) over a certain length of time (e.g., 4 measurement periods, immediately after the experiment and for every two weeks until the end of the semester, 20 therapy sessions, etc.).

An important aspect to consider in survival data is time. You have an event that either occurs or does not occur, and you have a measure of how much time passes, or how many intervals have gone by before the event occurs. Taking only one aspect of the data at a time does

not constitute survival data, and the chosen statistical analysis must be adjusted accordingly. If time is the only factor, and one does not care about whether an event occurred or did not occur, longitudinal data analysis would be more suited to analyze how these individuals changed over time. One popular technique utilized frequently today in response to these data would be a growth curve model, such as a latent growth curve or linear growth curve model (Hedeker & Gibbons, 2006).

If one only considers the binary nature of the event and ignores the time it takes to reach the event, logistic regression can be used to predict its occurrence with the independent variables that are included in the model. However, if survival data is analyzed with logistic regression, the odds ratio poorly estimates the hazard ratio produced by a survival analysis utilizing covariates (Cox model; Peduzzi, 1987). Only in very constrained settings (the event is rare and follow-time is short) does the odds ratio approximate the hazard ratio (Peduzzi, 1987). This gives additional evidence that the inclusion of time matters when analyzing survival data.

The Consideration of Censored Data

Another aspect to consider when dealing with survival data regards censoring. When someone is censored, we possess some information about them however, we do not know exactly when the event occurred for them (Singh & Mukhopadhyay, 2011). This can happen for multiple reasons: 1) the individual did not experience the event during the study interval (e.g., anytime between session 1 and 20), 2) the individual was lost in subsequent follow-ups before reaching the event, and 3) they withdraw from the study (Singh & Mukhopadhyay, 2011). There is right and left censored data. Left censored data occurs when the individual experiences the event before the study interval starts, thus not allowing the researcher to know when the event actually occurred (Singh & Mukhopadhyay, 2011). Right censored data occurs when they are lost during

the study interval, most probably to one of the three reasons listed above (Singh & Mukhopadhyay, 2011). In other analyses, censored data may be considered missing data as the techniques may not know what to do with them, and because of this, they introduce the possibility to bias the parameter estimates of those models.

Traditional statistical methods applied to survival data can lead to misleading results given the nature of the data and the weaknesses in traditional methods to fully accommodate it. For example, a *t*-test and multiple linear regression assume normality in the residuals and homogeneity in the variance between multiple groups. Survival data do not usually take this shape as survival curves are usually skewed and residuals are not expected to be normally distributed (Landau, 2002). Logistic regression also would not handle the data well because the analysis cannot accommodate the censored data, and it would treat it as missing. Losing a percentage of your data because it was considered missing by your analysis can have big ramifications for the statistical power of your model and resulting interpretations of covariates (Landau, 2002).

Asking Research Questions about Survival Data

Survival analytic techniques have been developed that can accommodate both aspects of this type of data (discrete and time), can handle censored and skewed data, and can provide proper results for data interpretation. With regards to answering research questions, survival analytic approaches are specifically geared to answer questions such as, "How long does it take (time) for someone to reach a certain threshold on a measure (event)?" Counseling psychologists may pose the question, "How many therapy sessions (time) does a client need before they have reached clinical change (event)?" This question informs the clinician about whether the

individual has reached a certain point in therapy in which termination could be a possibility, or more generally, how long until my client gets "better."

Notice that both elements of survival data are represented in these research questions. If a researcher wants to know whether a client reaches clinically significant change or not, they would be analyzing only one aspect and would be able to model it correctly using logistic regression. It would answer the research question, "Has my client achieved clinically significant change during any point in their therapy?" However, they would not have any available information regarding time. One could argue that since time is such a critical aspect of therapy, not including it as part of your data would be leaving out a crucial aspect of counseling. A client may reach clinically significant change, but they could reach it in 8 sessions or 80 sessions. This is not as informative to the clinician. If time is the only aspect under study, one could model total scores on outcome measures using growth curves or use other longitudinal data analytic techniques; however, it does not inform the researcher on whether the client has reached a particular event such as clinically significant change or reliable improvement.

Purpose of this Study

To illustrate the concepts presented in this chapter, an example will be examined in this proposed study using clinical data from a telepsychology clinic in the Southern United States that utilizes telepsychology technology to provide counseling services to a rural, underserved clientele. The population served by this clinic lends itself well to this study as it consists of a realistic pool of people and allows the researcher to assess response to "treatment-as-usual" counseling under realistic conditions. This study will utilize two outcome measures used by the clinic to track client progress after every session: The PHQ-9 and CORE-B. These measures will be used to create a series of events based on Jacobson and Truax's (1991) notion of clinically

significant change. Notice here that the event of interest is a positive one, so we ideally want clients to achieve it sooner as opposed to analyzing factors that may have allowed them to survive longer. Kaplan-Meier survival analysis will be utilized to depict overall survival trends in the data that can be generalized to everyone. A Cox model will then demonstrate how one can include independent predictors in the model to further analyze who is more likely to achieve an event and whether their time to said event is any different than another level of the independent variable. Assumptions of the Cox model will be tested to ensure its viability, and the results of the model will be interpreted. Last, logistic regression and growth curve modeling will be conducted to compare the results across methodologies, highlighting the strengths and weaknesses of each one. In this way, the Cox model can be shown to be superior over other methods in answering these research questions.

CHAPTER II

LITERATURE REVIEW

As counseling psychologists are in the business of change, the field must also ensure that what we do positively impacts the client and promotes therapeutic change. The impact of therapy on a client can be defined in multiple ways. One distinction is between efficacy and effectiveness. Efficacy regards the positive effects therapy can have in response to a randomized clinical trial, while effectiveness loosens its constraints and evaluates therapy's effect in more realistic and natural clinical settings (Seligman, 1995). In the field of counseling psychology over the past two decades, evidence-based practice and its emphasis on evidence-supported treatments has become a prominent view of how psychologists should approach therapy (Tucker & Reed, 2008). This approach also comes with more value placed on randomized controlled trials (RCTs) as the "gold standard" of evidence, and less priority being given to other methodological designs (Tucker & Reed, 2008). This pursuit of therapy efficacy has overshadowed and sometimes, prohibited the study of effectiveness. Because of this dynamic, research on the effectiveness of therapy is not as widespread (Minami et al., 2008).

Client Outcomes from Receiving Therapy

Conclusive evidence from psychotherapy research shows that clients get better after receiving therapy treatment (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009; Lambert & Ogles, 2004). In terms of measuring outcomes from therapy, generally, *outcome* refers to how a client is after therapy compared to how they reported before starting therapy. The client is affected by therapy in direct and indirect ways, so asking clients about severity of psychological distress or quality of life/well-being questions both before and after therapy, should reveal

differences between the two time points (Cooper, 2008). While we would like for all of our clients to improve, or in other words, have negative-valence scores decrease (e.g., depression severity) and positive-valence scores increase (e.g., health-related quality of life), clients do not always improve from therapy. The nature of a client's outcome from therapy (negative or positive patterns) can be considered a client's *response* to treatment. Hansen, Lambert, and Forman (2002) quote Howard et al.'s (1986) definition of a *response* as "whether a particular outcome event (e,g., clinically significant change) has taken place, as measured by change on one or more outcome measures" (p. 331). Ultimately as clinicians, we want to know how many people reach their desired outcome in therapy, which is most likely to have a significant decrease in severe psychological distress (Cooper, 2008).

Assessing Clients throughout Therapy

Beginning in the middle of the twentieth century, researchers such as Eysenck (1952) questioned the subjective nature of counselor judgments of client improvement from psychotherapy. The counseling literature base has expanded since Eysenck's critique in the 1950s, but researchers are still concerned with the best way is to measure client improvements during therapy. One idea by Howard, Moras, Brill, Martinovich, and Lutz (1996) is what they call "patient-focused" research as a paradigm to measure client improvement throughout the course of therapy (Lambert, Hansen, & Finch, 2001). In this paradigm, the goal is to provide the clinician with "real-time" feedback to assess whether the counseling treatment is working or whether some changes are necessary for the client to get the most out of their therapy.

One way to assess this is to use outcome measures that can be given weekly to clients and can provide some valuable information on how the client is doing—both currently after any particular session and over time as these scores can be compared with scores from previous

therapy sessions (Lambert et al., 2001). Lambert et al. (2001) describe characteristics of instruments that would be suited for this purpose; these characteristics included briefness (e.g., not too many items), ability to be administered and scored quickly and easily, sensitivity to a wide range of symptoms and/or diagnoses, possessing sound psychometric properties, sensitivity to changes in short periods of time, and inexpensive. Key components widely acknowledged as indicators of client improvement should be assessed: "…symptomatic functioning (mainly anxiety and depression), interpersonal problems (friendship and family relations), and social role performance (work adjustment and quality of life)" (Lambert & Hill, 1994; Strupp, Horowitz, & Lambert, 1997; Waskow & Parloff, 1975).

Dose-effect Model of Counseling

Two research questions have been outlined that relate to client change from therapy, "How much therapy is needed to achieve significant improvement, and how much do patients benefit from each session of therapy?" (Baldwin et al., 2009, p. 203; Hansen, Lambert, & Forman, 2002). This line of research focuses on how much clients change after receiving a "dose" of therapy, usually defined as a session or a number of sessions (Baldwin et al., 2009). Studies utilizing this framework are called dose-response studies and response to treatment is modeled as a result from a dose of therapy. The dose-effect model (Howard, Kopta, Krause, & Orlinsky, 1986) has been applied to therapy and posits that there is a relationship between "dose" and rate of change in response to therapy (Baldwin et al., 2009). This pattern is postulated as being negatively accelerating, meaning that clients tend to improve as sessions continue, but the effectiveness of therapy tends to go down as session "doses" reach higher amounts (Baldwin et al., 2009).

Other models have been proposed to explain clients' response to therapy; one other model described here is the good-enough level model (GEL; Barkham et al., 2006). This model makes a different assumption that clients who come for differing number of sessions change differently in response to therapy. Clients essentially stay in therapy until they have reach a sufficient level of improvement that is "good-enough," and in conjunction with their therapist, decide to terminate therapy. This model would predict that those receiving a low dose of therapy (only a few sessions) are those that change more rapidly in their response to therapy. Those who receive a higher dose (stay for a higher number of sessions) will be those clients who change more slowly in their response to therapy (Baldwin et al., 2009). The "dose" of therapy that people receive is related to their treatment response and not completely independent of it as the dose-effect model would attest (Baldwin et al., 2009).

Clinically Significant Change

Jacobson and Truax (1991) outline their definition of what they call clinically significant change (CSC). It consists of two major criteria: 1) after receiving therapy, clients must cross over from a clinical range (determined by normative data) into a non-clinical range (also determined by normative data), and 2) client outcomes must reliably improve between pre-therapy and post-therapy (Jacobson & Truax, 1991). Clients may get better after receiving therapy, and a quantifiable statistic of this change would be an effect size. They further adjust this effect size by taking into account the reliability of the measure being used to assess therapy outcome because they do not want to attribute to the client positive change that could in actuality be error variance from a measure that is not completely reliable in assessing its symptoms or construct (internal consistency reliability). They define this statistic as a reliable change index

(RCI). If the client's score is greater than the RCI, then the clinician can be more assured that the client's improvement is from the therapy received and not from error or other "noise."

The RCI's numerator is a simple mean difference between their most recent session score or the termination score and their intake score. The RCI's denominator is the standard error of the difference (S_{diff}) between two scores. The S_{diff} uses the standard error of measurement (*SEm;* an estimate of how much of the score's standard deviation can be attributed to error) to calculate the standard error of the difference between two scores that came from the same measure. Because the S_{diff} is technically a standard error, the RCI is a mix between an effect size and a *t*-statistic. This property is advantageous because the researcher can multiply the RCI by 1.96 (equivalent to a critical *z*-value at $\alpha = .05$) to obtain a standard of change that minimizes the chance of fluke improvements.

Hansen et al. (2002) discuss three critiques to the clinically significant change concept: 1) the CSC criterion is too stringent and thus impractical, making it very difficult for clients to reach even in the best of situations, 2) the criterion is improbable for clients with chronic conditions to meet, and 3) less distressed people are unable to reach the criterion because they started off in the lower range of distress. The strengths of clinically significant change include its ability to conceptualize meaningful change both at a group and individual client level and its ease of use for clinicians in measuring routine client outcomes and procedures in clinical practice (Hansen et al., 2002).

Client Responses to Therapy from a CSC Perspective

Jacobson and Truax (1991) describe several responses that clients can have as a result of therapy. If a client achieves CSC, they are called "recovered." Those who start in the clinical range are the only clients who are able to be deemed "recovered", but there are other good

outcomes that clients can have. Clients who start in the clinical range and improve by at least the RCI on an outcome measure, are considered to have "reliably improved." Clients who start in the non-clinical range who also improve by at least the RCI are also considered to have "reliably improved." Reliable change can also take a negative perspective; clients can reliably report more distress following therapy. Clients who start in either range who have achieved at least the RCI in the negative direction (positive-valence scores decrease or negative-valence scores increase) are considered to have "reliably deteriorated." The last response is "no change" or in other words, the client did not improve or deteriorate to a reliable degree.

Shape of the Therapy Response Curve

Discussion of response to therapy to this point has focused on a pre-post treatment scenario. However, one can also plot response to treatment over the duration of treatment. This plotting of outcomes from each session depicts a client's response curve to therapy. There has been some debate about what client therapy response curves should look like (Baldwin et al., 2009). The dose-response camp believes that the curve should be negatively accelerating (log-normal), indicating that the higher the number of sessions, the less effect each therapy dose or session should have. Lutz et al. (1999) described this negative accelerating curve as "lawful" (p. 571) and any individual client's response curve to therapy can be compared to the negative accelerating one as a standard (Baldwin et al., 2009). Inherent in this definition/conceptualization is the notion that the rate of change does not depend on the number of therapy sessions clients receive.

The GEL model predicts that those who receive lower doses of therapy are no more likely or less likely to achieve clinically significant change than those who receive high doses of therapy. Barkham et al. (2006) backed up this assertion and found that rates of clinically

significant change did not increase as session number increased. This model posits that if in any particular interval (e.g., 5 sessions of therapy), client improvement should be a linear curve as opposed to a negatively accelerating one. This model does not assume that rate of change is identical as session number increases; rather, it says that the rate of change depends on the number of sessions a client receives. Proponents of this model also argue that the negatively accelerating curve often found in research is an artifact of aggregating multiple clients together that are all changing at different rates and for different lengths in treatment. When constructing this curve, it appears that therapy is less effective as sessions increase because at the far end of the treatment number those who received therapy and rapidly changed have probably also terminated therapy leaving only those patients who change more slowly (Baldwin et al., 2009).

Survival Analysis for Clinically Significant Change Data

Hansen et al. (2002) discuss how clinically significant change can be statistically analyzed. They first acknowledge the binary nature of clinically significant change data: The client reached the criterion or they did not. They describe how probit regression can be used to "linearize" the binary variable and test linear patterns of independent variables and their ability to predict outcomes. This approach relies heavily on pre-therapy, post-therapy data to determine whether the event occurred. Researchers have argued that clinically significant change should not be analyzed only using a pre-post data approach, but rather, should be analyzed session-by-session (Anderson & Lambert, 2001; Kadera, Lambert, & Andrews, 1996). Analyzing data session-by-session allows for the event time to be directly known as opposed to being linearly interpreted from a probit or logistic regression (Hansen et al., 2002). They acknowledge that change does not occur in an equal distribution but most often in spurts (Tang & DeRubeis, 1999). They also outline that given the nature of this data, survival analysis is best suited because

it can handle both the event (binary) component and longitudinal data component (time) aspect of clinically significant change data. This also provides a better estimate of the number of sessions needed for clinical change to occur that pre-post data cannot accurately tell you (Hansen et al., 2002).

Very few studies have utilized survival analysis when analyzing clinically significant change data. Most notably is a study by Hansen and Lambert (2003) that utilized Kaplan-Meier survival analysis (a nonparametric type of survival analysis) to determine a median survival time (how many therapy sessions) it took for 50% of clients to reach clinically significant change (the event). They reported issues with not having enough people start in the clinical range that could obtain clinically significant change and had many clients that started already in the functional range; this limited their breakdown of results for some categories in their study. Their analyses focused on an employment assistance program (EAP), national health management organization (HMO), state community mental health clinic (CMHC), and local CMHC. Examining those who improved in some way, the EAP mean number of sessions needed for change across conditions ranged from seven to fifteen sessions. For the national HMO it ranged from eight to sixteen. The local CMHC ranged from nine to seventeen, and the state CMHC ranged from eleven to nineteen. They concluded that their findings were in concordance with previous findings that between 13 and 18 sessions are needed for 50% of clients to reach the criterion of clinically significant change (Hansen & Lambert, 2003).

Client Demographic Differences in Response to Therapy

One limitation of the previous survival analysis study by Hansen and Lambert (2003) is they did not consider client demographics at all in their study. They did not describe their sample, test any differences between different groups of clients, or test any demographic

variables for significant survival differences in their survival analyses. Lambert (1992) estimated the amount of variance in therapeutic improvement that can be attributed to different factors. Inside of this model, Lambert estimated that 40% of the variance can be attributed to "client variables and extra-therapeutic events" (Cooper, 2009, p. 60). It may be easy to focus on the therapist as the agent of change, but research has shown that it is the client who is heavily responsible for change in therapy (Bohart & Tallman, 1999; Duncan, Miller, & Sparks, 2004). Much of the research presented thus far in the discussion of clinically significant change ignores client factors in their conceptualization and/or analyses.

Much research has been undertaken to determine individual client differences in therapy. One recent study by Delgadillo, Moreea, and Lutz (2016) supported the notion that "different people respond differently to psychological therapy, based on their individual characteristics and circumstances" (p. 20). Notably, client self-reported disability status, unemployment, being younger (< 20 years old), and having severe psychosocial functional impairment were associated with persistence of depressive mood post-treatment. Further, possessing co-morbid depressive symptoms and holding the belief that therapy would not be beneficial in treating symptoms, contributed to persisting anxiety symptoms post-treatment (Delgadillo, Moreea, & Lutz, 2016). A recent study by Firth, Barkham, Kellett, and Saxon (2015) corroborate their findings. They found intake patient severity, patient unemployment, and treatment non-completion were detrimental to client outcomes. They also found a dose-effect curve that showed decreasing returns for clinical treatment when it extended beyond low-intensity treatment protocol (Firth et al., 2015; Richards & Whyte, 2009). They found a *U*-shaped curve that showed that benefits started to plateau between 6 and 8 sessions.

One final individual characteristic for consideration is the effect of trauma on counseling outcomes. There is "mounting clinical research [that] suggests that prior trauma histories—or stressful life events posttrauma—influence psychological outcomes and the likelihood that a trauma-related disorder will manifest" (Pimlett-Kubiak & Cortina, 2003, p. 529; see also Breslau, Chilcoat, Kessler, & Davis, 1999; Green et al., 2000; Kessler et al., 1995; King, King, Fairbank, Keane, & Adams, 1998). There is longstanding research (Brown & Harris, 1978; Dohrenwend & Dohrenwend, 1974) of the compounding effect that multiple "negative life events" can have. This research suggests that as the number of life stressors or victimizations increase, so, too, does psychopathology severity (Pimlett-Kubiak & Cortina, 2003, p. 529).

Characteristics of the Population Receiving Services

The present study will examine clinical outcomes observed in a telepsychology clinic that provides mental health services to an undeserved region of South Central Texas. This section will describe the issues in mental health disparities and the use of long-distance technologies to increase access to mental health services in this underserved area. This background will provide the context for the present study, the sample that will be studied, and the need to conduct outcome research in these clinical settings.

Texas has the largest rural-residing populations in the United States ("Growth in Urban Population," 2012), and it has the largest proportion of counties designated as Mental Health Professional Shortage Areas (MHPSAs; Trust for America's Health, 2014). The Brazos Valley has a total population of approximately 338,000. The largest county in this region is Brazos County, a metropolitan county with an estimated population of around 200,000. The remaining six surrounding counties are rural with populations ranging from approximately 13,500 to around 36,500 ("2010 Census: Population of Texas Counties," 2013).

In general, rural residents are considered a vulnerable population because they are more likely to endorse poorer health outcomes, have no health insurance, be diagnosed with a chronic health condition, and have less economic resources (lower socioeconomic status/living below the federal poverty level; Brossart et al., 2013). Rural residents are also more likely to need to travel greater distances to receive quality mental health care (Brossart et al., 2013). Women may be more at risk as they face unique problems in rural areas including limited social interaction, reduced occupational opportunities and consequent increase in poverty and/or issues in financial stability, and lack of access to adequate childcare. Residents with sensory loss face similar barriers and experience more depressive symptoms and lower health-related quality of life (Armstrong, Surya, Elliott, Brossart, & Burdine, 2016).

Minority group members in rural areas such as African American and Hispanic residents face even greater consequences and resulting health disparities compared to minority group members who live in urban areas (Brossart et al., 2013; Probst, Moore, Glover, & Samuels, 2004). In addition, they may face additional barriers such as experienced or perceived racism, various types of stigma, and acculturative stress (see Brown, Brody, & Stoneman, 2000; Kogan, Brody, Crawley, Logan, & Murry, 2007; Torres & Ong, 2010) along with possible culturally specific barriers such as fatalismo (Brossart et al., 2013; see Castillo & Caver, 2009). While there may be multiple available options in terms of how to address or mitigate these health disparity issues, one promising approach is the use of telepsychology services to address the mental health care needs of Brazos Valley residents.

Televideo services have been implemented successfully in rural areas to address lack of access to mental health professionals. The National Advisory Committee on Rural Health reported that 55% of the 3,075 rural counties in the country lacked practicing mental health

professionals including psychologists, psychiatrists, or social workers (National Advisory Committee on Rural Health, 2013; Nelson & Velasquez, 2011). Nelson and Velasquez (2011) discuss the needs of Kansas residents in that the majority of all counties in Kansas (100 out of 105) are designated as mental health professional shortage areas (full or partial designation). These authors highlight the work of the University of Kansas Center for Telemedicine and Telehealth (KUCTT) in implementing televideo in this area. This center has had positive outcomes including 1,130 psychology encounters across two years (2009 and 2010) and involving multiple psychologists (Nelson & Velasquez, 2011). Piloting telepsychology services in the Brazos Valley, McCord et al. (2011) describe their use of televideo to provide services to Brazos Valley residents. This project utilized remote sites in the region to provide services. The outcomes were generally positive with reports from both counselors and clients of being able to connect and form positive therapeutic relationships. Many clients in the area report that this project has made a difference in increasing access to mental health services there.

