EVALUATION OF TELEHEALTH VIDEOCONFERENCING PSYCHOTHERAPY IN RURAL PRIMARY CARE: A MIXED-METHODS RESEARCH STUDY

A Dissertation

by

GERARDO E. GONZALEZ

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,	Daniel Brossart
Committee Members,	Timothy Elliott
	William Rae
	Arnold LeUnes
Head of Department,	Victor Willson

August 2014

Major Subject: Counseling Psychology

ABSTRACT

The effectiveness of telehealth videoconferencing psychotherapy (TVCP) for a rural sample obtaining services through a primary care setting in Texas was examined by combining single-case and group research methods. Treatment-as-usual was delivered via TVCP to 41 patients for an average of 11 sessions (*SD* = 7.79) by doctoral level psychology students under supervision of licensed psychologists. Patients were assessed periodically over the course of treatment with the Clinical Outcomes in Routine Evaluation – Short Form B (CORE-B), Patient Health Questionnaire – 9 (PHQ-9), and SF-12 health survey. Group analyses included bootstrapped paired-samples *t* tests of pre- and post-treatment mean scores for all outcome variables. Reliable improvement (Improved) and clinically significant change (Recovered) was assessed for all patients on the CORE-B scales and the PHQ-9. Single-case analyses of four female patients included visual ratings of graphed CORE-B Global Distress scores and simple mean shift regressions of all CORE-B scales. Results of single-case analyses were compared with group results to uncover clinical insights.

TVCP produced statistically significant results on all mental health outcomes for the group despite declines in perceived physical health quality. On the CORE-B Global Distress scale, 27% of patients Improved and 32% Recovered. On the CORE-B subscales, a large percentage of patients made reliable improvements and clinically significant change in Risk (58% Improved, 8% Recovered), Well-Being (13% Improved, 52% Recovered), Symptoms (33% Improved, 29% Recovered), and Functioning (24% Improved, 33% Recovered). On the PHQ-9, 46% of patients Improved and 23%

Recovered. Single-case analyses of four female patients provided a more differentiated representation of treatment response and context for group results. Comparison of single-case and group results suggested treatment response was dependent upon type and severity of diagnoses, severity of physical health issues, and situational context. Clinical and methodological conclusions of the study were discussed with implications for scientists and practitioners.

DEDICATION

To my Mother and Father,

For your unconditional love, patience, and support throughout my life…

I am truly blessed and eternally grateful.

This is for you.

I love you.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Brossart, and my committee members, Dr. Elliott, Dr. Rae, and Dr. LeUnes, for their guidance, patience, and support throughout the course of this research. I would especially like to extend my gratitude to Dr. Brossart and Dr. Elliott for their constant support over the years and for giving me the opportunity to be involved in this project. I also want to extend my gratitude to Gloria McCarty and the Leon Health Resource Center staff who provided invaluable support in coordinating the daily functioning of the clinic and facilitating care for our patients. I also want to acknowledge Kirsten Salerno, my colleague and friend, for sharing a critical role in this project and for her encouraging conversations throughout our work together.

Thank you to all my family and friends who provided me with emotional support throughout the last four years. You truly made my experience special and enjoyable. I could not have done this without you.

To my wife, Nomi, thank you for coming into my life and walking with me every step of the way. To my sister, Flory, thank you for always believing in me and being there for me when things were difficult. To my daughter, Pepper, thank you for the years of being by my side providing endless moments of joy and love. I love you all from the bottom of my heart.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

CHAPTER I INTRODUCTION 1

Evaluating Psychotherapy Effectiveness by Combining Research Methods 4
Problem Statement 8
Purpose of the Study 11
Research Questions 12

CHAPTER II LITERATURE REVIEW 13

Rural Populations 13
Defining "Rural" 13
Context and Barriers to Rural Mental Health 15
Rural Mental Health Issues 17
History of Behavioral Telehealth in Rural Mental Health 18
Telehealth Videoconferencing Psychotherapy (TVCP) 20
TVCP Outcome Research with Rural Samples 20
Methodological Issues in Psychotherapy Research 31
Case Studies 32
Group Research 33
Single-Case Experimental Studies 37
Mixed-Method: Single-Case and Group Designs 42
Group Comparison Designs with Case-Based Studies 43
Group Comparison Designs with Single-Case Experimental Studies 45

CHAPTER III METHOD 48

Page

Study Methods..... 48
Context..... 48
Participants..... 49
Treatment..... 52
Outcome Measures..... 53
Procedures..... 57
Design..... 58
Analyses..... 59

CHAPTER IV RESULTS..... 65

Missing Data..... 65
Group Results..... 65
Summary of Group Results:
"The Average Patient"..... 76
Single-Case Results..... 77
Comparison of Single-Case and Group Results..... 85

CHAPTER V CONCLUSIONS..... 92

Summary of Results..... 92
Design Considerations and Limitations..... 102
Implications..... 105

REFERENCES..... 108

APPENDIX NOMENCLATURE..... 132

LIST OF FIGURES

FIGURE		Page
1	Study timeline, measurement schedule, and flow of participants through the study	59
2	Paired-samples *t* tests results for PHQ-9 (0-27), CORE-B Global Distress scale (0-40), MCS-12 (0-100), and PCS-12 (0-100)	67
3	Paired-samples *t* tests results for CORE-B subscales (0-40): Well-Being, Symptoms, Functioning, and Risk	68
4	Single-case design graphs of CORE-B Global Distress scale (0-40) for four female patients	79
5	Comparison of treatment response for the group and four patients on CORE-B scales as standardized mean difference effect size, Cohen's *d*	87

LIST OF TABLES

TABLE		Page
1	Demographics of patients receiving treatment	50
2	Paired-samples *t* tests results for PHQ-9, CORE-B Global Distress scale, MCS, and PCS	66
3	Paired-samples *t* tests results for CORE-B subscales	67
4	Clinically Significant and Reliable Change – PHQ-9	69
5	Clinically Significant and Reliable Change – CORE-B Global Distress scale	70
6	Clinically Significant and Reliable Change – CORE-B Risk subscale	71
7	Clinically Significant and Reliable Change – CORE-B Well-Being subscale	72
8	Clinically Significant and Reliable Change – CORE-B Symptoms subscale	73
9	Clinically Significant and Reliable Change – CORE-B Functioning subscale	74

CHAPTER I

INTRODUCTION

The availability of professional mental health providers in rural areas across the country is sparse (Smalley et al., 2010). It has been estimated that more than 85% of Mental Health Professional Shortage Areas (MHPSAs) are found in rural areas (Bird, Dempsey, & Hartley, 2001). In fact, more than half of all the counties in the United States (U.S.) do not have a psychologist, psychiatrist, or social worker available to provide mental health care to rural residents (American Psychological Association [APA], 2002; National Advisory Committee on Rural Health, 1993). These are troubling statistics given that approximately 20% of the U.S. population, or about 55 million people, live in rural areas (Health Resources and Services Administration [HRSA], 2005).