The delivery of services was provided by the Telehealth Counseling Clinic associated with the Texas A&M Health Science Center. In partnership with the Center for Community Health Development at the School of Public Health, the clinic was able to work collaboratively with the community for assessment of service needs, sharing of available resources, mobilizing community resources, and ultimately, having the goal of working with the community as opposed to coming in and working inside the community (McCord et al., 2011). More details about the clinic and its service delivery model are available in the literature (Chang, Frazier, & Elliott, 2013; Garney, McCord, Walsh, & Alaniz, 2016; Wendel, Brossart, Elliott, McCord, & Diaz, 2011).

In terms of measuring outcomes, research in this area found that residents typically had lower health-related quality of life that did not improve after receiving telehealth counseling; however, this could be due, in part, to co-occurring health conditions or chronic health conditions that would not be as impacted by therapy (Gonzalez & Brossart, 2015; Tarlow, McCord, Elliott, & Brossart, 2014). Residents in the region also report higher levels of mental health-related quality of life after receiving telehealth services (Gonzalez & Brossart, 2015; McCord et al., 2011; Tarlow, McCord, Elliott, & Brossart, 2014). Other outcomes have been reported including positive impacts on depression severity post-treatment (Gonzalez & Brossart, 2015; McCord et al., 2011), reduced symptom severity, decreased risk to self or others, increased well-being, and increased levels of functioning (Gonzalez & Brossart, 2015; as measured by the CORE-B outcome measure, CORE System Group, 1998).

Summary of Presented Literature

There is preliminary evidence showing the effectiveness of telepsychology services across the board in terms of client outcomes; so far, the research is encouraging (Hilty et al., 2013). While promising, there remains a need for additional research on the effectiveness of telepsychology services, specifically televideo services. The present study can potentially contribute to this literature in terms of the effectiveness of telepsychology in a real-world scenario, showing its strengths in helping to curb access issues in an underserved region, further explore client factors like gender, race, trauma history, or symptom severity and their potential impact on both client improvement and client deterioration, and provide both group and individual level outcomes. Further, overall trends in number of sessions needed for improvement can help clinics achieve developmental milestones and justify their place in institutions (institutalization) or continued grant funding. Outcome data can also inform clinic policies (e.g.,

building session limit policies based on data) as well as inform community stakeholders on the current status of one major form of mental health treatment in the region.

Assessment measures also play into how clinics and researchers are able to measure client outcomes over time. Outcome measures are sometimes limited in terms of what is accepted as evidence to grant funding agencies or institutions (e.g., colleges), and important decisions can depend on reliable, valid data for a specific measure with a specific population. The present study is also evaluating the effectiveness of the PHQ-9 and CORE-B outcome measures in capturing client responses to therapy over time within this region with these residents. If the measures are not well-suited to this endeavor, the assessment of client outcomes will be affected, and the decision to choose different outcome measures may be warranted based on the provided evidence.

There has been some debate regarding clients outcomes to therapy—both conceptually (how clients respond to therapy and achieve CSC; dose-effect model, GEL model) and visually (how these outcome curves should look over time). While clients may achieve CSC sometime within their treatment, it must not be understated the importance of including time as a factor in creating data as the binary nature of client response patterns is not enough to provide meaningful evidence. Practitioners for instance, want to know how many sessions until their clients will achieve CSC or reliable improvement.

Survival analytic approaches lend themselves well to answering research questions based on survival data surrounding client outcomes as it takes both elements (discrete and time) of the data into account. The present study will utilize survival analytic techniques and demonstrate how these results can be interpreted and applied to answer various research questions that are pertinent to clinics, practitioners, and researchers. Further, the analyses will delve deeper into

utilizing alternative statistical methodologies to elucidate the strengths of survival analytic approaches and the limitations of using other approaches.

In general, clients provide a survival curve for their individual treatment. Over a population, the overall survival curves can provide information about the community as a whole. Further, when looking at reliable deterioration, the clinic and its practitioners want to know the prevalence of clients who get worse after receiving therapy. Additionally, if individual factors play a role in determining whether someone is more likely to achieve reliable improvement versus reliable deterioration, this can be used potentially to address these further disparities to ensure that "at-risk" clients may be better served when providing future services.

Residents in this underserved region face considerable barriers to quality mental health care along with possessing many factors that negatively impact them, such as a chronic health condition, limited access to dependable transportation, and unemployment. Clinically significant change is one positive outcome that the TCC would want for their clients to achieve; however, reliable improvement is another well regarded outcome in terms of response to therapy. Given the criticisms of CSC, reliable improvement should not be seen as a second-rate outcome for clients, especially for residents who are already facing access barriers, have increased depression severity, and limited opportunities for medical and mental health treatment.

CHAPTER III

METHODS

Procedure

Participants included in the study were clients receiving services from the Telehealth Counseling Clinic (TCC) located in College Station, TX. The TCC serves residents from five underserved counties in the Brazos Valley in Texas. The most common referral sources for clients are from primary clinics, local health resource centers, and the local mental health authority, the MHMR Authority of the Brazos Valley. In order to receive services, clients must first complete a phone screening during which telepsychology services are described, demographic information is collected, and contact information is obtained from the client. During the screening, clients are also assessed for agency appropriateness with the exclusionary criteria being, 1) the client is actively psychotic and in need of antipsychotic medication, 2) are at a high enough risk for suicidality or homicidality that they cannot guarantee their safety before seeing a counselor, or 3) would be better served by an inpatient treatment facility such as in the case of substance abuse needing drug rehabilitation services. Clients are then either assigned to a waitlist or placed directly on a counselor's case load.

All clients must go to their remote county clinic for their intake appointment. During this intake appointment, they meet with a counselor using videoconferencing. Clients are then seen for future sessions either continuing over videoconferencing or using a telephone counseling modality. One reason for the use of telephone counseling is the barrier that transportation poses for many clients to drive to their remote county clinic. While the drive times are reduced considerably compared to driving into the micropolitan area of Bryan/College Station from their

rural location, there is still a great burden on clients to afford or find adequate transportation. In addition, telephone counseling services can be more practical for clients, specifically for those in the outlying areas. In comparison to other counseling clinics where there may be a pre-treatment phase consisting of the first three to four sessions to establish a treatment plan and intake evaluation report, counselors at the TCC complete these following a client's intake session. Additionally, length of therapy treatment is a collaborative decision between client and counselor while also taking into consideration the client's presenting concerns and the counselor's theoretical orientation.

The majority of TCC counselors come from Texas A&M University's doctoral counseling psychology program. During the first year of this program, students take counseling psychology courses covering counseling techniques, counseling theories, psychopathology, ethics, and multicultural counseling. They also complete a practicum at a psychological service training clinic under supervision of a licensed psychologist. After completing their first year in the program, students become eligible to apply to the TCC for a field practicum. At the TCC, they receive individual supervision along with group supervision, of which at least one semester provides overall training and an introduction to telepsychology. In addition, counselors are required to participate in training activities geared towards the community and population being served such as visiting the community or attending lectures/workshops on rural physical and mental health. Clinically, TCC counselors are trained in the scientist-practitioner model to be generalists (McCord, Saenz, Armstrong, & Elliott, 2015); they utilize a variety of treatment approaches including but not limited to, humanistic approaches, cognitive behavioral approaches, interpersonal approaches, and trauma-focused approaches. For more detailed information on the training model of the TCC, see McCord et al. (2015).

The Patient Health Questionnaire (PHQ; Kroenke, Spitzer, & Williams, 2001) and CORE-B Short Form (CORE System Group, 1998) are administered at each intake appointment. Clients then receive appropriate weekly counseling through videoconferencing or telephone by TCC counselors. On an ongoing basis, clients completed the CORE-B Short Form weekly and the PHQ-9 every two weeks at the time of their counseling sessions. Clients received videoconference counseling on a large high definition widescreen TV and a Cisco TelePresence teleconferencing unit, including a high definition camera and microphone. The high audio and video quality helps to avoid latency or lag in communication and enables real-time communication between the therapist and client. Clients receiving telephone counseling typically use their home or cell phones for sessions and were called from the clinic's phone in a private room.

Measures

Patient Health Questionnaire (PHQ). The Patient Health Questionnaire (PHQ; Kroenke et al., 2001) was developed as a screening tool originally designed to detect five common mental health disorders that present in primary care: depressive disorders (major and other), anxiety disorders (generalized anxiety and panic), alcohol use issues and other substance abuse problems (substance dependency/overuse), eating disorders (binging/purging and food restriction) and somatoform disorders. The diagnostic validity of the PHQ is comparable to the Primary Care Evaluation of Mental Disorders (PRIME-MD). It is a self-report measure with questions tied directly to diagnostic criteria according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). In addition, the PHQ asks questions relevant to counseling and research including those about experiencing trauma in the past and whether they were still experiencing symptoms surrounding any past traumatic events.

Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 (Kroenke et al., 2001) consists of the nine items assessing depression from the full PHQ measure and can be given on its own to assess depression diagnosis and severity over time. Clients complete the PHQ-9 every other counseling session to track changes in depression severity over time. First, the respondent is presented with this prompt: "Over the last 2 weeks, how often have you been bothered by any of the following problems?" Nine items follow that are each scored on a 4-point Likert-type scale from 0 (*Not at all*) to 3 (*Nearly every day*). Examples include: "Little interest or pleasure in doing things," "Feeling down, depressed, or hopeless," and "Trouble falling asleep, staying asleep, or sleeping too much." The scale is available in Appendix B. The PHQ-9 is appropriate for use in clinical settings and in surveys of community residents (Gilbody, Richards, Brealey, & Hewitt, 2007).

The primary validation study of the PHQ-9 had 3,000 patients originating from eight primary care clinics. Compared to diagnoses obtained from clinical interviews by mental health professionals, the measure's overall accuracy was 85%, its sensitivity was 75%, and its specificity was 90% (Spitzer, Kroenke, & Williams, 1999). The PHQ-9 has many strengths including limited time required for administration, acceptable levels of validity, and its widespread availability in other languages.

***Using community health surveys from the area*.** The Center for Community Health Development (CCHD) at Texas A&M University collaborated with community stakeholders in the Brazos Valley counties to conduct health assessments of residents to identify issues, circumstances, and needs that affect the health of residents (Burdine, Clark, Shea, Appiah, & Hollas, 2012). These community health surveys are conducted approximately every four years to obtain needed information that may inform policy and direct services for residents. Stakeholders

included measures of depression and health-related quality of life in the surveys. Depression was included because stakeholders wanted more detailed, specific information about the mental health needs of the region, and it is a frequent secondary complication of many health conditions. Psychologists participating in the development of the 2010 survey recommended the PHQ-9 because of its status as a well-validated, criterion-based instrument that could be used to obtain more accurate information about the rates of clinically significant depression in the region. Brossart et al. (2013) report results from these surveys in 2006 and 2010; based on the CESD-5 used in 2006, 17.9% of the sample reported depressive symptoms that merited mental health follow up, and in 2010, 10.9% of the sample reported depressive symptoms warranting further mental health investigation.

Previous studies of the PHQ-9 recommended using a cutoff total score of 10 ("yellow flag") to indicate that further follow up was needed. In a community sample, Gilbert et al. (2007) found that a cutoff score of 10 was as useful as the official scoring algorithm (that determines major or other depressive disorder) at detecting probable depression. For the present study, this cutoff score will be used and a clinical range will be those scoring 10 or greater (≥ 10) and a non-clinical range represents those scoring 9 of fewer (≤ 9).

Because the present sample resides in the Brazos Valley, a reliable change index can be calculated from data obtained from a larger, representative sample of residents throughout the communities in the region. PHQ-9 data from the 2010 Brazos Valley Health Survey includes respondents from the area served by the TCC and the PHQ-9 data obtained in the survey contains the full range of scores (0-27) and the naturally occurring variance of these scores. A reliable change index, in accordance with the procedures laid out by Jacobson and Truax (1991), was calculated for this sample as 5 points. This is fairly close to other studies that have calculated a

RCI for the PHQ-9 at 6 points (Delgadillo et al., 2013; Richards & Borglin, 2011). This reliable change index calculated for this study will be used to determine whether a client improved reliably, deteriorated reliably, and in establishing whether they reached clinically significant change.

The following client response patterns are possible for the PHQ-9:

- ***Clinically significant change.*** To achieve CSC, the client must have obtained a PHQ total score of 10 or greater at their intake session, see a decrease in their total score by at least 5 points, and have their most recent score equal 9 points or lower.
- ***Reliable Improvement.*** To achieve reliable improvement, the client must see a decrease in their total score by at least 5 points, regardless of whether they started in the clinical or non-clinical range, and the client must stay in the range they started in at their intake session.
- ***Reliable Deterioration.*** To achieve reliable deterioration, the client must see an increase in their total score by at least 5 points, regardless of whether they started in the clinical or non-clinical range.
- ***No Change.*** To achieve no change, the client must not see an increase or a decrease in their total score by at least 5 points, regardless of whether they started in the clinical or non-clinical range.

CORE-B Short Form. The CORE-B short form (CORE-SF/B; CORE System Group, 1998) was developed with the primary aim to provide a screening measure that could be used on a weekly basis. The scale includes 18 items divided into four major subscales; well-being (4 items), problems (6 items), functioning (6 items), and risk (1 self-directed risk item and 1 other-

directed risk item). Each item is scored on a 5-point Likert scale: Not at all (0), Only Occasionally (1), Sometimes (2), Often (3) and Most or all of the time (4). Subscale scores are calculated by averaging all items within them. The well-being subscale included items such as, "I have felt optimistic about my future" and "I have felt OK about myself". The problems subscale included items such as, "I have felt totally lacking in energy and enthusiasm" and "Tension and anxiety have prevented me from doing important things." Example functioning items include: "I have been able to do most things I needed to do to" and "Talking to people has felt too much for me" The risk subscale items were, "I made plans to end my life" and "I have threatened or intimidated another person". The scale is available in Appendix B. The total score (Global Distress Score) is an average of all 18 items. Subscale and total scores range between 0 and 4. Relevant items are also reverse-scored to reflect consistent interpretation that higher scores indicate worse indicators (higher symptoms or lower well-being) and lower scores indicate better indicators (lower risk or higher functioning).

Norms for clinical and non-clinical ranges were adapted from the parent scale, the CORE-OM to be used with the CORE-B (CORE System Group, 1998). Available data for the Brazos Valley was not available to determine cutoff scores; data from official normative studies reported in the technical manual (CORE System Group, 1986) were used. Cutoff scores are interpreted for each CORE-B subscale: well-being (above 1.37 for men and 1.77 for women), problems (above 1.44 for men and 1.62 for women), functioning (above 1.29 for men and 1.30 for women), and risk (above 0.43 for men and 0.31 for women; CORE System Group, 1998). In addition, a total score above 1.19 for men and 1.29 for women represents the clinical range when considering all items (CORE System Group, 1998). Another feature of the CORE-B Short Form

is that it includes the four well-being items directly from the CORE-OM, thus allowing for direct comparison to CORE-OM well-being domain scores.

The RCI had to be calculated also from normative data as available data on this population was not available. The RCI requires information on the spread of data (standard deviation, *SD*) and reliability information. It is not optimal to use reliability coefficients from another sample; however, it can be done to provide a close approximation. Barkham et al. (2001) gave the CORE-B to 2,700 participants in secondary care settings. This is a large normative sample and represented the full range of scores, including those who were distressed and not distressed. Standard deviations for men and women from this sample were used in the RCI calculations. Reliability information was obtained from Barkham et al. (2010) for the subscales and the total score. The total score is most reliable ($\alpha = .94$) and each subscale has acceptable reliability (Well-being: $\alpha = .76$; Problems: $\alpha = .89$; Functioning: $\alpha = .87$; Risk: $\alpha = .79$).

With both sources of information, a RCI was calculated for men and women on all subscales and the total score. The total score RCI did not differ between men and women and was .611 points. This is similar to a study by Stulz, Lutz, Leach, Lucock, and Barkham (2007) that also calculated the RCI for the total score as .63. Each subscale is slightly less reliable than the total score (less items—less reliability) and the spread of scores is slightly different, so the RCI's for each subscale vary from the .61 total score RCI. The Well-being subscale RCI was 1.29 (men) and 1.25 (women). The Problems subscale RCI was .80 (men) and .78 (women). The Functioning subscale RCI was 0.85 (men) and .80 (women), and the Risk subscale RCI was .95 (men) and .99 (women).

Each RCI can be slightly tweaked to accommodate the nature of the scale and allowable change on that scale. Because each subscale score is calculated by averaging the items in the

scale, the minimum amount of change in someone's score is dependent on the number of items in the score. The Well-being subscale has 4 items, so the minimum amount of change is .25. The RCI for men on Well-being was 1.29, a value that is closer to 1.25 than 1.50 the next possible value. This score will be adjusted to 1.25 to better approximate the true RCI level as putting it at 1.29 and requiring 1.50 amount of change may be slightly too stringent. The RCI for women on Well-being (1.25) is already at an appropriate interval (.25). The Problems and Functioning subscales contain 6 items; thus, the minimum amount of change would be .167. The Problems subscale RCI's (.80 and .78) are already at an appropriate level that would allow clients to achieve it in a reasonable reduction in scores (i.e., .667 < RCI < .833). The Functioning subscale RCI for men is .85 and was adjusted to .833 accordingly. The Risk subscale RCI's (.95 and .99) were kept requiring the client to at least change by 1 subscale point. For the total score, the minimum amount of change is .056. The closest amount of change is .616, so the RCI did not need to be adjusted.

The following client response patterns are possible on the CORE-B (total and subscales):

- ***Clinically significant change.*** To achieve CSC, the client must have obtained a CORE-B score that was greater than the established cutoff for that scale/subscale at their intake session, see a decrease in score by at least the RCI for that scale/subscale, and have their most recent score fall within the non-clinical range for that scale/subscale.
- ***Reliable Improvement.*** To achieve reliable improvement, the client must see a decrease in their CORE-B score by at least the RCI for that scale/subscale, regardless of whether they started in the clinical or non-clinical range for that

scale/subscale, and the client must stay in the range they started in at their intake session.

- ***Reliable Deterioration.*** To achieve reliable deterioration, the client must see an increase in their CORE-B score by at least the RCI for that scale/subscale, regardless of whether they started in the clinical or non-clinical range for that scale/subscale.
- ***No Change.*** To achieve no change, the client must not see an increase or a decrease in their score by at least the RCI for that scale/subscale, regardless of whether they started in the clinical or non-clinical range.

Participants

Participants included in the study were all former or current clients with the Telehealth Counseling Clinic. Data collected for clinical service purposes was de-identified and approved for research purposes by the Texas A&M University Institutional Review Board. Clients were given general consent forms for de-identified research at the start of treatment, and it was made aware to them that clinical services were not contingent on their consent to participate. All available participants will be utilized in this study that have been maintained in an IRB-approved research database. Not all of the data will be available for use in all analyses, but the breakdown of data and included participants for each analysis will be discussed in the results section.

Research Questions

The present study aims to answer research questions that a typical clinician or clinic would ask when possessing client outcome survival data. In particular, clinics may be especially constrained by organizational demands, lack of resources, or reporting requirements to third parties (i.e., grant funding organizations, college or university deans, etc.). Analyzing this type of

data can help clinics answer specific research questions both regarding their clients and client outcomes but also to help inform clinic policies or provide information to grant organizations. The present study will examine the following research questions:

(1) How many telehealth counseling sessions did clients receive on average until they reached clinically significant change or reliable improvement on the PHQ-9 or the CORE-B (total and subscale scores)?

(2) How did the client responses to therapy (CSC, reliable improvement, reliable deterioration, no change) present in this population receiving telehealth counseling services?

(3) Are there any recognizable patterns when comparing the outcomes of the PHQ-9 and the CORE-B (total and subscale scores)? Are the results concordant or discordant across the two measures?

(4) What individual client factors impact response to telehealth counseling?

(5) Are there any major differences between the survival analysis results and results from other, alternative statistical analyses?

Statistical Analyses

Demographic information including race, federal poverty level, age, gender, and insurance status were obtained for each participant. In addition, relevant clinical data were obtained from intake forms, clinical assessments, and review of client files. Descriptive statistics of the general sample will be provided. A series of survival analyses will be conducted and illustrated to answer the above research questions. Modeling events on both measures simultaneously is not feasible given current methodology, so each measure will have its own survival analysis series. In addition, once a client has reached an event in one analysis does not

mean they are excluded from reaching a different event in another analysis. This includes all six scales on the CORE-B, meaning that these events will be examined independently from one another. As opposed to comparing pre-post therapy data, the determination of whether an event occurred will be calculated after each session. Subsequent counseling session scores will be compared to scores obtained during the client's intake session.

To answer the first research question, Kaplan-Meier survival curves will be used to plot the overall survival for clients who achieve CSC or reliable improvement. Following, Kaplan-Meier survival curves will be plotted for each response pattern subsequently (i.e., first looking at CSC, then if they did not achieve CSC, did they achieve RI, etc.). Comparisons in survival curves between measures will be made by overlaying survival curves and visually analyzing any differences. After plotting these Kaplan-Meier curves to get a general sense of client outcomes, Cox proportional hazard models will be fitted to estimate the impact of independent variables (client factors; covariates) on clients' survival. Last, logistic regression models and linear growth curve models will be fitted to some of the previous analyses/data to explore any congruence or incongruence between the results. All survival analyses and logistic regressions will be conducted using Stata 14.1. Linear growth curve models will be fitted with the HLM 7 software package.

Key Concepts in Survival Analysis

Appendix A presents vital information and introduces key topics in conceptualizing survival analysis. Sections presented below use this as a basis for interpreting the analyses of the study.

Survival—the survival function. The first topic needed is how survival analysis combines events experienced by participants and time. Proportion formulae can be utilized in

estimating these proportions only when cases are complete, meaning that everyone experiences an event inside of the time interval of the study. When observations are censored, different analyses are needed; the Kaplan-Meier estimator being the most common. This is a nonparametric (i.e., not requiring an assumption of the underlying distribution) estimator used to estimate the *survival function* and plot the survival curve. Both numerically and visually, these curves/analyses provide information regarding survival of participants, including counts of individuals and trends of survival at each of the time points across the study interval. The Kaplan-Meier estimator (along with other survival analysis methods) also has the advantage of handling incomplete cases, meaning that individuals do not need to have data points at every time interval in the study as long as they have a baseline measurement and one additional data point. These concepts can be used to answer the first three research questions. Specifically, survival curves provide information regarding median survival, or the point in time in which we can expect 50% of clients to experience the event. Graphically, survival curves can be drawn together to compare how they are alike or different either within the same event (CSC) on different measures (PHQ, CORE-B Total) or across different events (CSC, RI, RD) on the same measure (PHQ).

Hazard—the hazard function. The last survival concept necessary is the hazard. Graphing the survival curves over time portrays overall trends, or in other words, what percentage of clients can one expect to have experienced the event at any particular time point given that they have not previously experienced the event. The survival function does not provide information regarding the probability that one individual will experience the event in the next time interval; this is conceptualized as the *hazard* (most often termed the hazard rate as further explained). The *hazard rate* is "a numerical quantity operationalizing the instantaneous

risk of the … event occurring at a particular time point (e.g., Session 2) to a particular client, given that the event for that client has not yet occurred" (Corning & Malofeeva, 2004, p. 356). In each time interval, if an individual has not experienced the event in the previous time interval, they are still "at-risk" to experience the event in the current time interval and belong to the *risk set*. The hazard rate then is "the proportion of the risk set that is expected to experience the event at a given time" (Corning & Malofeeva, 2004, p. 356). Because the hazard rate is a measure of instantaneous risk (i.e., the risk in one instant—session), it must be calculated at every time point. One conceptual point inherent in survival analysis made by Corning and Malofeeva (2004) is the acknowledgment that an individual who experiences the event after session one or session two is different than an individual who experiences the event later in therapy such as at session 10 or session 12. Since the hazard rate can be calculated at each time point separately, it is thus allowed to vary over time—over a course of therapy; a graphical depiction of the *hazard rate* at every time point is called the *hazard function* (or *cumulative hazard function*; Corning & Malofeeva, 2004). Given that both the survival function and hazard function are estimated given the same sample data, they are arithmetically linked to one another (explained further in Appendix A). The interest in the hazard function comes from the fact that it describes the "underlying survival process being studied" (Hosmer, Lemeshow, & May, 2008, p. 64). In other words, "if one is able to specify the hazard function, then the cumulative hazard, and hence, the survival function are also specified" (Hosmer et al., 2008, p. 64). Given this, the hazard function became a likely choice to create a statistical model with.

The goal of creating a statistical model around the hazard function was to "characterize the hazard function, not only as a function of time but also as a function of subject and other study covariates" (Hosmer et al., 2008, p. 64). The hazard function can be conceptualized as the

dependent variable in a regression framework and is predicted by a term consisting of information regarding time (conceptualized as an intercept, α, indicating how the hazard function changes as a function of time) and terms regarding the impact of study covariates (conceptualized as beta weights, ß's, indicating how the hazard function changes as a function of covariates). This is most commonly achieved through the use of a semiparametric analysis called the Cox proportional hazards model, or Cox model. The model conceptualizes differences between individuals as a ratio of their hazard functions, as explained by the two previously described functions, α and ß's. The α term is called the *baseline hazard function*, and the ß terms represent the differences in the hazard function between individuals depending on values of the covariates. The baseline hazard function (i.e., the probability distribution of the survival times) does not need to be explicitly specified; the analysis (estimation process) relies solely on the order in which events occur, and not the exact times they occur; this is what makes it a semiparametric statistical model (from Appendix A). Because the Cox model is conceptualized as a ratio, the α terms cancel out. What is left then are the ß terms for one level on the covariate (x_1) over the ß terms for the other level (x_0), for instance on a dummy covariate such as rural and urban; this is known as a hazard ratio (Hosmer et al., 2008). The interpretation of the hazard ratio is in terms of "relative risk" for the x_1 covariate group compared to the x_0 covariate group (Hosmer et al., 2008). For instance, if a hazard ratio for rural clients was 3, the interpretation would be, "rural clients are experiencing the event at three times the rate compared to urban clients."

CHAPTER IV

RESULTS

Descriptive Statistics of the Sample

The starting sample (*n* = 413) consisted of all clients who attended a session at the Telehealth Counseling Clinic and whose information were entered into the clinic's research database between September 24th, 2009 and December 21st, 2016. The survival analyses used in this study required participants to have attended at least two sessions. The clients who only attended one session did not contribute any information to the analyses because their measurements only include their baseline scores with no followup data. 258 clients had only one data point available, usually their intake session, and were excluded at this point from the sample. The remaining sample consisted of 155 clients who attended therapy at least twice and had two scores on at least one outcome measure. Given missing data, each analysis varied somewhat from each other based on what data was available. The average age of clients was 41.22 years (*SD* = 13.72, Min. = 19, Max. = 77). The majority of the sample identified as female (*n* = 107, 69.03%; male, *n* = 39, 25.16%; missing information, *n* = 9, 5.81%). The most common race identified by clients was White (*n* = 85, 54.84%) followed by African-American (*n* = 34, 21.94%), Hispanic (*n* = 21, 13.55%), Asian (*n* = 7, 4.52%), and bi-racial (*n* = 2, 1.29%) with 6 (3.87%) unidentified.

Most clients possessed no health insurance (*n* = 83, 53.55%), and 26 clients were on either Medicare (*n* = 16, 10.32%) or Medicaid (*n* = 10, 6.45%). 27 clients (17.42%) possessed private health insurance, and 19 clients (12.26%) did not have health insurance information available. Over half of the clients (*n* = 81, 52.26%) were below 100% of the federal poverty level

(FPL) guidelines for 2017. The rest of the clients ranged from 133% to 400% FPL (100% FPL and above: *n* = 39, 25.16%), and 35 clients (22.58%) did not have FPL information.

Client trauma history was also collected: the PHQ items 12j, 13, and 13 (write-in) were used as indicators. Item 12j reads, "*In the last 4 weeks, how much have you been bothered by any of the following problems?...Thinking or dreaming about something terrible that happened to you in the past – like your house being destroyed, a severe accident, being hit or assaulted, or being forced to commit a sexual act*". Item 13 reads, "*In the last year, have you been hit, slapped, kicked or otherwise physically hurt by someone, or has anyone forced you to have an unwanted sexual act?*" Item 13 (write-in) reads, "*History of Abuse (Child, Spouse, Physical, Sexual)* with a blank for answers to be written in. A trauma history indicator was created from a positive result to any of the three items; a score of "1" or "2" on item 12j, a "Yes" on item 13, or a written in answer on item 13 (write-in) with the value "1" representing trauma history and "0" representing no endorsement of any trauma history. 92 clients (59.35%) reported a trauma history, 40 clients (25.81%) did not report a trauma history, and 23 clients (14.84%) were missing information about their trauma history.

A series of analyses were conducted to compare differences between individuals who were kept in the study (*n* = 155; analysis sample) to those who were excluded (*n* = 258; excluded sample) based on not having enough data available (i.e., having only one data point available). Of the 258 clients who only had one data point available, 255 of them had demographic information available. The mean age of clients in the excluded sample was similar (*M* = 41 years, *SD* = 13.21, Min. = 18, Max = 78), and the difference in age of the samples was not statistically significant, $t(390) = 0.16$, $p = 87$. The majority of clients in the excluded sample identified as female (*n* = 181, 70.98%; male, *n* = 65, 25.49%; missing information, *n* = 9, 3.52%), and there

was no statistically significant difference between these two samples on sex, $\chi^2(1) = 1.42$, $p = .49$. Race identification looked similar as well with most clients identifying as White or Caucasian ($n = 171$, 67.06%) followed by African-American ($n = 38$, 14.90%), Hispanic ($n = 35$, 13.73%), Asian ($n = 1$, 0.40%), and bi-racial ($n = 1$, 0.40%) with 9 (3.53%) clients without racial identity information. There was a statistically significant difference between the two samples on client-identified race, $\chi^2(5) = 14.20$, $p = .014$, most likely due to the much higher number of clients who identified as White in the excluded sample ($n = 171$) compared to the analysis sample ($n = 83$).

Health insurance status for those in the excluded sample was similar to clients in the analysis sample with the exception that 123 clients did not have insurance information available compared to 19 clients in the analysis sample. Like the analysis sample, most clients had no health insurance ($n = 78$, 59.09%) followed by private health insurance ($n = 30$, 22.73%) and Medicare/Medicaid (Medicare, $n = 6$, 4.55%; Medicaid, $n = 18$, 13.64%). There was not a statistically significant difference on health insurance status between the analysis sample and excluded sample, $\chi^2(3) = 7.01$, $p = .072$. The vast majority of the excluded sample (n = 217, 85.10%) did not have federal poverty level information available compared to 33 in the analysis sample. Of the 38 clients with FPL information available in the excluded sample, the majority were <100% FPL ($n = 30$, 78.95%), and the rest were above 100% FPL ($n = 8$, 21.05%). There was not a statistically significant difference on FPL between the two samples, $\chi^2(7) = 6.20$, $p = .52$. In the excluded sample, over half of clients ($n = 130$, 50.98%) reported a trauma history, and 77 clients (30.20%) did not report a trauma history; 48 clients (18.82%) had missing information about their trauma history. For self-reported trauma history in the two samples, there was not a statistically significant difference between them, $\chi^2(2) = 3.09$, $p = .21$.

Overall Improvement: Joint CSC/RI Outcomes

Overall improvement can be conceptualized as looking at individuals who have achieved reliable improvement (RI) regardless of their starting or ending point. This does not have the stricter requirement of CSC that requires movement from the clinical to normal sample by the end of treatment. Because of this leniency, it is the best representation of overall counseling improvement. In this analysis, our event of interest is a positive one. The survival function (S(*t*)) estimates the probability that an individual survives longer than a certain time point, *t* (Kleinbaum & Klein, 2012); this is usually what you want when the event is something that an individual wants to avoid (i.e., you want to survive longer). In this case, we want to see the estimate depicting the probability that people experienced the event by *t* (i.e., you want to experience the event sooner). This is called the failure function and is represented by the complement of the survival function (i.e., 1 - S(*t*)). The failure function, in this context, is a misnomer as we are looking at positive outcomes from therapy. In more general terms, this is known as the distribution function and provides information about the events experienced by participants over time. For this study, this will be referred to as the *success* function to better represent the valence of events being modeled.

Our sample consists of all clients who received services at the Telehealth Counseling Clinic, and a part of their services involves receiving some in-person or group psychotherapy, usually in conjunction with telehealth services. One key question then is: "Do clients who received telehealth services differed in their outcomes compared to clients who received in-person services in addition to telehealth services?" The joint CSC/RI outcome is the most encompassing outcome, and thus, provided a broad inference on their comparability.

Patient Health Questionnaire. There were 111 clients included in this analysis, and 64 clients achieved joint CSC/RI over their course of therapy. Of those 64 joint CSC/RI events (successes), 52 clients only received telehealth services, and 12 clients received a combination of services. The survival (or success) functions can be graphed simultaneously by the two groups to see if they compare. A formal test of their comparison can be accomplished through a log-rank test. For the telehealth-only group, 52 joint CSC/RI events (successes) were observed and 48.25 events were expected, and for the combination group, 12 events were observed and 15.75 events were expected. With 1 *df*, this gives us, $\chi^2(1) = 1.46$, $p = .23$. This provides us with evidence that we can analyze these two groups together as their survival curves (and functions) are similar. Figure 1 depicts the success curve (a direct plot of the success function for the total sample) for the PHQ with joint CSC/RI as the event. A common statistic obtained from the survival/success function is the median survival time; this is the value of *t* where survival is 50% (probability of .5). Given the complementary nature of survival and success, median survival is equivalent to median success. The estimated median survival is 11 sessions with *SE* = 2.39 and 95% CI [7, 19]. From this, we can conclude for our sample that around 11 sessions were necessary before seeing reliable improvement on the PHQ. It is important to remember that the median survival time is a point estimate of the true population parameter value. However, it would not be unreasonable for us to expect around half of our future clients to experience reliable improvement on the PHQ some time before 20 sessons.

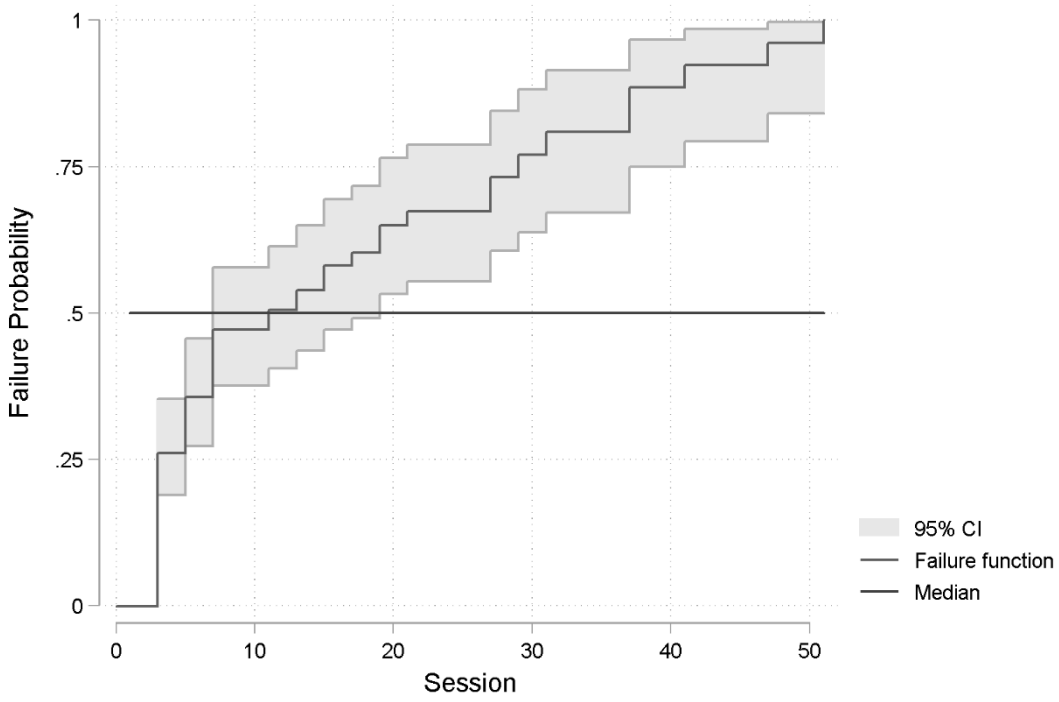

Figure 1.
Success Curve for the PHQ with Joint CSC/RI (event) with 95% CI's.

CORE-B. Similar tests to the PHQ were conducted to test the comparability of the two telehealth services groups on all scales of the CORE-B. All log-rank tests were statistically insignificant, and the combined sample will be used from here on out. Figure 2 depicts the success curves for the CORE-B total and subscale scores with joint CSC/RI as the event. For the total score, the median survival was 4 sessions with *SE* = .54, 95% CI [3, 6]. The risk subscale did not reach a median survival time and could not be estimated. The function subscale had a median survival of 6 sessions with *SE* = .89, 95% CI [5, 8]. The wellbeing subscale had a median survival of 7 sessions with *SE* = .75, 95% CI [5, 11]. The problem subscale had a median

survival of 4 sessions with *SE* = .60, 95% CI [4, 6]. Overall, clients improved on the CORE-B within 4 to 7 sessions of therapy.

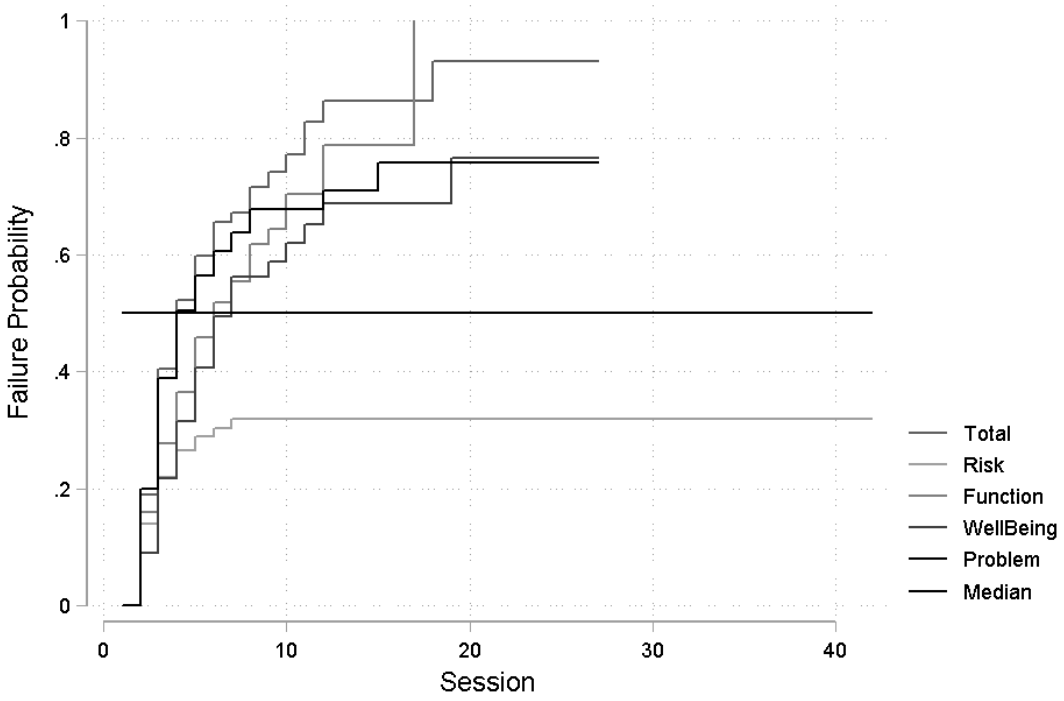

Figure 2.
Success Curves for the CORE-B with Joint CSC/RI (event).

Client Response Patterns to Therapy

Patient Health Questionnaire. There were 111 clients total who were analyzed on the PHQ. Out of these, 23 clients achieved CSC by the end of their treatment. Over half of them achieved CSC following their third therapy session. Out of the 88 clients remaining, 43 clients achieved reliable improvement. Their median survival time was 13 sessions with *SE* = 4.35, 95%

CI [7, 27]. The RI estimate was less precise than the joint CSC/RI analysis given its higher standard error. Of the 45 remaining clients, 13 experienced reliable deterioration. This left 32 clients who experienced no reliable change during their time in therapy. Figure 3 depicts the client response pattern success curves for the PHQ.

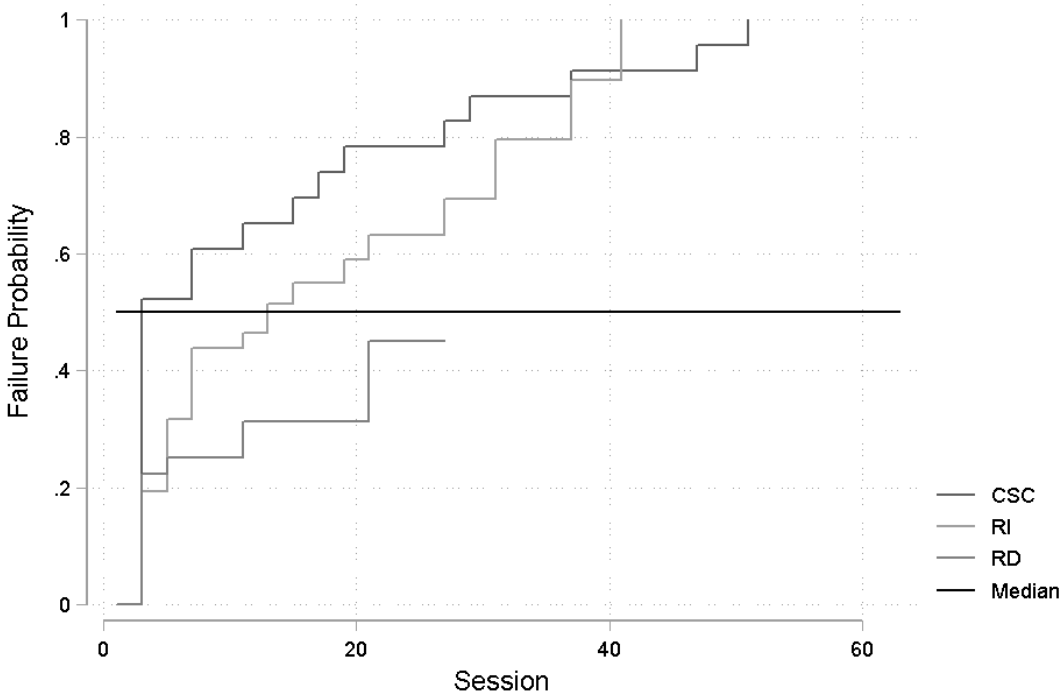

Figure 3.
Client Response Pattern Success Curves for the PHQ.

CORE-B. Client response patterns were assessed first by each subscale individually.

Total. 100 clients were analyzed on the CORE-B total scale and all subscales. 13 clients achieved CSC with a median survival of 2 sessions with *SE* = .18, 95% [2, 3]. Out of the 87

clients without CSC, 54 achieved RI with a median survival of 5 sessions with *SE* = .63, 95% [4, 7]. And of the remaining 33 clients, 5 clients experienced RD; however, the median survival time was indeterminate. 28 clients did not achieve reliable change on the CORE-B total scale in therapy. For the total score, CSC, RI, and RD patterns are displayed in Figure 4.

Risk. No clients achieved CSC on the risk subscale. Out of the 100, 29 clients experienced RI, and a median survival time could not be calculated. Of the 71 remaining subjects, 11 experienced RD and an indeterminate median survival. 60 clients did not achieve reliable change in their risk scores while in therapy. Figure 5 depicts the success curves for the risk subscale.

Function. 14 clients achieved CSC with a median survival of 2 sessions with *SE* = .20, 95% [2, 3]. Out of the remaining 86 clients, 41 achieved RI with a median survival of 8 sessions with *SE* = 1.38, 95% [5, 10]. 45 clients remained and 11 of them experienced RD; the median survival was 12 sessions with *SE* = .72. At 12 sessions, all clients were either censored or had experienced the event. This does not allow creating a proper confidence interval besides putting an estimate of the 95% lower bound at 12 sessions. 34 clients experienced no reliable change on the function subscale. Figure 6 depicts the success curves for the function subscale.