The problem posed by mental health provider shortages is further compounded by other barriers facing rural populations, including limited accessibility to services due to high poverty rates, insufficient means of transportation, minimal insurance coverage, and poor health (Gamm, Stone, & Pittman, 2010; Stamm et al., 2003; Wagenfeld, 2003). Furthermore, mental health treatment acceptance by rural residents may be impacted by a cultural stigma about obtaining psychological services and fear of decreased anonymity (HRSA, 2005; Letvak, 2002; Reese, Conoley & Brossart, 2006). These barriers may explain why rural residents are less likely to seek mental health care when compared to people in urban locations (Wagenfeld & Buffum, 1983) and why those rural residents who seek treatment enter care later and with more serious symptoms than

urban residents (Gamm, Stone, & Pittman, 2010; Rost, Fortney, Fischer, & Smith, 2002; Rost, Zhang, Fortney, Smith, & Smith, 1998; Wagenfeld, Murray, Mohatt, & DeBruyn, 1994).

Such barriers to mental healthcare treatment seem to contribute to the persistence of mental health disparities in rural populations (Gamm et al., 2010; Smalley et al., 2010). Approximately 34-41% of rural patients seen in primary care have a mental health disorder (Sears, Evans, & Kuper, 2003). One study of three rural primary care clinics found that psychological distress of patients accounted for functional impairment more than the severity of their chronic illnesses (Thurston-Hicks, Paine, & Hollifield, 1998). In fact, several studies have reported that the prevalence of mental health problems like depression, substance abuse, domestic violence, incest, child abuse, and suicide are serious risks to individuals in rural areas that often occur at equal or higher rates than in urban residents (Bushy, 1998; Cellucci & Vik, 2001; Eberhardt, Ingram, & Makuc, 2001; Eberhardt & Pamuk, 2004; Gamm et al., 2010; Rost et al., 1998).

Fortunately, telehealth services seem to be a viable mode of quality mental healthcare delivery to rural settings with barriers to much needed care (Backhaus et al., 2012; Elliott, Brossart, Berry, & Fine, 2008; McCord et al., 2011; Griffiths, Blignault, & Yellowlees, 2006; Reese et al., 2006; Richardson, Frueh, Grubaugh, Egede, & Elhai, 2009; Schopp, Demiris, & Glueckauf, 2006). Over 30 years ago, telehealth services were born as a means to attend to the needs of underserved populations in the United States (Jerome et al., 2000). *Telehealth* is defined as the "use of telecommunications and information technology to provide access to health assessment, diagnosis, intervention,

consultation, supervision, education, and information across distance" (Nickelson, 1998, p. 527). When telehealth is applied in the provision of psychological services, it is termed *behavioral telehealth* (Nickelson, 1998). *Telehealth videoconferencing psychotherapy* (TVCP) is one technological modality by which behavioral telehealth services can be delivered to individuals in remote locations without access to mental health services (Backhaus et al., 2012).

Recognition that behavioral telehealth services can help bridge gaps in mental health service disparities has resulted in strong federal support of such technology (Conrad, 1998; Wasem & Puskin, 2000). As one example, the President's New Freedom Commission on Mental Health (2003) recommended the "Use health technology and telehealth to improve access and coordination of mental health care, especially for Americans in remote areas or in underserved populations" (p. 25). This commission and other federal agencies have also recognized that mental health disparities for rural residents could be reduced by integrating behavioral health services into primary care settings (Office of the Surgeon General, 1999; National Association of State Medicaid Directors, 2008; Smalley et al., 2010). The Office of the Surgeon General (1999) recommended primary care settings as a point of entry which can provide opportunities for identifying and treating mental illness. In fact, some investigators have reported that about half of those people with depression are identified by primary care providers (Mitchell, Vaze, & Rao, 2009). As behavioral telehealth services expand and become integrated into primary care settings, there is an increased need to evaluate the effectiveness of this service modality for rural residents in primary care settings.

Empirical evaluation of behavioral telehealth and TVCP is still early in its development (Backhaus et al., 2012; Loberg, 2006). Early in the movement, behavioral telehealth research focused on patient and provider satisfaction (e.g. Reese et al., 2006) as well as differences in the therapeutic alliance between these technological approaches and face-to-face psychotherapy (Richardson et al., 2009). However, a recent review of TVCP funded by the National Institutes of Health stated that currently "there are few controlled efficacy and effectiveness studies and research explicitly conducted with racial minorities or people living in rural areas is minimal" (Richardson et al., 2009, p. 11). In effect, this report suggested that the evolution of behavioral telehealth is experiencing a trend toward addressing a recognized need for clinical outcome research which supports the effectiveness of TVCP in general, as well as with underserved groups like rural populations (Richardson et al., 2009).

Evaluating Psychotherapy Effectiveness by Combining Research Methods

In an effort to advance evidence-based practice in psychology (EBPP), the American Psychological Association Presidential Task Force on EBPP proclaimed the need for future investigations that utilize "multiple types of research evidence" to exhibit treatment efficacy and clinical utility of psychological treatments (i.e., with underrepresented groups) in which findings are generalizable (APA, 2006, p. 274). The Task Force recognized that there is much to be gained by obtaining evidence of treatment effects from various research perspectives (APA, 2006). The problem with using only one type of research method to understand a study's findings is that each individual method is uniquely suited to address certain kinds of questions (Greenberg &

Neuman, 1996). The hope is that if researchers use multiple types of methods there will be (a) a more comprehensive and reliable portrayal of treatment results, (b) balance between internal and external validity, and (c) enhanced communication between scientists and practitioners about treatment efficacy and its clinical utility (APA, 2006).