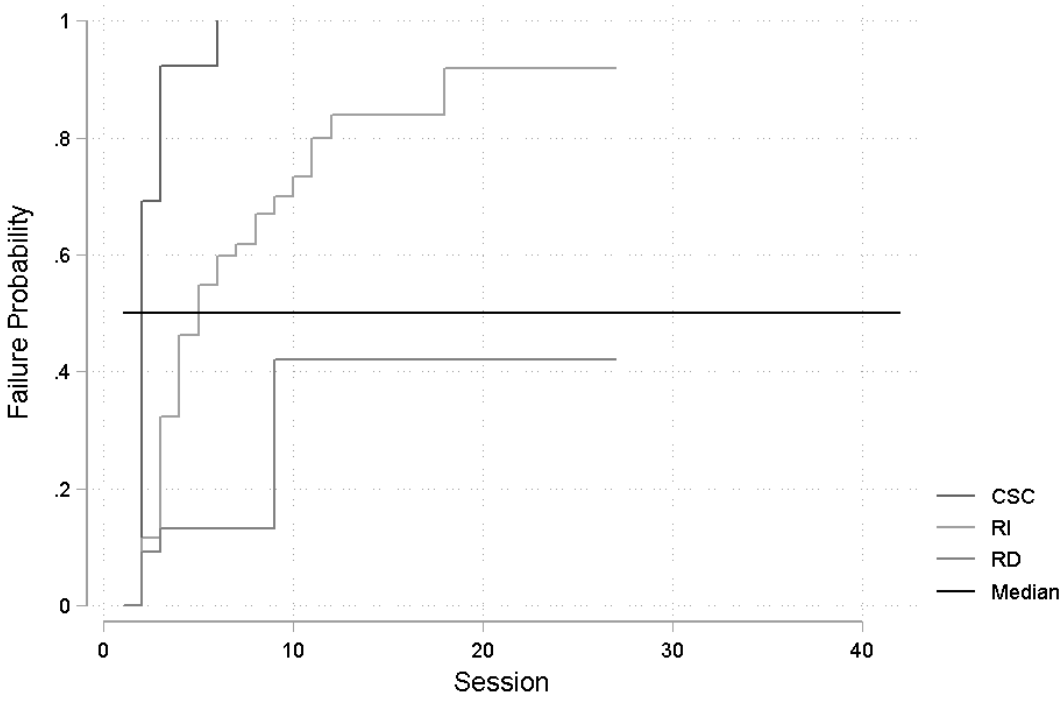

Figure 4.
Client Response Pattern Success Curves for the CORE-B Total Scale.

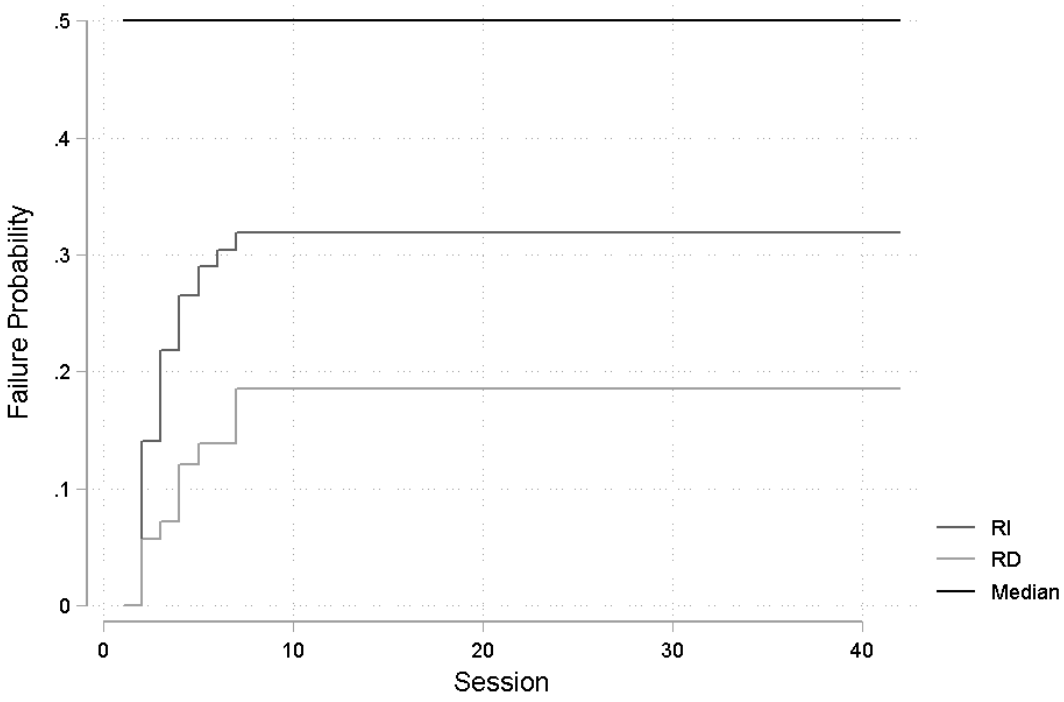

Figure 5.
Client Response Pattern Success Curves for the CORE-B Risk Subscale.

WellBeing. 14 clients achieved CSC with a median survival of 2 sessions with *SE* = .23, 95% [2, 3]. Of the 86 remaining clients, 36 achieved RI with a median survival of 9 sessions with *SE* = 1.83, 95% [6, 19]. Of the remaining 50 clients, 6 clients experienced RD with an indeterminate median survival. 44 clients experienced no reliable change on the wellbeing subscale. Figure 7 depicts the success curves for the wellbeing subscale.

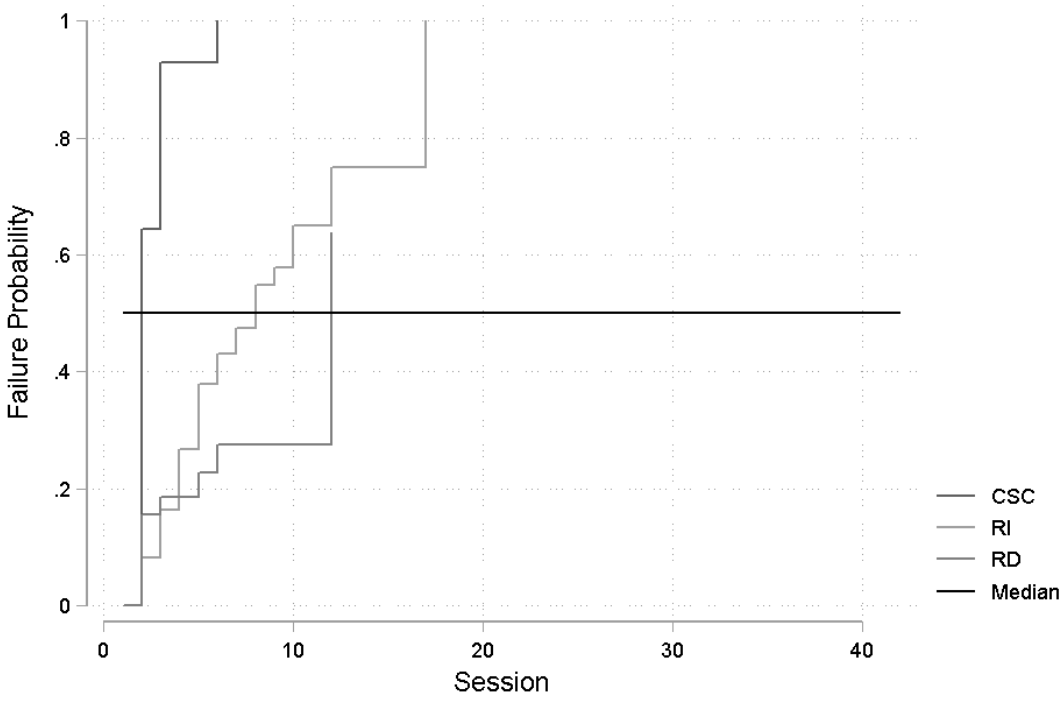

Figure 6.
Client Response Pattern Success Curves for the CORE-B Function Subscale.

Problem. 12 clients achieved CSC with a median survival of 2 sessions with *SE* = .17, 95% [2, -]. All clients experienced CSC at either 2 or 3 sessions, which does not allow for an accurate calculation of the 95% confidence interval except for a lower bound at 2 sessions. Of the 88 remaining clients, 49 achieved RI with a median survival of 5 sessions with *SE* = 1.12, 95% [4, 8]. Of the remaining 39 clients, 9 clients experienced RD with an indeterminate median survival. 30 clients experienced no reliable change on the problem subscale. Figure 8 depicts the success curves for the problem subscale.

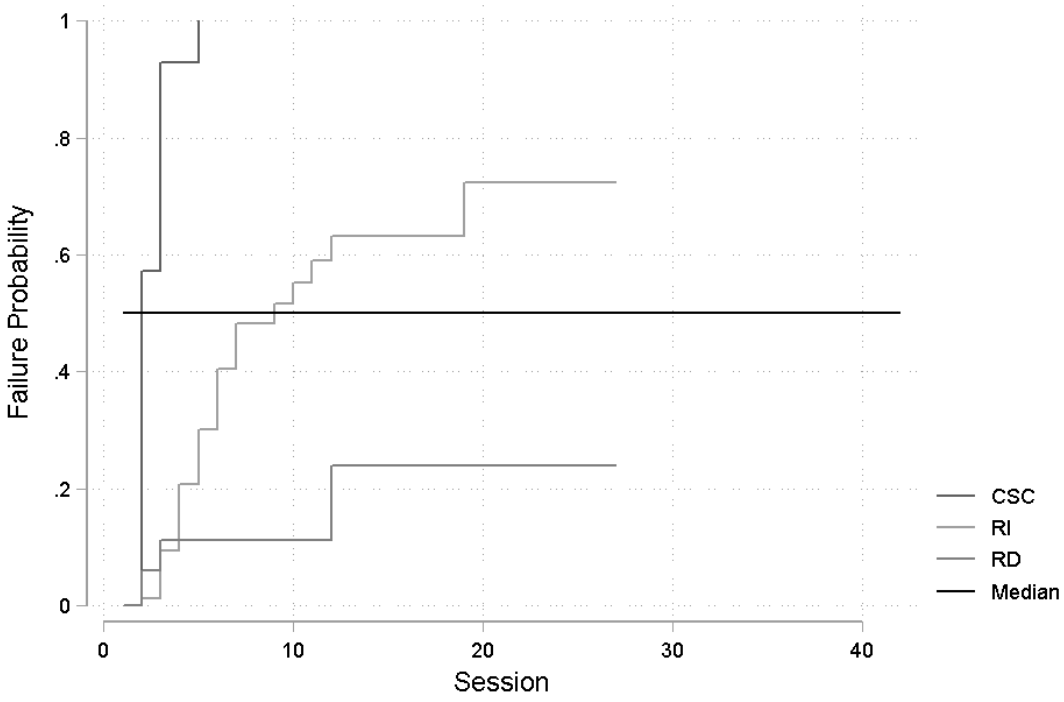

Figure 7.
Client Response Pattern Success Curves for the CORE-B WellBeing Subscale.

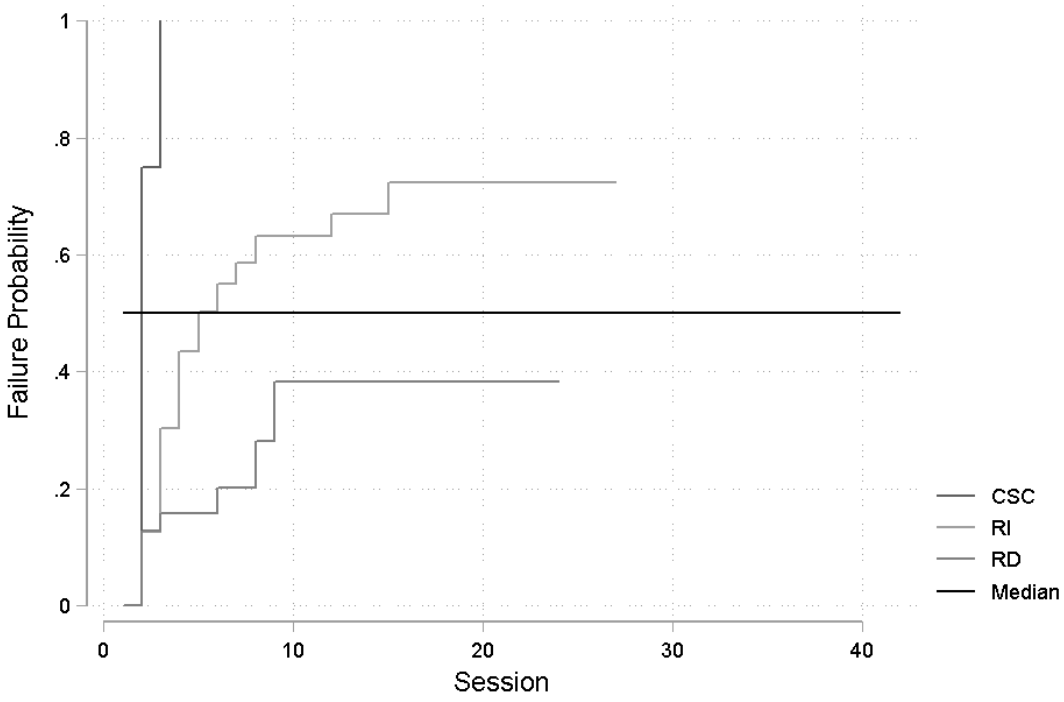

Figure 8.
Client Response Pattern Success Curves for the CORE-B Problem Subscale.

Summary. Overall improvement and client response patterns to therapy are summarized in Table 1. About 60% of clients improved on the PHQ, and two-thirds of clients improved on the CORE-B Total. The CORE-B Problem subscale also showed around 60% improvement. CORE-B Function and CORE-B showed slightly lower rates of overall improvement at 55% and 50% respectively. CORE-B Risk had the lowest improvement at 29%. See Table 1 for further information regarding client responses to therapy for each scale/subscale.

Table 1.

Summary of Joint CSC/RI and Client Response Patterns for all Scales/Subscales.

	Telehealth Only		Telehealth and In-Person/Group			
	CSC/RI Observed Events, *N*	CSC/RI Expected Events, *N*	CSC/RI Observed Events, *N*	CSC/RI Expected Events, *N*	χ^2	Total *N*
PHQ/PHQ-9	54	50.06	12	15.94	1.59	66
CORE-B Total	53	51.94	9	10.06	0.17	62
CORE-B Risk	26	22.86	2	5.14	2.57	28
CORE-B Function	46	42.83	6	9.17	1.53	52
CORE-B WellBeing	38	38.82	8	7.18	0.13	46
CORE-B Problem	49	48.80	9	9.20	0.01	58
	CSC/RI	CSC	RI	RD	NC	Total
Combined Sample	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N*
PHQ/PHQ-9	66 (59.45%)	23 (20.72%)	43 (38.73%)	13 (11.71%)	32 (28.83%)	111
CORE-B Total	67 (67%)	13 (13%)	54 (54%)	5 (5%)	28 (28%)	100
CORE-B Risk	29 (29%)	-	29 (29%)	11 (11%)	60 (60%)	100
CORE-B Function	55 (55%)	14 (14%)	41 (41%)	11 (11%)	34 (34%)	100
CORE-B WellBeing	50 (50%)	14 (14%)	36 (36%)	6 (6%)	44 (44%)	100
CORE-B Problem	61 (61%)	12 (12%)	49 (49%)	9 (9%)	30 (30%)	100

Note. χ^2 tests are 1 *df*. * = < .05.

Client Response Patterns Across Outcome Measures

To compare across scales/measures, success probability curves were overlayed with session number aligned to allow for direct comparison. Comparisons were made across all available scales for CSC, RI, and RD. Samples for CORE-B and PHQ were not combined in this

process, so inferences represent clients who possessed data on one measure or more than one measure.

Clinically significant change. With few clients achieving CSC across all measures, their success curves are less descriptive or informative. However, preliminary inferences can be made. Clients who achieved CSC on the CORE-B scales tended to do so in the first few sessions. This is a positive outcome in that we can clearly see that starting therapy for these clients produces a measurable decrease in symptoms and increase in well-being. One drawback however, information about client CSC change occurring later on in therapy or occurring after an initial CSC is not preserved due to clients being dropped from the analyses after reaching the event. Clients achieved CSC on the PHQ around three sessions, very similar to the CORE-B scales. With more variance regarding when CSC occurred, there is more information about CSC after four sessions. The majority of clients achieved CSC on the PHQ before session 20. Figure 9 presents the success curves for CSC on all measures.

Reliable improvement. Improvement was seen around sessions five, nine, and 13. Half of the clients saw improvement on two scales around session 5; improvement was observed in the CORE-B Total and CORE-B Problem. The CORE-B Total and CORE-B Problem scales had a similar success curve giving evidence that symptom reduction in the CORE-B Total may be from improvements in their reported problems. It takes a few more sessions (session 9) until they see improvement in their functioning and wellbeing. Minimal data was available on reliable improvement in risk; however, these data suggest that risk tended to improve in the first seven sessions. This is potentially related to the reduction in problems reported around this same time in therapy. Last, PHQ scores reliably improved around session 13. Figure 10 presents the success curves for RI on all measures.

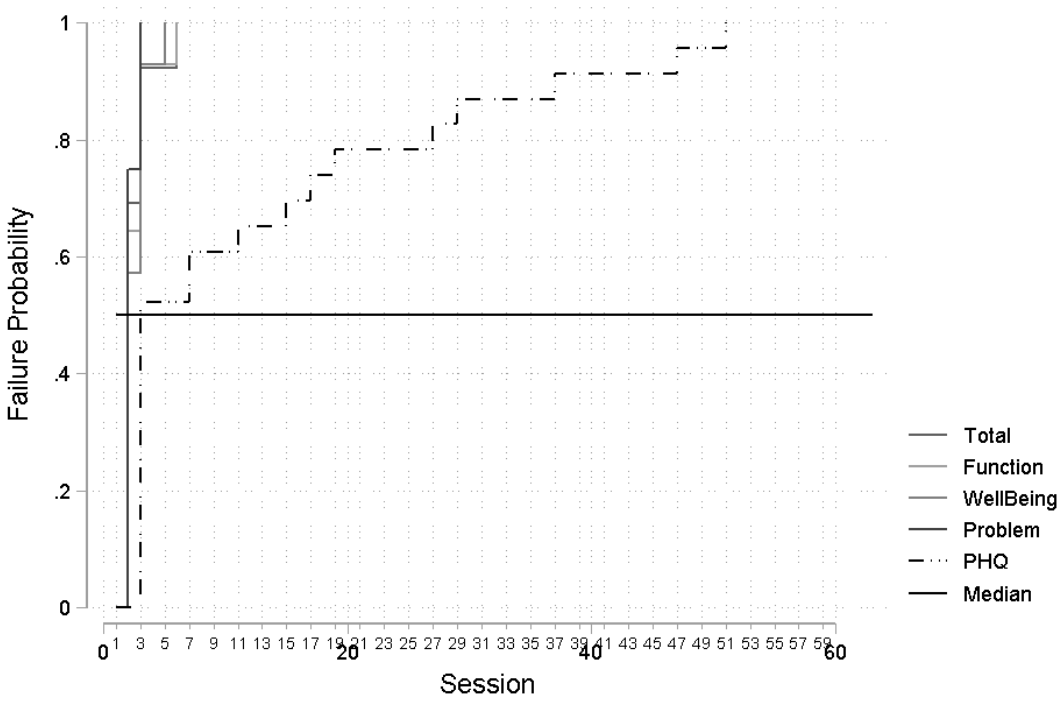

Figure 9.
Client Response Pattern Success Curves for CSC Across Scales.

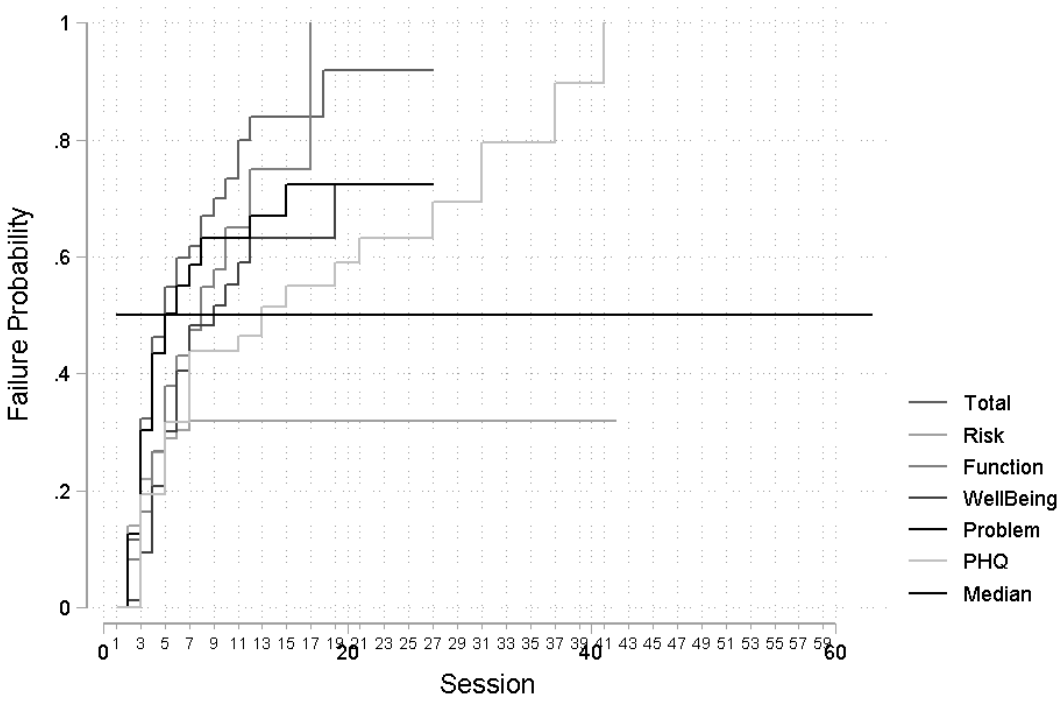

Figure 10.
Client Response Pattern Success Curves for RI Across Scales.

Reliable deterioration. There were very few clients who achieved reliable deterioration; the majority of clients who did not achieve one of the previous events were most likely to stay around where they started instead of getting worse through the course of therapy. Only one scale (CORE-B Function) possessed a median survival time (12 sessions). In addition, the various scales' success curves look different with only some portions overlapping. This means that there was not one point in time (i.e., start of therapy, session 1-4, or end of therapy, sessions 15-20) where clients were more likely to achieve reliable deterioration. It appears that clients who do deteriorate over the course of therapy do so at different times in different areas. If they are to

deteriorate, it also tends to be before session 20. Figure 11 presents the success curves for RD on all measures.

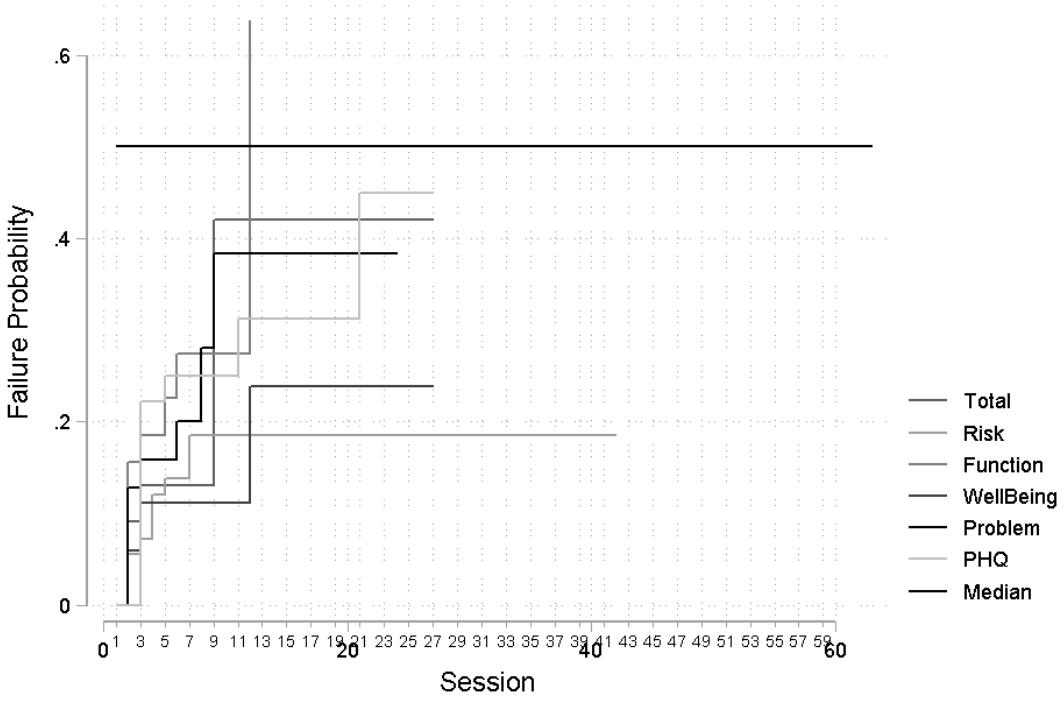

Figure 11.
Client Response Pattern Success Curves for RD Across Scales.

Impact of Demographic Characteristics on Therapy Outcomes

As described in Appendix A, underneath the survival function, is the hazard, which is mathematically related to the survival. It provides an instantaneous representation of risk at each time point given an immediate event right afterward. It does this by quantifying differences in the differing hazard functions given various covariates. As opposed to Kaplan-Meier estimates of

the survival and success curves that do not require large sample sizes, Cox proportional hazards models work better with moderate to large samples. While estimates can be obtained, coefficients and standard errors of the coefficients can be incredibly large or incredibly small. In these cases, caution is warranted when interpreting these models, and one may choose to modify the statistical model you are using (e.g., using less covariates, running them separately, etc.). Given the low number of events and general difficulty in interpreting the CSC and RD outcomes above, the following models used CSC/RI and RI only. In this case, the Cox models included all covariates simultaneously; this model best represented the client population and research questions. These covariates initially included race, sex, language, federal poverty level (FPL), insurance, trauma history, and age. Participants who identified as bi-racial were very scarce, and this prevented stable coefficients. They were excluded (nine participants only had one score; one participant had multiple scores on the outcome measure). The race, FPL, and insurance variables possessed several levels. Given low numbers on some of the levels across the following analyses, when applicable, a dummy variable was used instead. For race, this was clients who identified as White compared to everyone else. For FPL, this was clients who were <100% FPL compared to >100% FPL. For insurance, this was clients with no insurance compared to clients with Medicaid, Medicare, or private insurance. These effects were broken apart by level when possible to do so.

Patient Health Questionnaire. In the joint CSC/RI model, the covariates included in the model determined the sample tested and reduced it (i.e., covariates for clients must be complete to be included); the model included 72 clients and 41 successes. No covariates' hazard ratios were statistically significant in this model; the model is presented in Table 1. The RI model included 60 clients and 29 successes considering the covariates. This model is presented in Table

2. There were no noticeable effects for race, sex, or federal poverty level. Clients with insurance (Medicaid, Medicare, or private insurance) were more likely to experience reliable improvement compared to those without insurance (*HR* = 3.07, SE = 1.39, *p* = .013, 95% CI [1.26, 7.45]). Clients who self-reported a trauma history were also more likely to experience reliable improvement (*HR* = 7.21, *SE* = 4.93, *p* = .004, 95% CI [1.89, 27.54]). Age did not predict reliable improvement in this model, but it may have served to help aid in interpreting the other predictors within context. A submodel was ran that included all levels of insurance to explore these effects more in-depth along with all of the other predictors. Medicare (*HR* = 6.80, *SE* = 4.69, *p* = .005, 95% CI [1.72, 26.25]) and Medicaid (*HR* = 6.26, *SE* = 5.38, *p* = .033, 95% CI [1.16, 33.70]) clients had a much higher likelihood of achieving RI compared to clients without insurance. The clients with Medicaid and Medicare had hazard ratios with large standard errors; this makes their hazard ratios difficult to treat as reliable. More precision in these estimates should be obtained by adding additional clients before fully trusting these estimates. Clients with private insurance did not have a large hazard ratio (or one that was statistically significant) and were only slightly more likely to experience reliable improvement compared to clients with no insurance. Using a global test to test the proportional hazards assumption of the Cox regression revealed the model was viable and met this critical assumption, $\chi^2(10, N = 60) = 3.62$, $p = .96$. Given the fact that all clients were provided clinical services regardless of whether they possessed insurance, this opens up the possibility that insurance is acting as a proxy variable or "standing in" for something else or potentially distorting an effect or inflating an effect erroneously. An argument could be made that those with Medicaid are diferent than those with private insurance or Medicare. Those with Medicaid in the state of Texas have a lower household income and fewer resources. Those with Medicare are usually older but may have some available

resources. When clients with Medicaid and Medicare were taken out of the model, the insurance variable was no longer statistically significant (*HR* = 2.12, *SE* = 1.12, *p* = .16, 95% CI [0.75, 5.98]).

Table 2.
Cox Regression Results for Joint CSC/RI on the PHQ-9.

	Haz. Ratio	Std. Err.	z	P>z	[95% Conf. Interval]	
Race						
Caucasian	1	(base)				
African-American	0.76	0.31	-0.67	.50	0.35	1.68
Asian	0.70	0.60	-0.41	.68	0.13	3.79
Hispanic	0.53	0.42	-0.81	.42	0.11	2.47
Sex						
Male	1	(base)				
Female	0.89	0.36	-0.29	.77	0.40	1.97
FPL*						
<100%	1	(base)				
>100%	0.56	0.22	-1.47	.14	0.25	1.21
Insurance						
No Insurance	1	(base)				
Insurance	2.04	.74	1.94	.052	0.99	4.17
Trauma						
No Trauma History	1	(base)				
Trauma History	1.75	0.71	1.38	.17	0.79	3.89
Age	0.99	0.01	-0.69	.49	0.97	1.02

Note. *FPL = Federal Poverty Level.

Table 3.
Cox Regression Results for RI on the PHQ-9.

	Haz. Ratio	Std. Err.	z	P>z	[95% Conf. Interval]	
Race						
Caucasian	1	(base)				
African-American	1.00	0.45	0.00	.99	0.41	2.42
Asian	0.37	0.44	-0.83	.41	0.03	3.90
Hispanic	2.35	1.98	1.01	.31	0.45	12.30
Sex						
Male	1	(base)				
Female	0.99	0.49	-0.03	.98	0.37	2.63
FPL*						
<100%	1	(base)				
>100%	0.63	0.31	-0.93	.35	0.24	1.66
Insurance						
No Insurance	1	(base)				
Insurance	3.07	1.39	2.47	.013	1.26	7.45
Trauma						
No Trauma History	1	(base)				
Trauma History	7.21	4.93	2.89	.004	1.89	27.54
Age	0.99	0.02	-0.11	.91	0.96	1.03

Note. *FPL = Federal Poverty Level.

CORE-B. Cox models were estimated for the CORE-B by each subscale individually.

Total. On the CORE-B total scale, there were more clients available across the span of the predictor variables (for this model, n = 65 with 43 successes). Language was able to be estimated in the joint CSC/RI model along with the other predictors (race_dummy, sex, FPL_dummy, insurance_dummy, trauma history, and age). In the joint model, none of the

predictors significantly predicted an increase in occurrence of joint CSC/RI. In the RI model, also none of the predictors had statistically significant hazard ratios.

Risk. In the joint CSC/RI model (the only available one for this subscale), the race dummy hazard ratio was statistically significant (*HR* = 3.71, *SE* = 1.83, *p* = .008, 95% CI [1.41, 9.75]). Further exploration showed that clients who identified as African American had a higher rate of joint CSC/RI (*HR* = 5.80, *SE* = 3.04, *p* = .001, 95% CI [2.07, 16.22]). Clients who identified as Asian could not be calculated, and clients who identified as Hispanic, their hazard ratio was not statistically significant. None of the other predictors were statistically significant. This model met the proportional hazards assumption, $\chi^2(9, N = 65) = 3.34$, $p = .95$.

Function. In the joint CSC/RI, none of the predictors were statistically significant. In the RI model, none of the predictors were statistically significant.

WellBeing. In the joint CSC/RI model, none of the predictors were statistically significant. Also in the RI model, none of the predictors were statistically significant.

Problem. There were similar results in the Problem subscales models as with WellBeing. None of the predictors in either model was statistically significant.

Examples of Alternative Statistical Analyses – Key Examples

Logistic regression. The dichotomous nature of the event in this study would also allow for a logistic regression model to determine what factors could predict whether someone experienced an event or did not. One key finding from previous analyses was that on the PHQ, insurance status statistically significantly influenced whether someone experienced reliable improvement. Similar to a Cox model, a logistic regression was run to show how the same covariates would influence an analysis that generally looked at whether someone experienced reliable improvement in therapy at any time. A logistic regression was fit with reliable

improvement as the dependent variable and race, sex, language, FPL (dummy), insurance (dummy) , trauma history, and age. Similarly to the Cox model, the only statistically significant predictor was insurance status. When this is expressed as an odds ratio, (*OR* = 3.63, *SE* = 2.26, *p* = .038, 95% CI [1.07, 12.29]), it is interpreted as having insurance increases your odds of experiencing reliable improvement by 263% or you are 3.63 times more likely to experience reliable improvement compared to those without insurance. One crucial piece of information missing from this analysis was when this might occur in a course of therapy. Looking at the success curves for reliable improvement by insurance status illustrates this point. Figure 12 presents the success curves simultaneously for clients without insurance compared to those with insurance. The median survival time for those with insurance is around 7 sessions while the median survival for those without is around 20 sessions longer (*t* = 27). The beginning of therapy looks similar for these two groups; however, the impact appears earlier for those with insurance, and they soon depart from the other group in the middle sessions of therapy. The logistic regression and Cox model generally agree with regards to covariates' impact on reliable improvement. The logistic regression however, did not provide any further information regarding when this improvement might occur. The Cox model indicated a similar finding; however, afterward, success curves could be estimated to determine where RI would be most likely to occur.

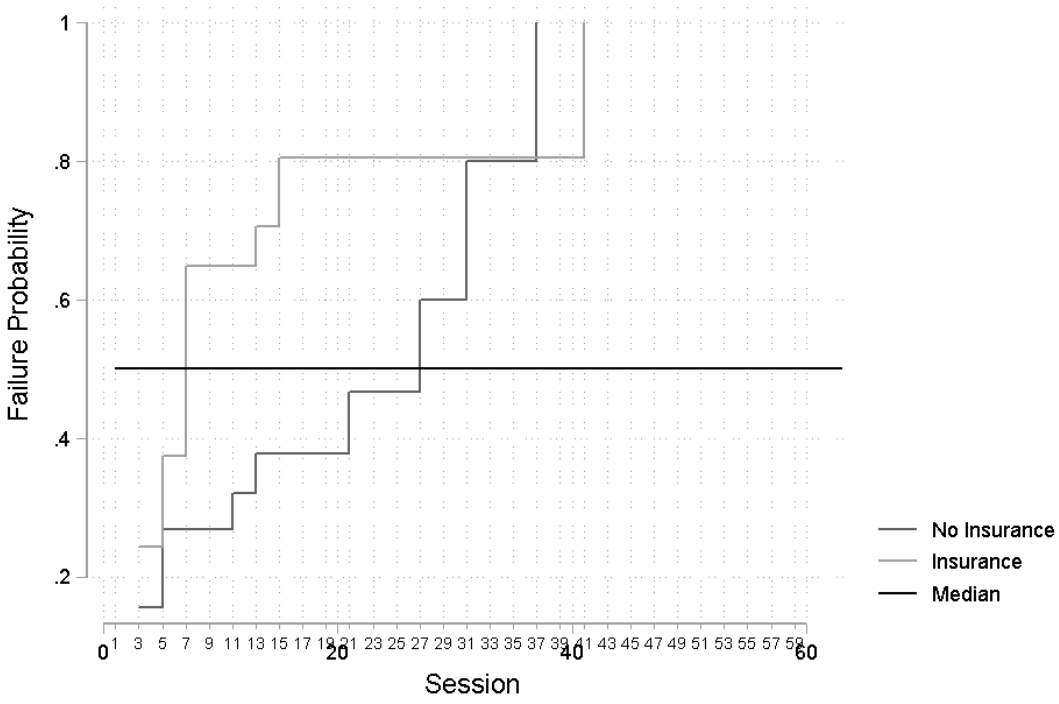

Figure 12.
PHQ-9 Success Curves by Insurance Status.

Linear growth curve. With the outcome measures being taken after each session over time, this also enabled the use of a hierarchichal linear model, specifically tailored towards longitudinal analysis, called a linear growth curve model. This allows for an outcome to be measured continuously over time and provides an overall trajectory of change in that outcome across time. Compared to traditional longitudinal approaches with more stringent assumptions, the linear growth curve model accommodates the interdependent nature of repeatedly measuring an outcome, as a future score could conceivably be somewhat related to the previous one. This analysis also has the advantage in using the PHQ total scores across all sessions and all available clients without missing data prohibiting the analysis. The linear growth curve model represents

all clients starting points in a random-effects variance component, or intercept. This allows for all clients to have different starting points, and the variance of this intercept component represents how much variability exists in these different starting points. The growth over time is represented by a linear variable, in this case, the session number associated with the score. Also representing growth, there is a quadratic growth variable equal to the linear growth variable squared (Time*Time). This captures any bend in the growth curve over time.

Figure 13 presents the growth curve over time for the PHQ-9 total scores. In general, the trend shows that clients' PHQ-9 total scores improve in therapy over time. On average, clients start with a total score of 13, and it gradually decreases over time. To drop from 13 to eight in their PHQ-9 total score takes somewhere between 16 to 20 sessions generally. At the individual level (in this case, Level-2 as scores are clustered within individuals), other covariates can be included in the model to see if they can explain variance in either individual's starting score (random intercept) or in their slopes (either fixed or random; fixed in this case).

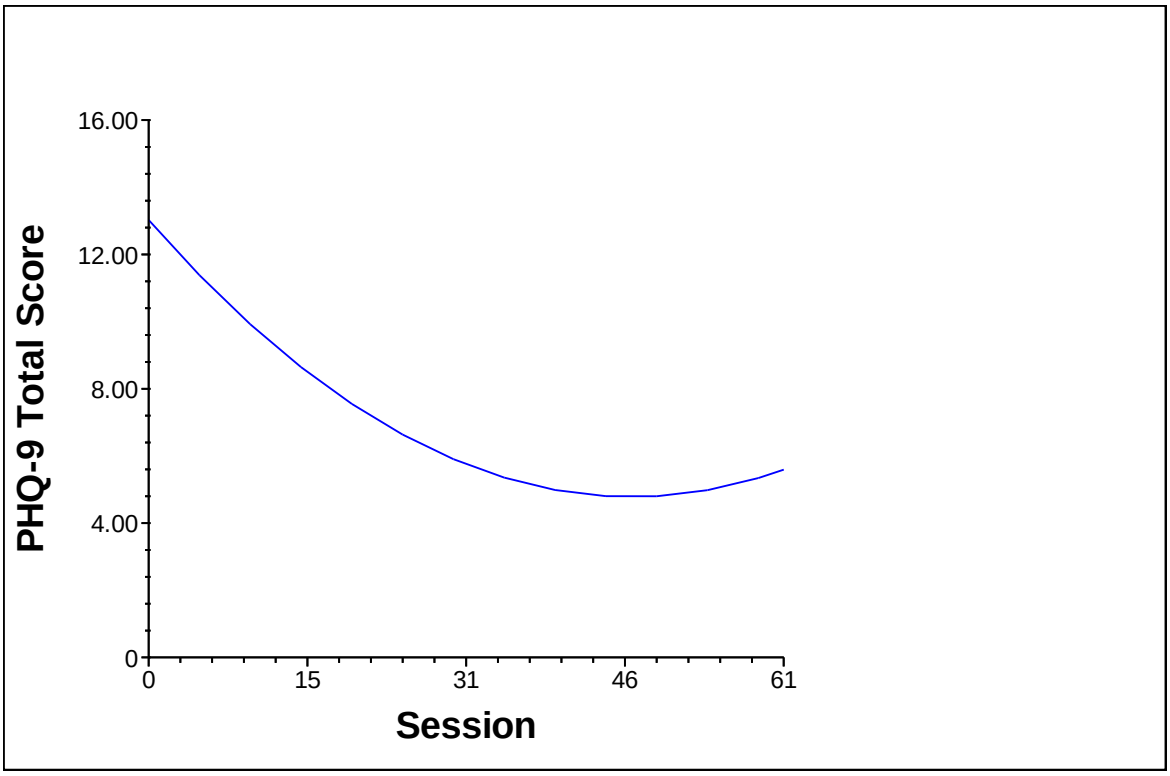

Figure 13.
Linear Growth Curve Model for PHQ-9 Total Scores.

From this model, growth curves can be drawn based on covariate values to illustrate differences. Insurance status (dummy) was added as a predictor to try and explain the random intercept and to see differences in growth over time for the two groups. Figure 14 depicts these two groups' growth curves together.

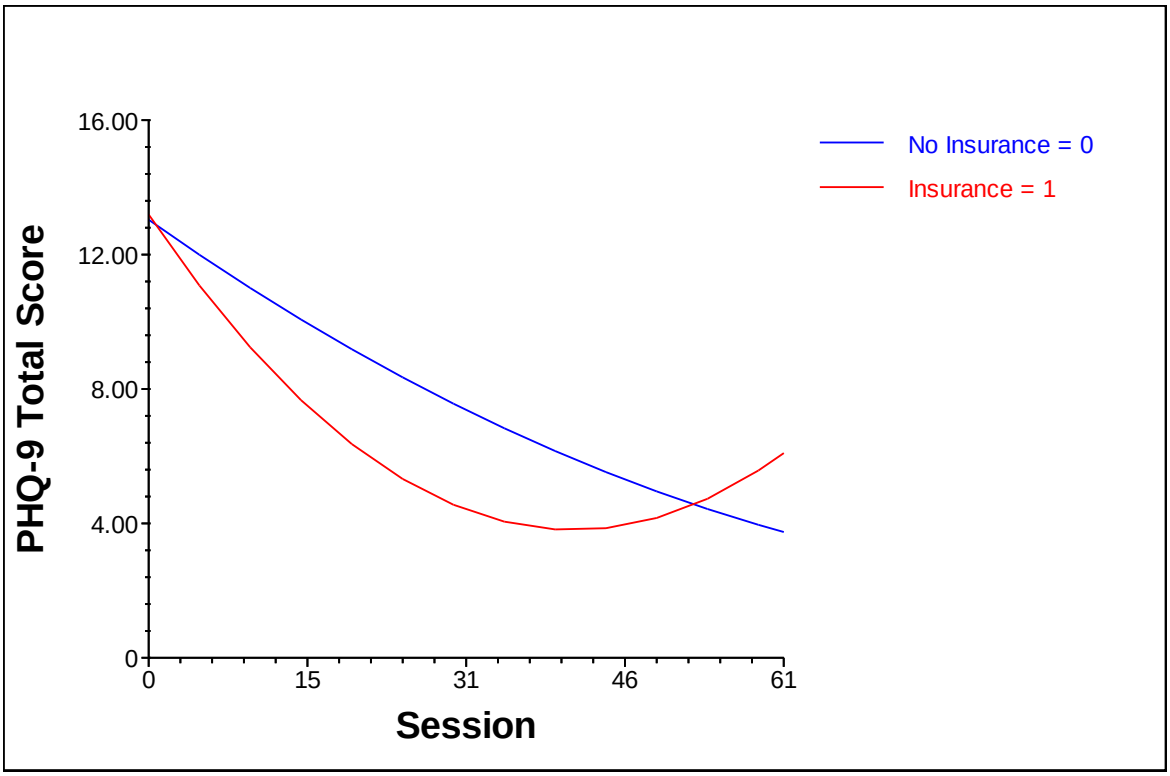

Figure 14.
Linear Growth Curve Model for PHQ-9 Total Scores by Insurance Status.

The model results, as displayed in Figure 14, show that individuals were similar towards the beginning of therapy but digressed from one another over time. For those with insurance, going from 13 to eight was achieved in 12 to 15 sessions, and for clients without insurance, they achieved a five point reduction somewhere around 27 sessions—1.5 to two times as long.

These two linear growth curve models show general growth over time for clients on the PHQ, and in general, clients get better after receiving therapy. Although one can include predictors in the linear growth curve model, as illustrated above, to try to parse out differences, the model does not utilize event data alongside growth over time. Predictors can be incorporated at level-1 or level-2 in a hierarchical linear model, and one could put in a predictor representing a

key event such as CSC or RI. However, by doing this, you are not obtaining the same information, and arguably, it is also not as useful information. This is because in survival analysis, the growth over time is dependent upon the events experienced by the clients. In other words, survival analysis jointly estimates event data with longitudinal data. Linear growth curve modeling accomplishes this in more of a piecewise fashion and does not have the dependencies/interrelatedness of the data built into its analytical framework.

CHAPTER V

DISCUSSION AND CONCLUSION

This study sought to analyze therapy outcome data from a telehealth counseling clinic to answer research questions that were pertinent to the clinic and provide information, both to the counseling psychology literature and to community stakeholders. Through the use of the Jacobson and Truax (1991) CSC paradigm, outcomes included CSC, RI, RD, and no reliable change. This perspective adds nuance in considering psychotherapy outcomes and whether clients improved after receiving psychotherapy. This allowed for more in-depth analyses, considering both the outcome measure used (*between-measure* outcomes) and the outcome being considered (*within-measure* outcomes). In addition to questions about improvement on the outcome measures in the study, client demographic and clinical characteristics were also considered in these analyses to help provide context and to investigate whether these outcomes can be generalized across different clients within the clinic's client population. These analyses were presented within the framework of survival analysis, and thus, key highlights of these results were compared with alternative statistical techniques further elucidating key points presented that survival analysis is best suited for this type of data.