Researchers have proposed that the overreliance on large group studies becomes especially problematic when results do not have a clear connection about how findings apply to individual patients (Goldfried & Wolfe, 1998). For instance, if we observe the effects of treatment on a group of people as an aggregated whole (nomothetic knowledge), our vantage broadens and blurs the details and context of each individual (idiographic knowledge), leaving a depiction of an on-the-average treatment effect experienced by a mythical "average patient". Such results are misleading because a group of truly homogeneous patients are all but a myth (Kiesler, 1966). Even Randomized Controlled Trials (RCTs), considered by some to be the gold standard methodology for evaluating psychotherapy treatment efficacy (Chambless & Hollon, 1998), are limited in their ability to generalize results for clinical utility (Westen, Novotny, & Thompson-Brenner, 2004). Other concerns about the overreliance on group comparison research have also been recognized for decades in the psychotherapy literature (e.g., Goldfried & Wolfe, 1998; Silverman, 1996). For instance, the ethics of control groups, practical issues in collecting large numbers of participants homogeneous for a targeted disorder or problem, statistical averaging of group results, and intersubject variability are all well-known limitations of this approach (Barlow, Nock, & Hersen, 2009).

However, in a recent article Barlow and Nock (2009) ask psychotherapy researchers if "Rather than simply critiquing nomothetic methodologies, can we enrich these methodologies with a complementary focus on the individual?" (p. 20). They advocate for the use of single-case experimental designs and state that:

> The flexibility and efficiency of these designs make them ideally suited for use by psychological scientists, clinicians, and students alike, given that they require relatively little time and few resources and subjects and yet they can provide strong evidence of causal relations between variables. The time now seems right to put more emphasis on idiographic strategies that can be integrated in a healthy way into existing nomothetic research approaches in both clinical and basic science settings. (p. 20)

Dattilio, Edwards, and Fishman (2010) also suggested that multiple types of research evidence can be used to evaluate the effectiveness of psychotherapy treatments so as to lead to "a more differentiated explanation of findings and extrapolation of their implications for application to and dissemination in practice" (p. 431). The authors proposed mandatory reporting of a small set of representative systematic case studies in the publication of RCTs. They posited that this reporting strategy would promote a balance of external validity and facilitate evaluation of treatment process and contextual factors that affect its delivery with specific individuals. This type of study design, in which integration of idiographic methods – like single-case research (SCR) designs – occurs alongside the more common group (nomothetic) approaches, appears to be a much needed and viable paradigm shift for the reporting of psychotherapy results.

Edwards, Dattilio, and Bromley (2004) have outlined a general model for how to combine group and single-case research designs into a mixed-methods approach. Their model seems reminiscent of earlier research studies by Strupp (e.g., 1980a, 1980b, 1980c, 1980d), in which he systematically selected cases from the Vanderbilt I and Vanderbilt II group design research projects to be analyzed in depth by single-case qualitative methods. However, although Strupp analyzed the cases thoroughly, the methods were solely qualitative. The model proposed by Edwards et al. (2004) has not yet been applied to single-case experimental designs as recommended by Barlow and Nock (2009).

The benefits obtained by combining quantitative SCR designs with group comparison designs are abundant. SCR designs utilize each individual patient as their own control, potentially allowing for causal inferences about treatment efficacy (APA, 2006; Kratochwill et al., 2010). Additionally, the frequent, systematic administration of outcome assessments is an advantage of SCR designs that results in time-series data which may best represent dynamic changes in individual functioning that serve to highlight patterns in psychotherapy that may otherwise go unnoticed (Borckardt et al., 2008). Thus, SCR studies are valuable because they can increase the internal validity of results while also promoting a greater understanding of individual patient change processes as they naturally occur throughout the course of psychotherapy (Borckardt et al., 2008; Kratochwill et al., 2010). Furthermore, in addition to visual displays and qualitative reports of patient progress traditionally presented in SCR, recent advancements in statistical methods allow for quantifiable measurement of score

improvement (effect sizes) at the individual level (Parker & Brossart, 2003; Parker & Hagan-Burke, 2007; Parker, Vanesst, & Davis, 2011).

The appeal of integrating this idiographic design is evident for practitioners, as both visual and statistical representations of treatment effects can be used to judge the effectiveness of an intervention for each individual entering treatment within their particular context. Together with more commonly used group approaches, greater acuity and breadth of a study's results may emerge, as idiographic methods explain the particulars of individual phenomena and nomothetic knowledge finds "generalities that are common to a class of particulars and deriving theories or laws to account for these generalities" (Robinson, 2011, p. 32). The proposed mixed-methods design (Barlow & Nock, 2009; Dattilio et al., 2010) can readily be conducted and is therefore likely to be an efficient way to provide mutual communication of psychotherapy outcome results to both scientists and practitioners simultaneously. As a more comprehensive and multidimensional evidence base is produced, communication between scientists and practitioners can improve EBPP in accordance with "…APA's century-long tradition of attention to the integration of science and practice" (APA, 2006, p. 273).

Problem Statement

Rural populations are geographically isolated from urban areas where mental health professionals tend to be situated (Gamm et al., 2010; Stamm et al., 2003). Unfortunately, rural residents are in great need of mental health services they cannot readily access (Gamm et al., 2010; Rost et al., 2002; Rost et al., 1998; Wagenfeld et al., 1994). With low socioeconomic status and limited means for transportation to obtain

mental health services creating barriers to treatment for rural residents, serious mental health issues occur at rates that are often higher than for urban populations (Gamm et al., 2010; Stamm et al., 2003). Luckily, mental health professionals providing behavioral telehealth services via TVCP have begun to address mental health disparities faced by rural populations (Backhaus et al., 2012). However, more research into the effectiveness of TVCP as a modality for providing mental health services to rural populations is still lacking (Backhaus et al., 2012; Richardson et al., 2009). Furthermore, the research evaluating TVCP provided to rural populations in a primary care setting is scant.