Dual Service-Delivery—In-Person and Telepsychology

Before examining any outcomes, it was important to ensure that all clients were generally comparable and combining results would be feasible—most importantly in terms of the services they received. Log-rank tests revealed that on the PHQ-9 and CORE-B scale/subscales, clients who received in-person or group services along with telehealth services possessed similar

outcomes to clients who only received telehealth services. This allowed for the largest sample to be analyzed and for the results to generalize across the largest number of people.

This provides preliminary evidence of effectiveness to clinics who provide a combination of services, or dual service-delivery clinics. With telepsychology continuing to be utilized and researched in the field, there may be pressure on clinicians or clinics to adopt some form of telepsychology in their own practices. One potential barrier of adoption however, is the need to continue to provide traditional services, either for financial reasons or to appease stakeholders. A dual service-delivery clinic would be able to provide both services in ratios that are suited for them to meet both needs accordingly. When clinics are positioned inside of communities and potentially receiving grant funding or community funding, community stakeholders are also interested in how the clinics are performing. This enables community stakeholders to justify funding, coordinate resources, gain understanding on what clients are being served, by what modality, and for what cost.

Overall Improvement

Client outcomes were positive, in general. Clients achieved either CSC or RI on the PHQ after an average of 11 sessions. For the CORE-B total score, clients improved in four sessions, and on the other subscales, clients improved in four to seven sessions. One reason why the CORE-B scale/subscales may have shown earlier CSC/RI events is because of its multifaceted nature. It encompasses both client problems and risk with positive elements such as wellbeing and functioning compared to the PHQ, a measure solely assessing depressive symptomology.

Client Response Patterns

Clients who achieved CSC on any outcome measure, tended to do so before four sessions of therapy, with some occurring as early as the second session. This may represent an early-

treatment spurt (Tang & DeRubeis, 1999) characterized by an initial reduction in symptoms to a significant degree. RI tended to take longer and may provide a more accurate depiction of improvement in therapy. RI occurred anywhere from five sessions (CORE-B Problem) to 13 sessions (PHQ). These results are generally aligned with previous research utilizing survival analysis to measure psychotherapy outcomes (Hansen & Lambert, 2003) and provide additional evidence for the presence of sudden gains in therapy for some clients (Stiles et al., 2003). RD was a difficult outcome to measure in this study due to limited sample size of clients who experienced it and a low number of events. In most analyses, a median survival for RD could not be calculated (CORE-B Function at 12 sessions being the only exception). In some ways, these are positive results as the goal of therapy is for clients to 1) improve and 2) not get worse. From the success curves, a 25th percentile success estimate could be interpreted if desired. Overall, it appeared that there was no one point in therapy where clients reliably deteriorated, and these deteriorating effects seemed to occur before 20 sessions.

Evaluation of Outcome Measures

The choice of outcome measure(s) is an important one, and certain characteristics of potential outcome measures should be considered when making this decision (Lambert et al., 2001). Internally, clinicians and supervisors may want specific information to help provide feedback to clients through their therapy course (Howard et al., 2006). Additionally, there may be other contextual factors that affect this choice such as affordability of the outcome measure and what data are required to meet certain guidelines for funding bodies. Considering the results of this study, the PHQ provided vital information, both with CSC and RI, as these success curves tended to be more "spread-out" supplying more information across the therapy course. With regards to CSC and the CORE-B, these curves were very "clustered" together around the first

few sessions, and this did not relay any information about CSC on the CORE-B in the middle or later parts of therapy. With RI, the CORE-B success curves were more "spread-out." Also, the subscales on the CORE-B achieved outcomes at different times, which provided detailed information on what areas of improvement are likely to occur and when.

Client Characteristics

Preliminary findings regarding client characteristics suggested that insurance status, trauma history, and race impacted client outcomes in some situations for this clinical population. For the PHQ possessing insurance (specifically Medicaid and/or Medicare) positively impacted the probability of achieving reliable improvement. This was not the case for private insurance clients who were more like the no insurance clients. As opposed to expected results, clients who possessed a trauma history also tended to have higher reliable improvement than those without. On the CORE-B, client characteristics tended to not result in any differences in client outcomes; this could be due to minimal effects on outcomes by these variables or insufficient power (i.e., low sample size of clients and/or low number of events) to detect these effects. One exception was considering joint CSC/RI on the CORE-B Risk scale, clients who identified as African American had a higher probability of experiencing an improvement (CSC or RI) on this scale.

Many of the previous studies on psychotherapy outcomes were within the context of undergraduate counseling centers, employee programs, and privately managed mental healthcare (e.g., Hansen et al., 2002; Hansen & Lambert, 2003; Lambert et al., 2001). This leaves a major subsample of the U.S. population underrepresented given the context in which they seek or have access to mental health services. For instance, low-income client households, ethnic minorities, such as African-Americans and Latinos, and rural residents are not often represented or are underrepresented. This study helps to address the dearth of literature regarding outcomes for

these clients, and these results are within the frame of treatment-as-usual psychotherapy effectiveness and provides for more generalizability to this specific population.

Impact on the Clinic and the Community

Not only do rural clients face many barriers to quality mental health care (Brossart et al., 2013), so too, do clinics who treat them. Often, they face limited resources to provide quality mental health care to these residents. One critical complaint often raised by mental health clinics is limited time. Going as far back as 1939 in describing psychiatrists' use of psychotherapy for institutionalized patients, Harris (1939) wrote:

> In attempting to treat the Briarean syndrome of psychoneurosis, economy of the physician's time is the most important desideratum. An extensive literature of psychotherapy indicates only too clearly that treatment is, with rare exceptions, a time-consuming task for physician and patient. (p. 1)

Optimizing available resources and providing briefer, yet effective psychotherapy, has been a continual theme throughout the 20th century into present day. Straker (1968) wrote about a reorganization that occurred in a university hospital in Montreal in the 1960s; for certain patients they provided 10 to 12, 30-minute psychotherapy sessions to aid with high client dropout rates, heavy clinic congestion, and long waiting times for services. He goes on to write: "Therapy is directed toward attainable goals. These include mobilizing the patient's own resources, termination of the therapy contact when feasible, and permitting the patient to go on unaided" (Straker, 1968, p. 1224). Optimization of clinic resources can be informed by data-driven decision-making.

Data-driven clinic policies. In adopting new policies, clinics may feel that they are "throwing darts in the dark" in regards to what to consider and what would be best for their

clientele. Community health surveys can provide region-level or community-level information (e.g., Brossart et al., 2013). Clinics can also utilize client and psychotherapy outcome data to inform their decisions that would suit their population's specific needs. For instance, clinics may want to implement a session-limit policy. They could generally utilize information from the literature (Hansen & Lambert, 2003) as a basis, and from there, incorporate contextual (community health information) and individual (outcome analyses) to further detail their proposed policy. Lambert (2007) argued that, "…in our mind, treatment length is most appropriately driven by patient treatment response rather than theoretical or cost-based decisions" (p. 3). He also summarizes data from over 6,000 patients analyzed with survival analysis techniques, the median survival was from 11 to 21 sessions, and the 75th percentile was between 25 to 45 sessions (Lambert, 2007). From the results presented above, 20 sessions would not be an unreasonable choice with room for more sessions in certain circumstances or upon meeting specific criteria guided by other available information (i.e., contextual and individual information). In addition, from analyses of client factors, certain clients may be allowed additional sessions given a general trend towards dropout or premature termination from therapy found previously. Furthermore, if certain clients were identified as being more "at-risk" for dropout or premature termination, specially tailored interventions and additional counselor efforts/clinic interventions could be implemented to help prevent premature termination, session no-show's, and client dissatisfaction with services. Clinics can also use these types of analyses as "hard evidence" to funding agencies for how they support their community, reduce usage of other vital community resources, such as emergency departments, help reduce barriers to quality mental health care, and provide information on continual improvements and areas needing additional attention/funding.

Clinics can also serve as a "home-base" for mental health resources in their community, both in terms of providing professional psychological services but also through encouraging and facilitating mental health literacy in their community. Jorm (2012) outlines areas of mental health literacy that could potentially be addressed by mental health clinics, 1) recognition of mental health disorders to enable faster treatment, 2) what resources are available and what effective treatments are available, 3) awareness of effective and viable self-help strategies (e.g., utilizing family/friend support, increasing exercise frequency and intensity, and doing more pleasurable activities), 4) training and provision of mental health first aid by key community members, and 5) information on helping to prevent mental health disorders. He recommends the following interventions, whole-of-community campaigns to increase community education about mental health, placing interventions in schools, colleges or universities, as these places tend to be well located and suited to facilitate these types of activities, hosting mental health first aid training events, and starting or supporting web-based programs, such as educational websites or other technological strategies/innovations to promote increased mental health literacy (Jorm, 2012).

Limitations

This study utilized data from a telehealth counseling clinic that provides counseling services to clients who are underserved. The data collected by the TCC was done so in the midst of providing clinical services; given this milieu, it is important to consider that data cannot always be gathered consistently as the primary focus is on providing psychological services. Survival analyses better accommodate data that may be incomplete; specifically, the Kaplan-Meier estimator utilizes data from clients who are present at any specific time in the study, given that they have not experienced the event or have been censored. One inherent limitation of

survival analysis is that clients must have 1) a baseline measurement and 2) at least one additional measurement occasion to contribute information to the analysis. This is not always the case or feasible in a clinic setting. In this study, this contributed to a major loss of potential participants who did not have adequate data for the analyses but were seen by the clinic.

When clients were excluded from the analyses, it also had a major impact on the study covariates. While survival analyses do not require the use of complete outcome data to analyze results, one essential assumption of including covariates in these analyses is that they contain complete data. As the number of covariates increase, the impact of this assumption also increases and led to decreased sample sizes in the Cox models. With decreased sample sizes, some hazard ratio coefficients possessed high standard errors from a lack of clients on that covariate or a low number of events (the source of information for the analysis), and thus, the model produced imprecise estimates in these cases. One benefit for clinics obtaining preliminary data results such as these is that estimates can be improved by including additional clients. As a clinic continues to see clients, they can re-analyze their data to increase the precision of these estimates combined with specific hypotheses or clinic policy questions in mind.

Future Directions

This study introduced survival analysis for use in counseling psychology outcome research. Basic conceptual and mathematical topics of survival analysis were presented from the perspective of conceptualizing counseling outcomes in terms of statistical modeling. This study could not present survival analysis in its entirety and should not be considered comprehensive, and some key points were omitted to aid in understanding. From this starting point however, there are continued avenues of exploration and additional techniques, ideas, and models that can be utilized. As opposed to explaining survival analysis in terms of procedural steps, this study

has presented it in terms of a way of thinking about data, research questions pertinent to the researcher, and the statistical language (analyses) necessary to analyze the data while considering the statistical model being tested. Data structures can be quite complex; given this, there are additional aspects of the data that can be incorporated into a model looking at survival data. For instance, the use of hierarchical linear modeling can be extended into survival analysis. Additionally, there are other forms of survival data models that could expand psychologists' thinking about their data and models they test in their studies.

HLM and survival analysis. In his tutorial article on hierarchical linear modeling (HLM or MLM) for counseling psychologists, Kahn (2011) expressed his opinion that "MLM [HLM] continues to be underutilized in counseling psychology" (p. 257). One potential direction for psychologists wanting to utilize survival analysis in their research on therapy outcomes would be to utilize multilevel survival analysis (Hox, 2010) to accommodate data that possess hierarchical structures. Analyses in this study generally took into account measurements *within* individuals in its statistical models. Not addressed in this study, but one improvement would be to look at how survival of clients may change when you consider therapists (clients being further clustered within therapist) as clients from the same therapist may be related in some way or have similar outcomes. In addition, given clinic location, specifically in this case—a rural/micropolitan clinic location/population—considering the impact of county or city on outcomes may also be warranted. Clients would be clustered within counties to accommodate similarities of these clients who may have comparable experiences. Access to HLM software is also increasing as the methodology continues to rise in popularity. The HLM7 package was utilized in this study; the student version (used here) is freely available (with some limitations). The common statistical software, IBM SPSS, also contains a HLM module that can be used. Other popular statistical

software containing this ability is STATA and SAS. In terms of modeling, HLM and survival analysis can also be modeled with the statistical modeling software, Mplus (Muthén & Muthén, 2017) as explained below.

Advanced survival analysis methodology. One important research question in this study was client deterioration (or RD). Given the data from the clinic, this proved difficult to model given low sample size and success rate. In some ways, these findings are positive. Having a median survival for clients on RD would practically and clinically translate to many individuals having a negative outcome from therapy. A related outcome to reliable deterioration not measured in this study is client drop-out. Corning and Malofeeva (2004) provide a tutorial article on survival analysis to counseling psychologists analyzing client drop-out. They formulate different events in much the same way here including premature termination, mutual termination, and arbitrary-end to therapy (Corning & Malofeeva, 2004). When outcomes are mutually exclusive, they discuss researchers can utilize *competing-risks* survival analyses that jointly estimate the hazard function over time for multiple events simultaneously (Corning & Malofeeva, 2004). This approach better "honors" the reality of drop-out of therapy clients by incorporating multiple possibilities into the analysis, which are more reflective of the reality of clinical service delivery. The clinically significant change methodology cannot currently model data in this way as CSC, RI, RD, and NC, are related to one another and are not mutually exclusive. However, as survival analysis methodology continues to improve, this mutually exclusive requirement may be lifted and modeling their interdependency might be possible.

This study utilized single event survival analysis techniques. When an individual experienced the event, they were dropped out of the study by the analysis regardless of whether they continued in therapy or not. Survival analytic methods can accommodate multiple events—

most notably, in a *recurrent-events* survival analysis or *frailty* model (Hosmer, Lemeshow, & May, 2010). One implicit assumption in the clinically significant change literature/approach is that clients will do their best at the end of therapy, and this is where you should see the change. In some ways, this does not consider that clients may get "better" along the way. With CSC being the most restrictive outcome, it would be difficult to model this in a different way; however, it would not be difficult to conceive that clients might obtain multiple instances of RI over their course of therapy. Another assumption made is that all error in the reliability of scores is measurement error. It does not consider that clients may get worse for other factors or have significant events in their life that could impact treatment. Further models and conceptualizations of therapy outcomes could directly build this into a statistical model that is more reflective of therapy. For example, viewing growth in therapy as nonlinear or possessing discontinuous patterns of change (Hayes, Laurenceau, Feldman, Strauss, & Cardaciotto, 2007), could better inform theory surrounding psychotherapy and the statistical models employed. In other words, there are different ways and considerations necessary when attempting to model variability within client scores over time.

Survival analysis modeling. Survival analysis, as with many statistical techniques, have traditionally been restricted in their ability to incorporate other techniques, data structures, or types of relationships hypothesized and tested. Mplus is one easily accessible statistical modeling software. Developed and currently maintained by Muthén and Muthén (2017), their approach involves the incorporation of different types of variables (i.e., continuous, categorical—binary and ordered, observed, and latent variables, among others) into a general statistical modeling environment/framework. Further, Bengt O. Muthén (2002) explains the importance of latent variables and notes that many statistical analyses implicitly use them; through the use of latent

variables, this framework capitalizes on them as they can “capture a wide variety of statistical concepts, including random effects, missing data, sources of variation in hierarchical data, finite mixtures, latent classes, and clusters” (p. 81). Historically:

> Structural equation modeling (SEM) took factor analysis one step further by relating the constructs to each other and to covariates in a system of linear regressions thereby purging the "structural regressions" of biasing effects of measurement error. The idea of using systems of linear regressions emanated from supply and demand modeling in econometrics and path analysis in biology. In this way, SEM consists of two ideas: latent variables and joint analysis of systems of equations. It is argued here that it is the latent variable idea that is more powerful and more generalizable. (Muthén, 2002, p. 82)

His emphasis on latent variables extends both to continuous latent variables and categorical latent variables (Muthén, 2002). With this inclusion, it allows for the unification and extension of many statistical analyses/techniques previously housed in their own traditions; this includes “SEM, growth curve modeling, multilevel modeling, missing data modeling, finite mixture modeling, latent class modeling. and survival modeling” (Muthén, 2002, p. 82). This expands researchers’ modeling possibilities with regards to survival analysis modeling. Examples from their website (www.statmodel.com) include discrete-time survival analysis using latent variable modeling (Raykov, Gorelick, Zajacova, & Marcoulides, 2017), discrete-time survival mixture analysis (Muthén & Masyn, 2005), and continuous-time survival analysis in latent variable models (Asparouhov, Masyn, & Muthén, 2006; Asparouhov, 2014; Muthén, Asparouhov, Boye, Hackshaw, & Naegeli, 2009). This framework allows for more advanced

statistical models allowing the researcher to model survival data within the context of other pertinent variables and model relationships between them.

Summary

"Psychotherapy is, for the most part, a longitudinal process occurring over a series of sessions" (Corning & Malofeeva, 2004, p. 355).

Counseling psychologists are in the best position to impact change through providing therapy to clients and by conducting research targeted towards understanding clients, therapy, the change process, and the greater context and implementing interventions that are going to be effective, targeted to the client, and meaningful. One way psychologists do this is in the use of theories and models.

Theory, models, and the postrevolution counseling psychologist. Counseling psychologists are keen on utilizing theories in their work with clients and in their role as researchers. Clinicians utilize counseling theories, such as Roger's person-centered therapy, or cognitive behavioral therapy (CBT). Rychlak (1968) defines theory as used most commonly in psychology as, "a coherent set of hypothetical, conceptual (meaning, in psychology, operationally defined), and pragmatic (predictive) principles forming the general frame of reference for a field of inquiry" (p. 11). By extension, when clinicians form case conceptualizations of their clients they are developing a "theory of the client." Inherent in this definition of theory is the acknowledgment that it must link back to reality. Given this framework and psychologists' already existing comfort with theory, employing a model-based perspective is not too far of a jump. Rodgers (2010) generally defines a model as, "1. A model matches the reality that it describes in some important ways. 2. A model is simpler than that

reality" (p. 5). You might hear a counseling psychologist use the phrase "theory of a client," but the phrases "theory of my study" or the "theory of my data" are used much less. Further, a question about the mathematical model underlying the research being done might receive shrugs or confused looks. They would however, be more able to describe the patterns of behavior (a model) exhibited in one of their clients. Making this link, Rodgers (2010) suggests two roles for statistical models in psychology:

> Two different roles exist for postrevolution statistical models. The first is a model-comparison framework based on the application of existing statistical methods like ANOVA and SEM, used and applied by researchers who study and develop models of behavior. The second involves the development of mathematical models to match topics of explicit interest to researchers. Within this second framework, substantive scientists study behavior and from that process develop mathematical models specific to their research domain. Here, statistical methods are used to compare and evaluate these mathematical models. Both approaches are important for a successful methodology within psychology. (p. 8)

One implicit aspect of both theories and models is their requirement of possessing utility. G. E. P. Box (1979) wrote, "Models of course are never true, but fortunately it is only necessary that they be useful" (p. 2; Rodgers, 2010).

The study outlined basic concepts in survival analysis by examining data from a telepsychology clinic. Research questions were formulated and answered through a survival analytic framework. The choice of questions by the psychologist is of importance. Clinically, research questions can be formulated that center around client outcomes, and more systemically, clinic outcomes can be measured that have real-world impact on the services that can be

provided, funding for the clinic, and the bigger community at large. After forming research questions, the analysis of a data structure should reflect the characteristics of the data. For survival analysis, this was the incorporation of both discrete events and time. An example of this was presented looking at client outcomes from the TCC. There is needed research looking at "treatment-as-usual" psychotherapy outcomes because it best represents the reality of providing clinical services, especially in a rural, underserved area. The clinically significant change paradigm was used as it provides a structure of measuring responses to therapy, and there was literature available to compare with afterward. Of which, results from this study generally support other research showing that 11 to 14 sessions of psychotherapy lead to clinically significant change or reliable improvement on outcome measures, and additionally, there is much client growth in the first few sessions. The impact of client demographics remains unclear; however, insurance status appears to be a preliminary factor positively affecting clients in this region.

The use of survival analysis in the counseling psychology literature is minimal, with a few exceptions (Corning & Malofeeva, 2004). However, many psychologists would not argue with a conceptualization of psychotherapy as a longitudinal process that occurs across a series of psychotherapy sessions (Corning & Malofeeva, 2004). Moving forward, counseling psychologists are well suited to expand their clinical services through the use of telepsychology and to expand their thinking and research to involve statistical models that better represent the reality they are trying to capture and understand.

REFERENCES

2010 Census: Population of Texas counties, arranged in alphabetical order. (2013, May 21). Retrieved from https://www.tsl.texas.gov/ref/abouttx/popcnty12010.html

Anderson, E. M., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology*, *57*(7), 875-888. doi: 10.1002/jclp.1056

Armstrong, T. W., Surya, S., Elliott, T. R., Brossart, D. F., & Burdine, J. N. (2016). Depression and health-related quality of life among persons with sensory disabilities in a health professional shortage area. *Rehabilitation Psychology*, *61*(3). 240-250. doi: 10.1037/rep0000083

Asparouhov, T., Masyn, K., & Muthén, B. (2006). Continuous time survival in latent variable models. In *JSM proceedings, biometrics section* (pp. 180-187). Alexandria, VA: American Statistical Association.

Asparouhov, T. (2014). *Continuous - time survival analysis in Mplus.* Retrieved from http://www.statmodel.com/download/Survival.pdf

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose–effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology*, *77*(2), 203-211. doi: 10.1037/a0015235

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*(2), 184-196. doi: 10.1037/0022-006X.69.2.184

Barkham, M., Connell, J., Stiles, W., Miles, J. N., Margison, F., Evans, C., & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: The good enough level. *Journal of Consulting and Clinical Psychology, 74*(1), 160-167. doi: 10.1037/0022-006X.74.1.160

Barkham, M., Mellor-Clark, J., Connell, J., Evans, C., Evans, R., & Margison, F. (2010). Clinical outcomes in routine evaluation (CORE) - the CORE measures and system: Measuring, monitoring and managing quality evaluation in the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 175-219). Hoboken, NJ: John Wiley & Sons.