The large majority of TVCP research conducted to date has been group research (Backhaus et al., 2012). The advantage of group research is that large numbers of participants allow for strong statistical power to find a statistically significant average effect for a treatment's effectiveness (Cohen, 1988), and if strong designs like RCTs are used, they can establish causal relations as well (Chambless & Hollon, 1998). However, the results of group research are presented in the context of on-average treatment effectiveness for an average patient that is typically not representative of those presenting for care in everyday clinical practice (Barlow et al., 2009). This disconnect between group research findings and their practical application to a particular patient has historically been a point of distaste and debate for clinicians and academics that has contributed to the scientist-practitioner divide in the field of psychology (Barlow et al., 2009; Goldfried & Wolfe, 1998).

SCR has been recognized as an important methodological tool that can be used to promulgate the effectiveness of psychotherapy (APA, 2006; Chambless et al., 1998;

Chambless & Hollon, 1998), and as a design that is useful for studying technology-based behavioral health interventions (Dallery, Cassidy, & Raiff, 2013). This design has a strong appeal to clinicians and is methodologically sound (Barlow et al., 2009). Although a few case studies have documented evidence for the effectiveness of TVCP with rural and geographically isolated patients (Cowain, 2001; Manchanda & McLaren, 1998; Nelson & Bui, 2010), only one study utilized a single-case research design (Himle et al., 2006). This study provided graphical displays of results typical of SCR, but it was limited by the absence of quantitative analyses that could describe the magnitude and statistical significance of the treatment effect (Himle et al., 2006). More SCR evaluating the effectiveness of TVCP with rural patients is needed.

As American psychotherapy research evolved away from the case study design and its focus on the individual, traditional group comparison research approaches have since dominated the psychotherapy literature (Barlow et al., 2009). The result has been a rift in the field between scientists and clinicians with the scientist-practitioner divide ever-growing in the struggle for a research design that could be useful to those in applied settings and based on the individual patient (Barlow et al., 2009). As Carl Jung (1957) so aptly stated,

> And if the psychologist happens to be a doctor who wants not only to classify his patient scientifically but also to understand him as a human being, he is threatened with a conflict of duties between the two diametrically opposed and mutually exclusive attitudes of knowledge on the one hand and understanding on the other. This conflict cannot be solved by an either/or but only by a kind of

two-way thinking: doing one thing while not losing sight of the other. (p. 7)

The implication is that there are clear disadvantages to both group and single-case research designs that could be remediated by combining them into a mixed-methods design (Barlow & Nock, 2009; Dattilio et al., 2010). Using a mixed-methods approach to evaluate psychotherapy outcomes may provide a more balanced view of change that can promote internal and external validity while serving to promote enhanced communication between scientists and practitioners. However, this type of mixed-methods model, which combines quantitative results of single-case and group comparison designs, has not been applied to psychotherapy treatment studies to date.

Purpose of the Study

This study presents findings from an on-going assessment of psychotherapeutic outcomes of a rural sample receiving TVCP through a primary care setting in the Brazos Valley of Texas. A mixed-methods (single-case and within-subjects group) design was used to present a snapshot of psychotherapy progress on a variety of psychological and physical health indicators in a study that lasted one year and three months. The purpose of this study is twofold: (1) to assess the effectiveness of psychotherapy on a variety of psychological and physical health indicators when delivered via videoconferencing to a rural sample obtaining mental health treatment through a primary care clinic in Central Texas; and (2) to illustrate the benefits of using single case visual and quantitative analyses in combination with group analyses as a multidimensional and integrative evaluation of psychotherapeutic outcomes that appeals to researchers and clinicians alike.

Research Questions

To achieve the aforementioned objectives, this study evaluated the effectiveness of TVCP with the following research questions using a mixed-methods design:

(1) Will TVCP produce statistically significant group improvements in mean depression symptoms, global psychological distress, risk level, symptom distress, emotional well-being, functioning, perceived mental health quality, and perceived physical health quality as measured by the PHQ-9, CORE-B scales (Global Distress, Risk, Symptoms, Well-Being, and Functioning), and the SF-12 (Mental and Physical Health Component Summary scores)? What percentage of patients will make reliable and/or clinically significant change on the PHQ-9 and CORE-B scales?

(2) What are the typical characteristics (i.e., demographics, diagnostic profile, and symptom severity) and treatment response of the average patient as described by group results on the PHQ-9, CORE-B, and SF-12?

(3) What are the unique characteristics (i.e., demographics, diagnostic profile, and symptom severity) and treatment responses of four patients as described by single-case quantitative results on the PHQ-9, CORE-B, and SF-12?

(4) How are the characteristics and treatment response of the average patient similar and/or different than those of the individual patients analyzed by single-case quantitative methods?

(5) What clinical insights emerge by comparing and contrasting the characteristics and treatment response of four individual patients with each other and to the average patient with a mixed-methods design?

CHAPTER II

LITERATURE REVIEW

Rural Populations

Defining "Rural"

To many people, the word "rural" is suggestive of country scenes with agricultural landscapes separating small towns of friendly people with their own cultural milieu and slow pace of life from the hustle-and-bustle pace found in large, densely populated cities. Such a colloquial definition is likely appropriate in many ways, and yet the complexity of rural populations and the multifaceted concept of "rurality" poses a problem for social science researchers and policy makers who understand that the various idiosyncrasies of rural areas in the U.S. create obstacles to arriving at universal agreement for the definition of rural (Hart, Larson, & Lishner, 2005). There are immensely large variations in the demography, economics, culture, and environmental make-up of rural areas across the country (Hart et al., 2005). For example, the proximity of urban centers from rural areas in this country is anywhere from just a few miles to hundreds of miles (Hart et al., 2005). Even the range of population size of rural towns can vary significantly, from towns with only a handful of residents to larger cities with tens of thousands (Hart et al., 2005). Some rural communities are involved in farming, whereas the large majority of rural communities are not (Hart et al., 2005). In essence, the definition of rurality may be appropriate for one purpose and inadequate or inappropriate for another (Hart et al., 2005).

Interestingly, the federal government has various definitions for rural, but does not define rural explicitly (HRSA, 2013). In the first definition, the federal government uses two major classifications of urban in the U.S. Census Bureau to describe what is not rural (HRSA, 2013). According to the U.S. Census Bureau, urban is defined as (a) urbanized areas of 50,000 people or more and (b) urban clusters of at least 2,500 and less than 50,000 people (HRSA, 2013). "Rural encompasses all the population, housing, and territory not included within an urban area" (HRSA, 2013). According to this definition, the 2010 Census indicated that 19.3% of the population and 95% of the land area was classified as rural (HRSA, 2013).

The second definition of rural by the White House's Office of Management and Budget (OMB) designates counties as Metropolitan (population of 50,000 or more), Micropolitan (population of 10,000-50,000), or Neither (HRSA, 2013). All counties not meeting the Metropolitan areas are classified as rural (HRSA, 2013). In other words, "Micropolitan counties are considered non-Metropolitan or rural along with all counties that are not classified as either Metro or Micro" (HRSA, 2013). Using this definition, 17% of the population and 74% of land area was classified as non-metro or rural (HRSA, 2013). However, a third definition derived from the previous also exists. The Office of Rural Health Policy (ORHP) uses its own classification by including all non-metro counties as rural (HRSA, 2013). The ORHP "accepts all non-metro counties as rural and uses an additional method of determining rurality called the Rural-Urban Commuting Area codes" (HRSA, 2013). With its definition, the OHRP found that 20% of the population and 91% of the land area in the U.S. is rural (HRSA, 2013).

This study utilized the OMB and ORHP definitions because both are dichotomous (metro vs. non-metro) classifications of rural which represent Leon County, Texas as non-metro. Leon County is a seven-city county spread across 1072 square miles (Wendel, Brossart, Elliott, McCord, & Diaz, 2011). The largest city in the county is composed of approximately 1800 residents, and the overall county population is approximately 16, 344 (Wendel et al., 2011). Thus, according to the OMB and ORHP definitions, Leon County is considered a Micropolitan/rural area.

Context and Barriers to Rural Mental Health

Vast stretches of rural land areas in the U.S. are often coupled with the high poverty rates for rural residents which create obstacles of accessibility to mental health treatment centers (Smalley et al., 2010). High unemployment and poverty rates are common in rural populations (Smalley et al., 2010). In 2009, the unemployment rate for rural populations was 9.6% and the poverty rate (in 2008) was 13%, similar to that of urban populations (Kusmin & Hertz, 2010). In 2013, the unemployment rate dropped to 7.8% for rural populations (7.5% for urban populations), but the poverty rate had increased to 17.7% (14.5% for urban populations) by 2012 (Kusmin, 2013). These financial obstacles make transportation to areas where the majority of mental health professionals are located extremely challenging (Smalley et al., 2010). The distances required for rural residents to obtain mental health treatment are typically very large, and studies have found these distances to be associated with less outpatient visits and an increased chance of hospitalization (Fortney, Booth, Blow, Bunn, & Cook, 1995; Fortney, Owen, & Clothier, 1999).

In addition to vast stretches of rural land areas and financial obstacles limiting accessibility to mental health care treatment, rural residents face other barriers related to availability and acceptability of such care. One major obstacle facing rural residents is the devastating shortage of mental health professionals in rural areas (Smalley et al., 2010). The availability of mental health providers for rural residents is significantly less than for urban residents, often leading to greater reliance on primary care physicians for mental health issues (Lambert & Agger, 1995). Another major obstacle which limits acceptability of care for rural residents seeking needed mental health treatment is the cultural stigma and fear of decreased anonymity of this group (Helbok, 2003; HRSA, 2005; Letvak, 2002; Smalley et al., 2010). One study by Hoyt, Conger, Valde, and Weihs (1997) found that cultural stigma is a major concern for rural residents which is inversely related to population size, such that smaller populations tend to hold more stigmatized beliefs about mental health.

Barriers such as these may help to explain why rural residents do not utilize treatments at the same rates as urban residents. In fact, a study by Wang et al. (2005) found that rural residents with mental health disorders are significantly less likely to utilize services when compared to urban residents. Rueter, Holm, Burzette, Kim, and Conger (2007) found similar resistance to treatment with their rural sample, of which only 27% of those diagnosed with a mental health disorder received treatment. The impact of help-seeking behavior for rural residents is important because it can contribute to mental health disparities. Together, the problems with accessibility, availability, and

acceptability of mental health treatment contribute to a state of poor mental health for many rural residents (Smalley et al., 2010; Wagenfeld et al., 1994).

Rural Mental Health Issues

Prevalence rates and comorbidity of psychiatric disorders for rural residents are generally very high. For instance, Rueter et al. (2007) examined a sample of 536 young adults (ages 19-23) living in the Midwestern U.S. and found that more than 60% of this sample met criteria for a psychiatric disorder at some point during their lifetime, and 24% had experienced a disorder within the year preceding their evaluation. Comorbidity of lifetime disorders for this rural young adult sample was approximately 30%, with a significant amount of the sample diagnosed with alcohol use disorders. These statistics are alarming because studies have shown that rural residents with serious mental illness have worse symptom outcomes than urban residents, especially with co-occurring substance abuse (Fischer, Owen, & Cuffel, 1996).

As another example of problematic rural mental health issues, Brossart et al. (2013) assessed 2,207 rural residents in the Brazos valley of Texas for various cultural barriers that are known to contribute to higher prevalence of depression in this population. The authors found that, in 2010, 10.9% of the rural sample was experiencing depression symptoms that merited follow-up by a mental health professional. In addition, this study found that depression rates were significantly higher for rural women than their urban counterparts. Also, the researchers found an inverse relationship (r_s = -.22, p < .01) between income level and depression scores in this sample. High depression rates and an association of depression with low socioeconomic status pose serious

mental health concerns in rural communities with limitations to treatment. The results of Brossart et al. provided evidence of a need for mental health services for this population and the potential for mental health disparities to result in their absence.

This is especially alarming if one considers that the potential lethality of untreated depression is often suicide. Disparities in suicide rates also exist when comparing rural residents to urban individuals. Research indicates that suicide rates among rural males are higher than males surveyed in urban settings across all four regions of the nation (Eberhardt et al., 2001). Furthermore, suicide attempts are also significantly higher for depressed rural adults than adults in urban areas (Rost et al., 1998).