Bohart, A. C., & Tallman, K. (1999). *How clients make therapy work: The process of active self-healing*. Washington, DC, US: American Psychological Association. doi: 10.1037/10323-000

Box, G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association, 74*(365), 1-4.

Breslau, N., Chilcoat, H. D., Kessler, R. C., & Davis, G. (1999). Previous exposure to trauma and PTSD effects of subsequent trauma: Results from the Detroit Area Survey. *American Journal of Psychiatry, 156*(6), 902-907. doi: 10.1176/ajp.156.6.902

Brossart, D. F., Wendel, M. L., Elliott, T. R., Cook, H. E., Castillo, L. G., & Burdine, J. N. (2013). Assessing depression in rural communities. *Journal of Clinical Psychology, 69*(3), 252-263. doi: 10.1002/jclp.21949

Brown, G. W., & Harris, T. (1978). *Social origins of depression: A study of psychiatric disorder in women.* London: Tavistock.

Brown, A., Brody, G., & Stoneman, Z. (2000). Rural black women and depression: A contextual analysis. *Journal of Marriage and Family, 62*(1), 187-198. doi: 10.1111/j.1741-3737.2000.00187.x

Burdine, J., Clark, H., Appiah, B., Shea, L., & Hollas, C. (2012). Health assessment in rural communities: A critical organizing and capacity building tool. In R. Crosby, R. Vanderpool, M. Wendel, & B. Casey (Eds.), *Rural populations and health: Determinants, disparities and solutions* (pp. 171-190). San Francisco, CA: Jossey-Bass and Pfeiffer.

Castillo, L. G., & Caver, K. A. (2009). Expanding the concept of acculturation in Mexican American rehabilitation psychology research and practice. *Rehabilitation Psychology, 54*(4), 351-362. doi: 10.1037/a0017801

Chang, J. E., Frazier, C. G., & Elliott, T. R. (2013). Using videoconferencing to provide psychological services to a rural clinic: A unique town and gown partnership. *Texas Psychologist, 65*(2), 9-14.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist,49*(12), 997-1003. doi: 10.1037/0003-066X.49.12.997

Cooper, M. (2008). *Essential research findings in counselling and psychotherapy: The facts are friendly.* London: Sage Publications.

CORE System Group. (1998). *CORE system (information management) handbook*. Rugby, UK: Core System Group.

Corning, A. F., & Malofeeva, E. V. (2004). The application of survival analysis to the study of psychotherapy termination. *Journal of Counseling Psychology, 51*(3), 354-367. doi: 10.1037/0022-0167.51.3.354

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose–response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*(1), 114-130. doi: 10.1111/bjc.12031

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, *79*, 15-22. doi: 10.1016/j.brat.2016.02.003

Dohrenwend, B. S., & Dohrenwend, B. P. (Eds.). (1974). *Stressful life events: Their nature and effects.* New York, NY: Wiley.

Duncan, B. L., Miller, S. D., & Sparks, J. A. (2004). *The heroic client: A revolutionary way to improve effectiveness through client-directed, outcome-informed therapy*. San Francisco, CA: Jossey-Bass.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*(5), 319-324. doi: 10.1037/h0063633

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, *5*(1), 75-98. doi: 10.1177/0959354395051004

Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. *Behaviour Research and Therapy*, *69*, 54-62. doi: 10.1016/j.brat.2015.04.001

Garney, W. R., McCord, C. E., Walsh, M. V., & Alaniz, A. B. (2016). Using an interactive systems framework to expand telepsychology innovations in underserved communities. *Scientifica, 2016, Article ID 4818053*. doi: 10.1155/2016/4818053

Gigerenzer, G., & Krauss, S. (2001). Statistisches denken oder statistische rituale: Was sollte man unterrichten? In M. Borovcnik, J. Engel, & D. Wickmann (Eds.), *Anregungen zum stochastikunterricht: Die NTCM-standards 2000. Klassische und Bayessche sichtweise im vergleich* (pp. 53-62). Franzbecker.

Gilbody, S., Richards, D., Brealey, S., & Hewitt, C. (2007). Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): A diagnostic meta-analysis. *Journal of General Internal Medicine, 22*(11), 1596-1602. doi: 10.1007/s11606-007-0333-y

Gonzalez Jr, G. E., & Brossart, D. F. (2015). Telehealth videoconferencing psychotherapy in rural primary care. *Journal of Rural Mental Health, 39*(3-4), 137-152. doi: 10.1037/rmh0000037

Green, B. L., Goodman, L. A., Krupnick, J. L., Corcoran, C. B., Petty, R. M., Stockton, P., & Stern, N. M. (2000). Outcomes of single versus multiple trauma exposure in a screening sample. *Journal of Traumatic Stress, 13*(2), 271-286. doi: 10.1023/A:1007758711939

Growth in urban population outpaces rest of nation, Census Bureau reports. (2012, March 26). Retrieved from https://www.census.gov/newsroom/releases/archives/2010_census/cb12-50.html

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research, 7*(1), 1-20.

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. doi: 10.1093/clipsy.9.3.329

Hansen, N. B., & Lambert, M. J. (2003). An evaluation of the dose–response relationship in naturalistic treatment settings using survival analysis. *Mental Health Services Research, 5*(1), 1-12. doi: 10.1023/A:1021751307358

Harris, H. I. (1939). Efficient psychotherapy for the large outpatient clinic. *New England Journal of Medicine, 221*(1), 1-5. doi: 10.1056/NEJM193907062210101

Hayes, A. M., Laurenceau, J. P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review, 27*(6), 715-723. doi: 10.1016/j.cpr.2007.01.008

Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley-Interscience.

Hilty, D. M., Ferrer, D. C., Parish, M. B., Johnston, B., Callahan, E. J., & Yellowlees, P. M. (2013). The effectiveness of telemental health: A 2013 review. *Telemedicine Journal and e-Health, 19*(6), 444-454. doi: 10.1089/tmj.2013.0075

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time to event data*. Hoboken, NJ: Wiley.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*(2), 159-164. doi: 10.1037/0003-066X.41.2.159

Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*(10), 1059-1064. doi: 10.1037/0003-066X.51.10.1059

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

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. doi: 10.1037/0022-006X.59.1.12

Jorm, A. F. (2012). Mental health literacy: Empowering the community to take action for better mental health. *American Psychologist, 67*(3), 231-243. doi: 10.1037/a0025957

Kadera, S. C., Lambert, M. J., & Andrews, A. A. (1996). How much therapy is really enough? A session-by-session analysis of the psychotherapy dose–effect relationship. *Journal of Psychotherapy Research and Practice, 5*(2), 132-151.

Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology*, *58*(2), 257-271. doi: 10.1037/a0022680

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*(12), 1048-1060. doi: 10.1001/arch-psyc.1995.03950240066012

King, L. A., King, D. W., Fairbank, J. A., Keane, T. M., & Adams, G. A. (1998). Resilience–recovery factors in posttraumatic stress disorder among female and male Vietnam veterans: Hardiness, postwar social support, and additional stressful life events. *Journal of Personality and Social Psychology*, *74*(2), 420-434. doi: 10.1037/0022-3514.74.2.420

Kogan, S. M., Brody, G. H., Crawley, C., Logan, P., & Murry, V. (2007). Correlates of elevated depressive symptoms among rural African American adults with type 2 diabetes. *Ethnicity & Disease*, *17*(1), 106-112.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613.

Lambert, M. J. (1992) Implications of outcome research for psychotherapy integration. In J. C. Norcross & M. R. Goldstein (Eds.), *Handbook of psychotherapy integration* (pp. 94-129). New York, NY: Basic Books.

Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72-113). New York, NY: Wiley.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*(2), 159-172. doi: 10.1037/0022-006X.69.2.159

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). Chicago, IL: John Wiley & Sons.

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. doi: 10.1080/10503300601032506

Landau, S. (2002). Using survival analysis in psychology. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *1*(4), 233-270. doi: 10.1207/S15328031US0104_03

Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: An application of random coefficient regression models to depicting the response of a patient to outpatient

psychotherapy. *Journal of Consulting and Clinical Psychology, 67*(4), 571-577. doi: 10.1037/0022-006X.67.4.571

McCord, C. E., Elliott, T. R., Wendel, M. L., Brossart, D. F., Cano, M. A., Gonzalez, G. E., & Burdine, J. N. (2011). Community capacity and teleconference counseling in rural Texas. *Professional Psychology: Research and Practice, 42*(6), 521-527. doi: 10.1037/a0025296

McCord, C. E., Saenz, J. J., Armstrong, T. W., & Elliott, T. R. (2015). Training the next generation of counseling psychologists in the practice of telepsychology. *Counselling Psychology Quarterly, 28*(3), 324-344. doi: 10.1080/09515070.2015.1053433

Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76*(1), 116-124. doi: 10.1037/0022-006X.76.1.116

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81-117. doi: 10.2333/bhmk.29.81

Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics, 30*(1), 27-58. doi: 10.3102/10769986030001027

Muthén, B., Asparouhov, T., Boye, M., Hackshaw, M., & Naegeli, A. (2009). Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. Technical Report. Retrieved from http://www.statmodel.com/download/lilyFinalReportV6.pdf

Muthén, L. K., & Muthén, B. O. (2017). Mplus (Version 8). Los Angeles, CA: Muthén & Muthén.

National Advisory Committee on Rural Health. (1993). *Sixth annual report on rural health*. Rockville, MD: U.S. Department of Health and Human Services, Health Resources and Services Administration, Office of Rural Health Policy.

Nelson, E., & Velasquez, S. E. (2011). Implementing psychological services over televideo. *Professional Psychology: Research and Practice, 42*(6), 535-542. doi: 10.1037/a0026178

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241-301. doi:10.1037/1082–989X.5.2.241

Peduzzi, P., Holford, T., Detre, K., & Chan, Y. K. (1987). Comparison of the logistic and Cox regression models when outcome is determined in all patients after a fixed period of time. *Journal of Chronic Diseases*, *40*(8), 761-767. doi: 10.1016/0021-9681(87)90127-5

Pimlott-Kubiak, S., & Cortina, L. M. (2003). Gender, victimization, and outcomes: Reconceptualizing risk. *Journal of Consulting and Clinical Psychology*, *71*(3), 528-539. doi: 10.1037/0022-006X.71.3.528

Probst, J. C., Moore, C. G., Glover, S., & Samuels, M. (2004). Person and place: The compounding effects of race/ethnicity and rurality on health. *American Journal of Public Health*, *94*(10), 1695-1703. doi: 10.2105/AJPH.94.10.1695

Raykov, T., Gorelick, P. B., Zajacova, A., & Marcoulides, G. A. (2017). On the potential of discrete time survival analysis using latent variable modeling: An application to the study of the vascular depression hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(6), 926-935. doi: 10.1080/10705511.2017.1315305

Richards, D., & Whyte, M. (2009). *Reach out: National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions* (2nd ed.). UK: Rethink.

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, *133*(1-2), 51-60. doi: 10.1016/j.jad.2011.03.024

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*(1), 1-12. doi: 10.1037/a0018326

Rychlak, J. F. (1968). Fundamental dimensions of theoretical orientation. *Philosophy of science for personality theory* (pp. 11-41). Boston, MA: Houghton, Mifflin and Company. doi: 10.1037/10535-000

Savalei, V., & Dunn, E. (2015). Is the call to abandon p-values the red herring of the replicability crisis? *Frontiers in Psychology*, 06 March 2015, Article 245. doi: 10.3389/fpsyg.2015.00245

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The consumer reports study. *American Psychologist*, *50*(12), 965-974. doi: 10.1037/0003-066X.50.12.965

Singh, R., & Mukhopadhyay, K. (2011). Survival analysis in clinical trials: Basics and must know areas. *Perspectives in Clinical Research*, *2*(4), 145-148. doi: 10.4103/2229-3485.86872

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*, *282*(18), 1737-1744. doi: 10.1001/jama.282.18.1737

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D. A., ... & Hardy, G. E. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Practice-based evidence. *Journal of Consulting and Clinical Psychology*, *71*(1), 14-21. doi: 10.1037/0022-006X.71.1.14

Straker, M. (1968). Brief psychotherapy in an outpatient clinic: Evolution and evaluation. *American Journal of Psychiatry*, *124*(9), 1219-1226. doi: 10.1176/ajp.124.9.1219

Stulz, N., Lutz, W., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology*, *75*(6), 864-874. doi: 10.1037/0022-006X.75.6.864

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (1997). *Measuring patient changes in mood, anxiety, and personally disorders: Toward a core battery.* Washington, DC, US: American Psychological Association. doi: 10.1037/10232-000

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive–behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*(6), 894-904. doi: 10.1037/0022-006X.67.6.894

Tarlow, K. R., McCord, C. E., Elliott, T. R., & Brossart, D. F. (2014). Health-related quality of life of rural clients seeking telepsychology services. *International Journal of Telemedicine and Applications*, *2014*, Article ID 168158, 7 pages. doi: 10.1155/2014/168158

Torres, L., & Ong, A. D. (2010). A daily diary investigation of Latino ethnic identity, discrimination, and depression. *Cultural Diversity and Ethnic Minority Psychology*, *16*(4), 561-568. doi: 10.1037/a0020652

Trust for America's Health. (2014). *Key health data about Texas.* Retrieved from http://healthyamericans.org/states/?stateid=TX

Tucker, J. A., & Reed, G. M. (2008). Evidentiary pluralism as a strategy for research and evidence-based practice in rehabilitation psychology. *Rehabilitation Psychology*, *53*(3), 279-293. doi: 10.1037/a0012963

Waskow, I.. E., & Parloff, M. B. (1975). *Psychotherapy change measures.* (DHEW Publication No. ADM No. 74-1207). Washington, DC: U.S. Government Printing Office.

Wendel, M., Brossart, D., Elliott, T., McCord, C., & Diaz, M. (2011). Use of technology to increase access to mental health services in a rural Texas community. *Family & Community Health*, *34*(2), 134-140. doi: 10.1097/FCH.0b013e31820e0d99

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, *38*(5), 343-352. doi: 10.3102/0013189X09339056

APPENDIX A

BASIC SURVIVAL ANALYSIS METHODOLOGY*

The Survival Function

The *survival function* $S(t)$ is defined as the probability that an individual's survival time, T, is greater than t, that is,

$$S(t) = \Pr(T > t). \tag{A1}$$

The graph of $S(t)$ against t is known as the *survival curve*. Two hypothetical examples are shown in Figure A1. The gradual decline of Survivor Curve 2 indicates longer survival times than the steeper decline of Survivor Curve 1.

The survival curve can be thought of as a particular way of displaying the frequency distribution of the event times, rather than by, say, a histogram. When there are no censored observations in the sample of survival times, the survival function can be estimated as

$$S(t) = \frac{\text{number of participants surviving longer than } t}{\text{total number of participants}} \tag{A2}$$

Because every participant is "alive" at the beginning of the study, and nobody is observed to survive longer than the largest of the observed survival times, then,

$$S(0) = 1 \quad \text{and} \quad S(t_{\max}) = 0 \tag{A3}$$

* Reprinted in its entirety (including all figures and tables) with permission from "Using survival analysis in psychology" by Landau, S., 2002. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1*(4), 233-270, Copyright 2002 by Taylor & Francis Group. www.tandfonline.com.

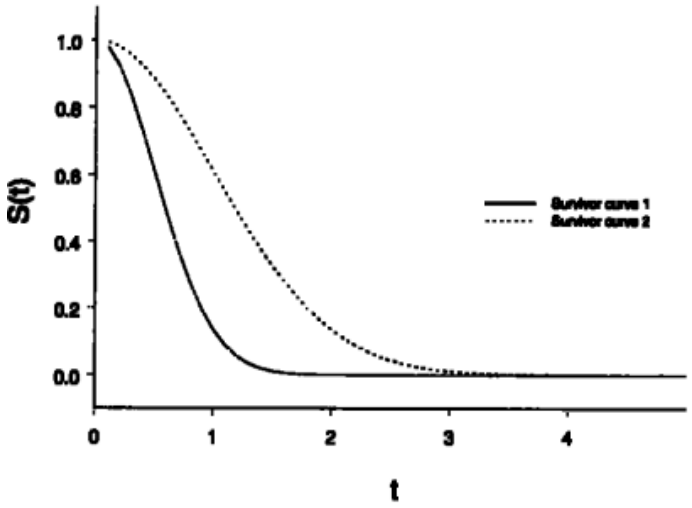

Figure A1.
Two Theoretical Survivor Curves.

The estimator in Equation A2 is simply a proportion, so that the confidence intervals can be obtained for each time t by using the variance estimate

$$\hat{\text{var}}\left[S(t) \right] = S(t)\left[1 - S(t) \right] / n, \tag{A4}$$

Where n is the total number of participants.

This simple method of estimating the survival function can be used only if all the individuals are followed up until the particular event of interest has happened to each. A number of methods are available for estimating the survival function for survival data containing censored observations, of which the most common is the *Kaplan-Meier* or *product limit estimator* method (Kaplan & Meier, 1958). The essence of this approach is the use of a product of a series of conditional probabilities. This involves ordering the r sample "death" times from the smallest to the largest such that

$$t_{(1)} \le t_{(2)} \le \cdots \le t_{(r)} \tag{A5}$$

Then the survival curve is estimated from the formula

$$S(t) = \prod_{j | t_{(j)} \le t} \left(1 - \frac{d_j}{r_j} \right), \tag{A6}$$

where r_j is the number of individuals at risk at $t_{(j)}$ and d_j is the number experiencing the event of interest at $t_{(j)}$. (Individuals censored at $t_{(j)}$ are included in r_j.)

A small example will help to clarify the estimation procedure, and for this the data in Table A1 are used.

- $S(12.8)$ is the proportion of participants surviving longer than 12.8 weeks (or one minus the opposite event of not surviving longer than 12.8 weeks) and is estimated by

$$S(12.8) = P_1 = \left(1 - \frac{1}{20}\right) = 0.95$$

- $S(15.6)$ is the proportion of participants surviving longer than 15.6 weeks and can be estimated by the product of the proportion surviving 12.8 weeks and the proportion of participants surviving longer than 15.6 weeks given they have survived 12.8 weeks, P_2. The estimate becomes

$$S(15.6) = P_1 \times P_2 = S(12.8) \times \left(1 - \frac{1}{20}\right) = 0.9$$

- Up to this point the estimates are identical to those given by the method described earlier for data containing no censored observations. The estimate for the next event time point, however, has to take into account the elimination of the observation censored at Week 24 from the participants at risk at 26.4 weeks.

Table A1.
Demonstration of Kaplan-Meier Estimation of Survival Probabilities.

*Time**	*Status*	*Death Time* (t_j)	d_j	*Survival Probability* $S(t)$	*SE* $\sqrt{\hat{V}ar(S(t))}$	*Cumulative Events*	*No. Remaining* $r_j - d_j$	*No. at risk* r_j
12.8	1	12.8	1	.950	.049	1	19	20
15.6	1	15.6	1	.900	.067	2	18	19
24.0	0					2	17	
26.4	1	26.4	1	.847	.081	3	16	17
26.7	0					3	15	
27.1	1	27.1	1	.790	.094	4	14	15
28.0	1	28.0	1	.734	.103	5	13	14
29.7	1	29.7	1	.678	.109	6	12	13
30.2	0					6	11	12
30.3	1	30.3	1	.616	.115	7	10	11
30.4	0					7	9	
35.1	0					7	8	
36.7	1	36.7	1	.539	.124	8	7	8
37.0	1	37.0	1	.462	.128	9	6	7
39.5	1	39.5	1	.385	.128	10	5	6
42.7	1	42.7	1	.308	.123	11	4	5
45.8	1	45.8	1	.231	.114	12	3	4
54.3	1	54.8	1	.154	.099	13	2	3
61.2	1	61.2	1	.077	.074	14	1	2
76.0	1	76.0	1	.000	.000	15	0	1

Note. $n = 20$. *Given in weeks.

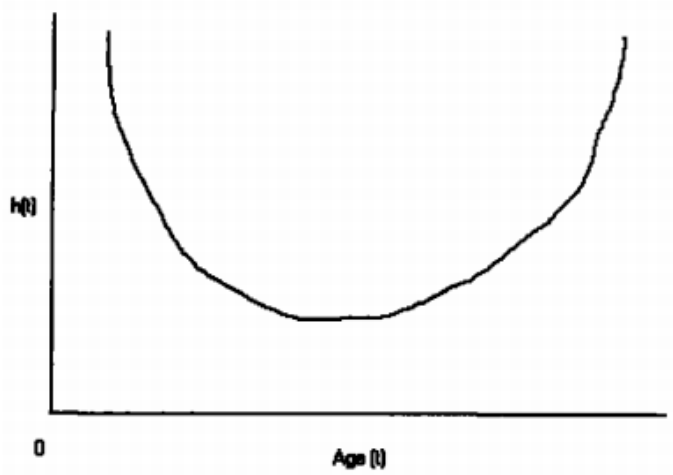

Figure A2.
Demonstration of Kaplan-Meier Estimates of Survival Curve for Example Latency Data.

Applying equation A6 gives

$$S(26.4) = P_1 \times P_2 \times P_3 = S(15.6) \times \left(1 - \frac{1}{19}\right) = 0.85$$

Similar calculation for the remaining data lead eventually to the estimated curve shown in Figure A2. The curve is a step function with decreases at the time points when events were observed. The censored observations in the data are indicated by the "cross" marks on the curve.

The variance of this estimator of the survival curve can itself be estimated from the following formula:

$$\hat{\text{var}}\left[S(t)\right] = \left[S(t)\right]^2 \sum_{j|t_{(j)} \le t} \frac{d_j}{r_j\left(r_j - d_j\right)} \qquad \text{(A7)}$$

This formula has been used to obtain standard errors of survival probabilities in Table A1.

Comparing Survival Functions

The estimated survival curve for a sample of survival times provides a useful description of the distribution and main characteristics of these times (e.g. the median latency), but generally it is the comparison of the survival curves of different groups of participants that is of most interest. Plotting the estimated survival functions of different groups of individuals is a good first step in comparing the survival experience of the groups, but a more formal means of making this comparison is often required, so that the hypothesis $H_0 : S_1 = S_2$ can be tested, where S_1 and S_2 are the population survival curves of the two groups.

In the absence of censored observations, standard nonparametric tests might be used to compare the survival times of each group (or even an independent-samples *t* test, if the distributions were approximately normal). When the data do contain censored observations, there are a number of modified tests, both parametric and nonparametric, that can be used. Here the most commonly used of these—namely the log-rank test—is described (Peto & Peto, 1972). This nonparametric test compares the observed number of "deaths" at each particular time point with the number to be expected if the survival experience of the two groups is the same.