Such disconcerting findings are reflective of the national mental health problems facing rural communities across the country. Despite the obstacles to mental health treatment, surveys of rural residents suggest that accessibility to mental health professionals is a high ranking concern for this cultural group (Smalley et al., 2010). If treatment is available and offered through low-cost settings, it is possible that mental health treatment could make a difference in remedying serious health disparities. In this vein, behavioral telehealth services have begun to meet the need for treatment in rural communities.

History of Behavioral Telehealth in Rural Mental Health

As early as 1959, the University of Nebraska School of Medicine experimented with telehealth to provide psychiatric health services between the Nebraska Psychiatric Institute and Norfolk State Hospital (Benschoter, 1967; Preston, 1993). In the mid-

1960s, telehealth was used to provide mental health consultative services for patients (Benschoter, Witson, & Ingham, 1965). Video technology eventually reached a number of rural areas in the mid-1970s, but due to the costs of these systems and limited provider acceptance, most of these projects went out of business (Bashur, 1997; Dwyer, 1973).

During the early 1990s, as technology costs decreased and the public interest in the Internet grew in popularity, the federal government began developing plans for institutional telecommunication systems (Nickelson, 1998). With the success of telehealth systems by military and correctional facilities in the late 1980s (Brecht, 1997; Edwards, 1997) and the recognition by policy makers that such an infrastructure could reduce health care costs, rural advocates began to take note (Nickelson, 1998). Successful rural advocates impacted many national health reform proposals (e.g., H.R. 5300, 1994; S. 2375, 1994). In fact, in 1987 the Office of Rural Health Policy and the National Rural Health Advisory Committee were formed within the Health Resources and Services Administration (HRSA) to provide mental health services where they were sparse (DeLeon, Wakefield, Schultz, Williams, & VandenBos, 1989). Over time, such federal agencies have supported many initiatives to facilitate the funding of behavioral telehealth for rural residents (Nickelson, 1998). As behavioral telehealth has expanded to rural areas, psychologists have become integrated into this movement to fill the need for such service (Nickelson, 1998).

Telehealth Videoconferencing Psychotherapy (TVCP)

Telehealth videoconferencing psychotherapy (TVCP) is one promising modality for service delivery of mental health treatment of geographically and economically isolated populations with limited access to mental health professionals (Backhaus et al., 2012). A systematic review of the TVCP literature (821 articles sorted for inclusion) conducted by Backhaus et al. (2012) found that TVCP is a feasible and productive alternative to traditional face-to-face psychotherapy which also has a high satisfaction of utility by both providers and patients. In addition, analyses of 42 empirical articles included in this study suggested that TVCP is equally effective as traditional face-to-face psychotherapy.

In this review, Backhaus et al. found that about half of all empirical studies used Cognitive-Behavioral Therapy (CBT), but about a quarter did not describe the intervention or described it as eclectic. Whether with CBT or treatment-as-usual, the bulk of these controlled and uncontrolled studies showed that TVCP helps different patient populations that present with a variety of disorders. To gain a better understanding of the ways in which TVCP has been effective for treating the mental health needs of rural populations, it is important to review relevant outcome research.

TVCP Outcome Research with Rural Samples

The following is a review of seven articles that addressed TVCP effectiveness research conducted with rural samples. Four group studies (Bouchard et al., 2000; Bouchard et al., 2004; Germain, Marchand, Bouchard, Drouin, & Guay, 2009; Tuerk, Yoder, Ruggiero, Gros, & Acierno, 2010) used a structured CBT protocol for a specific

disorder (e.g., Prolonged Exposure therapy for PTSD). Two group studies (McCord et al., 2011; Shepherd et al., 2006) adapted treatment to meet the needs of individual patients. One single-case research study (Himle et al., 2006) of TVCP for non-rural patients was also reviewed because it met the basic standards to qualify as a single-case research design. Although a couple of case studies of rural patients have documented successful treatment using TVCP (Cowain, 2001; Nelson & Bui, 2010), they were not reviewed because they lack mechanisms for establishing internal validity and quantitative treatment response data that can be assessed for statistical, practical, and clinical significance.

A pilot study by Bouchard et al. (2000) examined the effectiveness of a 12-session weekly protocol of CBT for Panic Disorder and Agoraphobia delivered via TVCP by three trained psychotherapists of varying experience levels (Ph.D., M.A., B.S.) to a small group of eight participants residing in rural Canada. Five men and three women ranging in age from 23 to 63 years were treated for panic disorder and agoraphobia by therapists via TVCP at a remote mental health clinic in Maniwaki – a small rural town approximately 85 miles from Ottawa. Ethnicity and education demographic data was not reported by the authors. Although the main diagnosis for all patients was panic disorder, every patient was also diagnosed with a co-occurring disorder. Several patients taking selective serotonin reuptake inhibitors (SSRIs) for more than six months were included in the study because they agreed not to make changes to their prescription during treatment.

In this pilot study, Bouchard et al. (2000) evaluated number of panic attacks, panic apprehension, global assessment of the severity of panic disorder with agoraphobia, self-efficacy to control panic attacks, trait anxiety, and global functioning via self-report measures and patient diaries. Non-parametric analyses of pre- and post-treatment measurements were conducted as a result of the small sample size. Results indicated that the group of eight patients made statistically significant reductions (all Wilcoxon's Z significant at $p < .05$) in number of panic attacks ($Z = -2.02$), panic apprehension scores ($Z = -2.36$), trait anxiety ($Z = -2.19$), global severity of panic and agoraphobia scores ($Z = -2.2$) along with improvements in self-efficacy to control panic ($Z = -2.52$) and global functioning ($Z = -2.17$). The authors concluded that despite the small sample size, TVCP was effective for delivering a panic disorder treatment protocol to help patients in rural areas obtain access to mental health care.

Bouchard et al. (2004) conducted a follow-up study of Bouchard et al. (2000) to see if results would replicate with a larger sample. The effectiveness of a 12-session weekly protocol of CBT for Panic Disorder and Agoraphobia was delivered by four trained clinicians (Ph.D. B.S., M.A., M.A.) to 21 participants (71% female, 29% male). Eleven patients resided in rural Canada and received treatment at a remote site in Maniwaki via TVCP while 10 others obtained face-to-face psychotherapy in the city of Ottawa. The 11 patients who received treatment at the remote site had a mean of 10.6 years of education, while the urban residents had a mean education of 14.6 years. Ethnicity of the participants was not reported by the authors. Sixty-six percent of this total sample was diagnosed with a co-occurring disorder.

Bouchard et al. (2004) assessed many of the same variables from the previous study (Bouchard et al., 2000) via self-report and patient diaries. Results of the Repeated-Measures Analysis of Variance (RMANOVA) of pre-and post-treatment measurements for the group of 21 participants showed no significant differences between face-to-face and TVCP conditions. Results confirmed findings from the pilot study of Bouchard et al. (2000) with large statistically significant improvements (all η^2 significant at $p < .01$) in panic attack frequency ($\eta^2 = .44$), panic apprehension ($\eta^2 = .51$), agoraphobic cognition ($\eta^2 = .63$), self-efficacy to control a panic attack ($\eta^2 = .64$), state anxiety ($\eta^2 = .54$), trait anxiety ($\eta^2 = .60$), depression ($\eta^2 = .66$), and global functioning ($\eta^2 = .76$). The authors concluded that TVCP was an effective modality for delivering this treatment protocol for panic disorder. They suggested that TVCP can improve access to mental health services for patients limited by their disorder (i.e., agoraphobia) as well as for rural patients with limited accessibility to providers because of financial or geographical barriers.

In addition to panic disorder treatment for rural patients, TVCP has also been used for the delivery of CBT for rural patients with Posttraumatic Stress Disorder (PTSD; Germain et al., 2009). Germain et al. (2009) examined the effectiveness of a CBT protocol administered to 48 participants (60% female, 40% male) suffering from PTSD, of which 62.5% had a co-occurring disorder. Treatment was delivered weekly in 60-minute sessions for a range of 16 to 25 weeks by psychologists with several years of experience. The participants ranged in age between 18 and 65 years. Approximately 40% of this sample had a college degree and 60% had a high school degree. Ethnicity of patients was not reported by the authors. Eleven rural residents were treated via TVCP

in a remote clinic 200 kilometers from Montreal, Canada, five urban residents were treated via TVCP in Montreal, and the other 32 participants obtained face-to-face psychotherapy.

All participants in this study were assessed for symptom changes during a waiting period as well as at pre- and post-treatment periods. Several self-report measures were used, including the Modified PTSD Symptom Scale (MPSS), the Beck Depression Inventory (BDI), Beck Anxiety Inventory (BAI), and the Assessment of Current Functioning (ACF) scale (Germain et al., 2009). RMANOVA indicated statistically significant improvements (all η^2 significant at $p < .001$) for the group on the MPSS ($\eta^2 = 0.64$), BDI ($\eta^2 = 0.35$), BAI ($\eta^2 = 0.37$), and ACF ($\eta^2 = 0.31$) after treatment, without any statistically significant differences between the face-to-face and TVCP groups. The authors also calculated clinically significant change using the normative cut-off method prescribed by Jacobson and Truax (1991), but evaluated participants on all measures combined using their own cut-off scores. According to their method, 50% of participants in the TVCP condition and 62.5% of participants in the face-to-face psychotherapy condition made clinically significant change after treatment. The authors concluded that TVCP is an effective modality for administering a treatment protocol for PTSD, which can increase the availability of psychological services for populations that are remotely located and lack access to mental health care.

The use of TVCP by Veterans Affairs medical centers across the country has been growing over the last decade to address a range of mental health disorders troubling Veterans (Cully, Jameson, Phillips, Kunik, & Fortney, 2010). The treatment of PTSD for

combat Veterans is one important example of how TVCP can help reduce mental health disparities to rural communities (Tuerk et al., 2010). For example, one study of 47 Veterans (94% male, 6% female) diagnosed with combat-related PTSD evaluated the effectiveness of weekly, 90-minute sessions of manualized Prolonged Exposure (PE) therapy (Tuerk et al., 2010). Treatment was delivered by two experienced psychologists to a sample of 12 rural patients via TVCP and 35 patients face-to-face for eight to fifteen sessions. The 12 rural Veterans who received treatment via TVCP did so at a Veteran's Affairs community-based outpatient clinic. Sixty-four percent of Veterans in the total sample were Caucasian and 34% were African-American. The mean age for the group was 39 years.

Veterans were assessed for symptom severity at pre- and post-treatment using the PTSD Check List (PCL) and Beck Depression Inventory – II (BDI-II). Paired-samples *t* test results indicated that the 12 Veterans treated with TVCP made large statistically significant reductions on the PCL ($d = 2.9$, $p < .001$) and the BDI-II ($d = 2.3$, $p < .001$). Veterans treated in the face-to-face psychotherapy condition made larger gains ($d = 4.2$) on the PCL than those with TVCP, but differences in effect sizes between the groups were not statistically analyzed for significance due to sample size limitations. Veterans treated in both conditions made clinically significant improvement in PTSD and depression symptoms according to the PCL and BDI-II, but the exact percentage was not calculated using methods like those prescribed by Jacobson and Truax (1991). The authors concluded that PE was effective and could be safely administered via TVCP to rural Veterans to treat PTSD. They suggested that these results are important because

many rural Veterans are limited in accessibility to health care for financial reasons and/or may avoid in-person treatment because they are uncomfortable around large crowds like those typically found in hospital waiting rooms.

In the aforementioned studies, treatment delivered via TVCP to rural patients exhibited promising results for this modality. TVCP outcome studies reviewed to this point (Bouchard et al., 2000; Bouchard et al., 2004; Germain et al., 2009; Tuerk et al., 2010) were composed of experimental group designs which evaluated the effectiveness of CBT protocol treatments for specific disorders (i.e., panic disorder, PTSD), some in comparison to face-to-face psychotherapy. Findings from these types of studies are important because these empirically supported treatments (ESTs) are best practices (APA, 2006) that require evaluation of whether these treatments can be effective if administered via TVCP. However, the exclusive use of ESTs targeted for specific disorders without consideration of the individual client's cultural context and needs has been a major point of controversy in the field of psychotherapy research (Goldfried & Wolfe, 1998).