The log-rank test is first illustrated using the small data set shown in Table A2. First a series of 2 x 2 tables are constructed giving the number of individuals dying and the number remaining alive and at risk (i.e., not censored) for different times at which "deaths" occur. This is illustrated for the first two death times in Table A3.

Table A2.
Example Data Set: Survival Times in Two Groups.

Participant	*Time*	*Status*	*Group*
1	5	1	2
2	6	1	2
3	8	1	2
4	9	1	1
5	10	1	2
6	12	1	2
7	13	1	1
8	14	0	1
9	16	0	2
10	18	1	1
11	23	1	1
12	24	1	2
13	27	1	2
14	28	0	1
15	30	1	2
16	31	1	1
17	33	1	2
18	34	1	1
19	43	1	2
20	45	0	1
21	46	1	2
22	48	1	1
23	161	0	1

Note. $n = 23$.

Table A3.

Demonstration of Log Rank Test for Data in Table A2.

	No. of Deaths at Time		No. Surviving Beyond Time		No. at Risk Just Before Time	
Group	5	6	5	6	5	6

(a) Examples of cross tabulations at failure times

Time of first death $t_{(1)} = 5$

1	$d_{11} = 0$	$r_{11} - d_{11} = 11$	$r_{11} = 11$
2	$d_{21} = 1$	$r_{21} - d_{21} = 11$	$r_{21} = 12$
Total	$d_1 = 1$	$r_1 - d_1 = 22$	$r_1 = 23$

Time of second death $t_{(2)} = 6$

1	$d_{12} = 0$	$r_{12} - d_{12} = 11$	$r_{12} = 11$
2	$d_{22} = 1$	$r_{22} - d_{22} = 10$	$r_{22} = 11$
Total	$d_2 = 1$	$r_2 - d_2 = 21$	$r_2 = 22$

j	*Death Time* $t_{(j)}$	*No. Deaths in Group 1* d_{1j}	*No. at risk in Group 1* r_{1j}	*No. Deaths in Group 2* d_{2j}	*No. at risk in Group 2* r_{2j}	*No. Deaths* d_j	*No. at Risk* r_j	*Expected No. Deaths in Group 1* e_{1j}	v_{1j}
(b) Quantities obtained from cross tabulations									
1	5	0	11	1	12	1	23	0.478	0.250
2	6	0	11	1	11	1	22	0.500	0.250
3	8	0	11	1	10	1	21	0.524	0.249
4	9	1	11	0	9	1	20	0.550	0.248
5	10	0	10	1	9	1	19	0.526	0.249
6	12	0	10	1	8	1	18	0.556	0.247
7	13	1	10	0	7	1	17	0.588	0.242
8	18	1	8	0	6	1	14	0.571	0.245
9	23	1	7	0	6	1	13	0.538	0.249
10	24	0	6	1	6	1	12	0.500	0.250
11	27	0	6	1	5	1	11	0.545	0.248
12	30	0	5	1	4	1	9	0.556	0.247
13	31	1	5	0	3	1	8	0.625	0.234
14	33	0	4	1	3	1	7	0.571	0.245
15	34	1	4	0	2	1	6	0.667	0.222
16	43	0	3	1	2	1	5	0.600	0.240
17	46	0	2	1	1	1	3	0.667	0.222
18	48	1	2	0	0	1	2	1.000	0.000
Total		7						10.562	4.137

(c) Calculation of test statistic

$$\chi^2 = (O_1 - E_1)^2 / \hat{\text{v}}\text{ar}(O_1) = \left[\sum_{j=1}^{r} d_{1j} - \sum_{j=1}^{r} e_{1j}\right]^2 / \sum_{j=1}^{r} v_{1j}$$

$$= \frac{(7 - 10.562)^2}{4.137}$$

$$= 12.688 / 4.137 = 3.067 ,$$

on 1 *df*, resulting $p = .08$.

Assuming the distribution of survival times is identical for both Group 1 and Group 2, the relevant expected values for each table can be calculated in the usual way as when testing for independence in a 2 x 2 contingency table. The expected numbers of deaths in Group 1 at failure times are included in Table A3. The observed and expected numbers of deaths in Group 1 are then summed to give the total numbers shown in Table A3. The totals can be compared using the following test statistic:

$$\chi^2 = (O_1 - E_1)^2 / \hat{\text{var}}(O_1) \qquad \text{(A8)}$$

The variance estimate of the total number of observed deaths in Group 1, $\hat{\text{var}}(O_1)$, can be calculated as the sum of the variance estimates of the observed number of deaths in Group 1 at each time point v_{1j}. The latter have also been included in Table A3. Here the test statistic becomes $\chi^2 = 3.07$ (see Table A3). Under the null hypothesis that the survival distributions in Groups 1 and 2 are the same, χ^2 has, approximately, a chi-square distribution with a single degree of freedom. For our small sample data set a trend toward a difference in survival times between groups is detected $(p = .08)$. (Here Group 1 was chosen to calculate the value of the test statistic; the choice of group is arbitrary and does not affect the value of the test statistic.)

The log-rank test can be used to compare survival times in more than two groups, and its general formulation for the *K* group situation is as follows:

- For each failure time $t_{(j)}$ and group $k = 1,...,K$, let r_{kj} denote the number of participants at risk and d_{kj} denote the number of deaths. The total number of individuals at risk at time $t_{(j)}$ is then $r_j = \sum_k r_{kj}$ and the total number of deaths $d_j = \sum_k d_{kj}$.

- If the survival experience is the same in all groups, the estimated probability of a person in group k dying at time t_j given that the individual is still alive is d_j / r_j and the expected number of deaths in each group is

$$e_{kj} = \frac{r_{kj} d_j}{r_j}, k = 1,\ldots,K. \qquad \text{(A9)}$$

- To define the test statistic, some matrix notation needs to be introduced. Let $d_j = \left[d_{1j},\ldots,d_{k-1,j} \right]^{\mathrm{T}}$ denote the $(K-1)$-vector of observed deaths at time j in Groups 1, ... , K-1 and $e_j = \left[e_{1j},\ldots,e_{k-1,j} \right]^{\mathrm{T}}$ the vector of respective expected deaths. Furthermore, let $V_j = \hat{cov}(d_j)$ denote the estimated $(K-1)\times(K-1)$ covariance matrix of the observed numbers of deaths at the jth failure time. Then the test statistic is given by

$$\chi^2 = \left[\sum_{j=1}^{r}(d_j - e_j)\right]^{\mathrm{T}}\left[\sum_{j=1}^{r}V_j\right]^{-1}\left[\sum_{j=1}^{r}(d_j - e_j)\right] \qquad \text{(A10)}$$

(The elements of V_j can be derived from a multivariate hypergeometric distribution. This was deliberately refrained from here; for details see Hosmer & Lemeshow, 1999. Also note that the choice of the first groups $K-1$is for convenience. Any other subset $K-1$groups produces the same value for χ^2.)

- Under the null hypothesis that the distribution of survival times is the same in each group, this statistic has, approximately, a chi-square distribution with $K-1$degrees of freedom.
- For the two-groups case $K=2$ the test statistic simplifies to

$$\chi^2 = \frac{\left[\sum_{j=1}^{r}(d_{1j} - e_{1j})\right]^2}{\sum_{j=1}^{r}v_{1j}} \qquad \text{(A11)}$$

Where

$$v_{1j} = \frac{r_{1j}\left(r_j - r_{1j}\right)d_j\left(r_j - d_j\right)}{r_j^2\left(r_j - 1\right)}$$

on 1 degree of freedom.

The log-rank test can be generalized to give differential weights to the failure times. The generalized Wilcoxon test (or Breslow test; Breslow, 1970) uses weights equal to the number at risk and therefore puts relatively more weight on differences between the survival curves at smaller values of time. The log-rank test, which uses weights equal to 1 at all time points, places more emphasis on differences at larger values of time. The *Tarone-Ware* test (Tarone & Ware, 1977) uses a weight function intermediate between those, namely $w_j = \sqrt{r_j}$. For the example data in Table A2 the generalized Wilcoxon test, $\chi^2(1) = 2.3, p = .13$, and the Tarone-West test, $\chi^2(1) = 2.59, p = .11$, would not detect a group difference in survival times indicating that large time points contribute to the difference detected by the log-rank test.

The Hazard Function

In the analysis of survival data it is often of some interest to assess which periods have the highest and which the lowest chance of death (or whatever the event of interest happens to be), among those people alive at the time. In the very old, for example, there is a high risk of dying each year among those entering that stage of their life. The probability of any individual dying in their 100th year is, however, small because so few individuals live to be 100 years old.

The appropriate quantity for such risks is the *hazard function*, $h(t)$, defined as the probability that an individual experiences an event in a small time interval s, given that the individual has survived up to the beginning of the interval. The hazard function therefore represents the instantaneous death rate for an individual surviving to time t. it is a measure of how likely an individual is to experience an event as a function of the age of the individual. The hazard function may remain constant, increase or decrease with time, or take some more complex form. The hazard function of death in human beings, for example, has the typical "bathtub" shape shown in Figure A3. The hazard function is relatively high immediately after birth, declines rapidly in the early years, and then remains pretty much constant until it begins to rise during late middle age.

In formal, mathematical terms, the hazard function is defined as the following limiting value:

$$h(t) = \frac{lim}{s \to 0}\left[\frac{Prob(event\,in\,t, t+s \,|\, \text{event time} > t)}{s}\right] \qquad \text{(A12)}$$

For a sample of survival data, the hazard function can be estimated as the proportion of individuals experiencing an event in an interval per unit of time, given that they have survived to the beginning of the interval, that is:

$$h(t) = \frac{\text{no. of individuals experiencing an event in the interval beginning at time } t}{(\text{no. of individuals surviving at } t)(\text{interval width})} \qquad \text{(A13)}$$

Figure A3.
Schematic Hazard Function of Death in Human Beings.

More formally, the hazard function in the intervals between failure times, $t_{(j)} \le t \le t_{(j+1)}$, can be estimated using the ratio

$$h(t) = \frac{d_j}{r_j(t_{(j+1)} - t_{(j)}} \tag{A14}$$

The hazard function is related to the survival curve by the relation

$$S(t) = exp\left[-\int_0^t h(x)\,dx \right] \tag{A15}$$

The integral term is known as the *integrated hazard* and is, in general, more useful than the hazard function itself, which is rarely plotted because it is simply too "noisy." The integrated hazard, $H(t)$, can be estimated as

$$H(t) = -\sum_{j|t_{(j)} \le t} \ln\left(\frac{r_j - d_j}{r_j} \right) \tag{A16}$$

The Proportional Hazards Model—Cox Regression

The previous sections looked at ways of summarizing and plotting survival data as well as a simple test for comparing the survival experience of different groups. In this section, modeling survival times when there are several explanatory variables of interest is discussed. The main technique is due to Cox (1972) and known as the proportional hazards model or, more simply, Cox's regression. In essence, the technique acts as the analogue of multiple regression for survival times containing censored observations, for which multiple regression for survival times containing censored observations, for which multiple regression itself is clearly not suitable. In a Cox regression it is the hazard function that is modeled. Central to the procedure is the assumption that the hazard functions for two individuals at any point in time are proportional. In other words, if an individual has a risk of death at some initial time point that is twice as high as that for another individual, then at all later times the risk of death remains twice as high. Cox's model is made up of an unspecified baseline hazard function, $\alpha(t)$, which is then multiplied by a suitable function of an individual's explanatory variable values, to give the individual's hazard function. Formally, for a set of *p* explanatory variables, $x_1, x_2, \ldots, x_p$ the model is

$$h(t) = \alpha(t) exp\left[\beta_0 + \sum_{i=1}^{p} \beta_i X_i \right] \tag{A17}$$

where the terms $\beta_0, \beta_1, \ldots, \beta_p$ are the parameters of the model that have to be estimated from sample data.

Consider two individuals, with covariate values $x_{11}, x_{12}, \ldots, x_{1p}$ and $x_{21}, x_{22}, \ldots, x_{2p}$; the ratio of their hazards, $h_1(t)$ and $h_2(t)$, under this model becomes

$$\frac{h_1(t)}{h_2(t)} = \frac{\alpha(t)exp\left[\beta_0 + \sum_{i=1}^{p}\beta_i x_{1i}\right]}{\alpha(t)exp\left[\beta_0 + \sum_{i=1}^{p}\beta_i x_{2i}\right]} = exp\left[\sum_{i=1}^{p}\beta_i(x_{1i} - x_{2i})\right] \qquad \text{(A18)}$$

The ratio does not depend on t. the interpretation of the parameter β_i is that $\exp(\beta_i)$ gives the relative risk change associated with an increase of one unit in x_i, all other explanatory variables remaining constant. Specifically, in the case of comparing hazards between two groups, $\exp(\beta)$ measures the hazard ratio between the two groups.

Cox's regression is considered a semiparametric procedure because the baseline hazard function, $\alpha(t)$, and, by implication, the probability distribution of the survival times, does not have to be specified explicitly. The estimation process relies only on the order in which events occur, not the exact times they occur. Details of procedures for parameter estimation in a Cox model can be found in Kalbfleisch and Prentice (1980). Parameter estimates are derived at by assuming continuous survival times. In most applied settings, however, event times are measured in discrete units, and there are often ties. Methods for dealing with such ties were described by Hosmer and Lemeshow (1999).

Checking Assumptions of a Cox Regression

As in multiple regression, residuals play a key role in assessing the adequacy of a model. A number of different residuals have been proposed for use with the Cox model, and a detailed review of these and further methods for checking model assumptions can be found in Hosmer and Lemeshow (1999). Here two types of residuals commonly used, the *Martingale residuals* and the *Schoenfeld residuals*, are introduced. These are important practical tools because they allow assessment of the two main assumptions underlying a Cox regression, namely, that:

- The effect of the covariates is additive and linear on a log-hazard scale.
- The ratio of the hazards of two individuals is the same at all times.

Cox and Snell (1968) suggested the use of the Martingale residuals; for the *k*th participant the residual is defined as

$$r_k = c_k - H(t_k) \qquad \text{(A19)}$$

where

$$H(t_k) = H_0(t_k) exp\left[\beta_0 + \sum_m \beta_m x_{km} \right]$$

Here $C_k = 1$ for uncensored observations and zero otherwise. $H_0(t_k)$ is an estimator of the baseline integrated hazard function at the failure time of the *k*th participant. This estimator is not based on the Kaplan-Meier estimator of the survival curve but uses the proportional hazards assumption instead (for details, see again Hosmer & Lemeshow, 1999).

Martingale residuals take values between $-\infty$ and unity, with the residuals for censored observations being negative. For large samples they are uncorrelated with each other and have a mean of zero. However, the residuals are not symmetrically distributed around zero; rather, the Cox-Snell residuals $r_k^* = c_k - r_k$ have an exponential distribution with a mean of 1 if the fitted model is correct, regardless of the actual distributional form of $S(t_k)$. Covariate-specific Martingale residuals can be constructed by calculating the residuals from a model in which the effect of the covariate of interest is set to zero. Therneau, Grambsch, and Fleming (1990) suggested plotting these residuals against the covariate. A smooth curve of the Martingale residuals (e.g., the lowess smooth) then provides an estimate of the functional form of the covariate in the model.

Schoenfeld (1982) proposed another set of residuals for use with a fitted proportional hazards model. For the *k*th participant and the *i*th covariate the residual is defined as

$$r_{ki}^{**} = c_k \left(x_{ki} - \bar{x}_{ki} \right) \qquad \text{(A20)}$$

Where

$$\bar{x}_{ki} = \frac{\sum_{j \in R(t_k)} x_{ji} \exp\left(\beta_0 + \sum_m \beta_m x_{jm} \right)}{\sum_{j \in R(t_k)} \exp\left(\beta_0 + \sum_m \beta_m x_{jm} \right)}$$

$R(t_k)$ is the set of individuals at risk at failure time t_k. Software packages set the value of the residual to missing for participants whose observed survival time was censored. Scaled versions of this residual are available that aim to standardize its variance (e.g., see Grambsh & Therneau, 1994).

The scaled Schoenfeld residuals of the *i*th covariate can be plotted against a function of time to assess the proportional hazards assumption for covariate i. Under the proportional hazards assumption the effect of the covariate is constant over time, and the plot should show a horizontal line. Residuals are commonly plotted against log-time, $\ln(t)$.

References

Anderson, P. K. (1992). Repeated assessment of risk factors in survival analysis. *Statistical Methods in Medical Research,1,* 297-315.

Breslow, N. E. (1970). A generalized Kruskal-Wallace test for comparing *K* samples subject to unequal pattems of censorship. *Biometrika, 57,*579-594.
Collett, D. (1994). *Modelling survival data in medical research.* London: Chapman & Hall.
Cox, D. R. (1972). Regression models and life tables (with discussion). Journal *of the Royal Statistical Society: Series B, 34,*187-220.
Cox, D. R. & Snell, E. J. (1968). A general definition of residuals with discussion. *Journal of the Royal Statistical Society: Series A, 30,* 248-27S.
Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests in diagnostics based on weighted residuals. *Biometrika, 81,* 515-526.
Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis.* New York: Wiley.
Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data.* New York: Wiley.
Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association, 53,* 457-481.
Klein, J. P., & Moeschberger, M. L. (1997). *Survival analysis techniques far censored and truncated data.* New York: Springer.
Maier, H., & Smith, J. (1999). Psychological prediction of mortality in old age. *Journal of Gerontology, 54,* 44-54.
Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariance test procedures (with discussion). *Journal of the Royal Statistical Society: Series A, 135,*185-206.
Schmidt, U., Evans, K., Tiller, J., & Treasure, J. (1995). Puberty, sexual milestones and abuse: How are they related in eating disorder patients? *Psychological Medicine, 25,* 413-417.
Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression. *Biometrika, 69,* 239-241.
Tarone, R. E., & Ware, J. (1977). On distribution free tests of the equality of survival distributions. *Biometrika, 64,* 156-160.
Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model.* New York: Springer.
Therneau, T. M., Grambsch, P. M., & Fleming, T. R. (1990). Martingale-based residuals for survival model. *Biometrika, 77,* 147-160.
Wechsler. D. (1974). *Wechsler Intelligence Scale for Children-Revised.* New York: Psychological Corp.
Weisz, J. R., Walter, B. R., Weiss, B., Fernandez, G. A. & Mikow, V. A. (1990). Arrests among emotionally disturbed violent and assaultive individuals following minimal versus lengthy intervention through North Carolina's Willie M program. *Journal of Consulting and Clinical Psychology, 50,* 720-728.
Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology, 61,* 952-965.
Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971*). Manual for the Embedded Figures Tests.* Palo Alto, CA: Consulting Psychologists Press.

APPENDIX B

OUTCOME MEASURES*

The CORE-B SF/B and PHQ-9 scales are included for reference and are reproduced here without alteration.

* CORE outcome measures, copyright of CORE System Trust. Reprinted with permission of Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) licence. https://www.coresystemtrust.org.uk/. PHQ and PHQ-9 reprinted with permission of Pfizer Inc. Copyright Pfizer Inc.

CLINICAL OUTCOMES in ROUTINE EVALUATION (SFB)

Site ID

letters only | numbers only

Client ID

D D / M M / Y Y

Date Completed

Stage Completed

F First Therapy Session
D During Therapy
L Last Therapy Session

Session Number
(first *therapy* session = session 001)

SHORT FORM B

IMPORTANT -PLEASE READ THIS FIRST

This form has 18 statements about how you have been OVER THE LAST WEEK.
Please read each statement and think how often you felt that way last week.
Then tick the box which is closest to this.
Please use a dark pen (not pencil) and tick clearly within the boxes.

	Over the last week	Not at all	Only Occasionally	Sometimes	Often	Most or all the time	OFFICE USE ONLY
1	I have felt terribly alone and isolated	0	1	2	3	4	F
2	I have difficulty getting to sleep or staying asleep	0	1	2	3	4	P
3	I have felt optimistic about my future	4	3	2	1	0	W
4	I have felt totally lacking in energy and enthusiasm	0	1	2	3	4	P
5	I made plans to end my life	0	1	2	3	4	R
6	I have been troubled by aches, pains or other physical problems	0	1	2	3	4	P
7	I have been happy with the things I have done	4	3	2	1	0	F
8	Talking to people has felt too much for me	0	1	2	3	4	F
9	I have felt OK about myself	4	3	2	1	0	W
10	Tension and anxiety have prevented me doing important things	0	1	2	3	4	P
11	I have been disturbed by unwanted thoughts and feelings	0	1	2	3	4	P
12	I have felt overwhelmed by my problems	0	1	2	3	4	W
13	I have felt I have someone to turn to for support when needed	4	3	2	1	0	F
14	I have felt like crying	0	1	2	3	4	W
15	I have threatened or intimidated another person	0	1	2	3	4	R
16	I have been able to do most things I needed to	4	3	2	1	0	F
17	I have thought I have no friends	0	1	2	3	4	F
18	I have thought I am to blame for my problems and difficulties	0	1	2	3	4	P

Total Scores

Mean Scores
(Total score for each dimension divided by number of items completed in that dimension)

(W)	(P)	(F)	(R)	All items	All minus R

PHQ-9 — Nine Symptom Checklist

Patient Name _____ **Date** _____

1. Over the last 2 weeks, how often have you been bothered by any of the following problems? Read each item carefully, and circle your response.

   a. Little interest or pleasure in doing things
   Not at all **Several days** **More than half the days** **Nearly every day**

   b. Feeling down, depressed, or hopeless
   Not at all **Several days** **More than half the days** **Nearly every day**

   c. Trouble falling asleep, staying asleep, or sleeping too much
   Not at all **Several days** **More than half the days** **Nearly every day**

   d. Feeling tired or having little energy
   Not at all **Several days** **More than half the days** **Nearly every day**

   e. Poor appetite or overeating
   Not at all **Several days** **More than half the days** **Nearly every day**

   f. Feeling bad about yourself, feeling that you are a failure, or feeling that you have let yourself or your family down
   Not at all **Several days** **More than half the days** **Nearly every day**

   g. Trouble concentrating on things such as reading the newspaper or watching television
   Not at all **Several days** **More than half the days** **Nearly every day**

   h. Moving or speaking so slowly that other people could have noticed. Or being so fidgety or restless that you have been moving around a lot more than usual
   Not at all **Several days** **More than half the days** **Nearly every day**

   i. Thinking that you would be better off dead or that you want to hurt yourself in some way
   Not at all **Several days** **More than half the days** **Nearly every day**

2. If you checked off any problem on this questionnaire so far, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

   Not Difficult at All **Somewhat Difficult** **Very Difficult** **Extremely Difficult**