Administering standardized techniques is not always recommended for providers in rural communities because urban-centered standards of treatment often do not directly apply to rural communities (Wagenfeld & Buffum, 1983). This is important to recognize because acceptability of treatment can easily lead to drop out (Bernal, Jimenez-Chafey, & Rodriguez, 2009). Consideration of cultural context is an important ethical standard of the APA, and understanding alternative treatment sources for minority cultures is one important goal of competent treatment to rural communities (APA, 1993, 1995).

Group research examining the effectiveness of treatment-as-usual delivered via TVCP to rural patients is also valuable because rural patients often present with unique needs (HRSA, 2005) not always amenable to protocol-driven ESTs. As one example, Shepherd et al. (2006) evaluated the utility of TVCP for 25 patients living in rural Australia who were experiencing psychological distress as a result of cancer. Patients were treated by a psychologist with various short-term CBT interventions (e.g., problem-solving, activity scheduling, and controlled breathing) via TVCP for a variety of psychological issues (i.e., anxiety, depression) for between one and six weekly or bi-weekly sessions. Patients participated in TVCP sessions from remote sites outside the Sydney metropolitan area. Patients were 64% female and 36% male, with a range in age of 28 to 70 years. Forty-eight percent of patients had an education level at or below 10^{th} grade. Ethnicity of the participants was not reported by the authors. Eighty-four percent of patients were receiving cancer treatment and two patients were also prescribed antidepressant medication.

Patients were assessed at pre- and post-treatment periods as well as after a one-month follow-up using the Hospital Anxiety and Depression Scale (HADS) and the Functional Assessment of Cancer Therapy – General (FACT-G). RMANOVA indicated that TVCP was effective for this sample of rural cancer patients, reducing anxiety (η^2 = .33, p = .01) and depression (η^2 = .08, p = .38) as well as the overall symptom score on the HADS (η^2 = .24, p = .04). In addition, emotional well-being, as measured by the FACT-G, was significantly improved for the group after treatment (η^2 = .40, p =.003). The authors concluded that psychological support for rural patients in Australia is

important and that TVCP is an effective, acceptable, and pragmatic option for delivering mental health services to these geographically isolated populations in need of support. The authors also recognized that despite not providing the "adequate dose" of CBT delivered in standard practice, the brief adapted interventions tailored to each patient's needs produced successful outcomes (Shepherd et al., 2006, p. 459).

In another important example of TVCP delivered to rural patients with unique needs, McCord et al. (2011) examined the effectiveness of treatment-as-usual provided by doctoral psychology students to a rural sample of 68 patients through a primary care clinic in rural Texas. The rural primary care clinic is situated approximately 70 miles from the city of Bryan. It was established by a community partnership to address mental health disparities for rural residents. Doctoral students in the Counseling Psychology program of Texas A&M University (TAMU) provided TVCP from the training clinic in Bryan to patients in this rural primary care clinic. The patient sample was composed of approximately 71% females and 29% males ranging in age from 9 to 73 years (M = 40.50, SD = 14.10). With regards to ethnicity, the sample was approximately 80% Caucasian, 7% each of African-American and Hispanic individuals, and 4.4% identified as biracial. Patients presented with a variety of psychological disorders and issues, ranging from bereavement, relational and adjustment problems to more serious mental health disorders like depression, anxiety, and PTSD.

Patients in this study were provided treatment-as-usual psychotherapy according to their needs. Treatment included CBT, cognitive processing therapy, person-centered, and humanistic interventions. All patients were measured at pre-treatment and after

every four weeks for depression symptoms and perceived mental health quality using the PHQ-9 and SF-12 instruments, respectively. Paired-samples *t* tests revealed that patients made statistically significant decreases in depressive symptoms ($M_{diff.}$ = 5.88, *SD* = 7.16, $p < .05$) along with an increase in perceived quality of mental health ($M_{diff.}$ =11.39, *SD* = 7.94, $p < .05$) after four sessions. The authors concluded that the community partnerships created the capacity for sustainable mental health services in which TVCP proved effective as a modality for delivering psychological support to a rural area otherwise bereft of much needed services. This study, like Shepherd et al. (2006) showed that delivering TVCP to rural patients with unique needs can be effective even when treatment is not protocol-driven.

In addition to group outcome studies of TVCP, Himle et al. (2006) conducted an experimental single-case research design to test the effectiveness of a 12-week CBT protocol treatment (Prolonged Exposure and Response Prevention) for Obsessive-Compulsive Disorder (OCD) delivered via TVCP to three female patients of ages 19, 29, and 39 years. Patients exhibited a variety of obsessions and compulsions (i.e., checking behaviors, contamination concerns, and hoarding), and all were diagnosed with major depressive disorder (MDD) in addition to OCD. Two of the three patients were taking SSRIs. Ethnicities of the three patients were not reported by the authors, nor were the qualifications of the therapists.

The authors utilized a multiple baseline design (MBD) with these three patients using a one, two, and three-week baseline period to which patients were randomly assigned (Himle et al., 2006). The researchers assessed patients' symptom improvements

weekly using the Yale-Brown Obsessive Compulsive Scale (YBOCS). They also utilized the Hamilton Depression Rating Scale (HAM-D) at pre-, post-, and follow-up treatment periods. Neither visual nor statistical analyses of phase contrasts were conducted by the researchers, but informal observation of the data suggested that all three patients improved over the 12-week course of treatment. In addition, patients' decreased YBOCS scores (Patient 1: pre = 29, post = 16; Patient 2: pre = 31, post = 14; Patient 3: pre = 30.5, post = 17) and self-reports of reductions in compulsive behaviors seemed to verify their conclusions. At follow-up, patients maintained improvements on the Y-BOCS, but only one patient maintained the reduction in depression symptoms according the HAM-D (Patient 1: pre = 5, post = 0, follow-up = 11; Patient 2: pre = 6, post = 5, no follow-up; Patient 3: pre = 20, post = 14, follow-up = 8). The authors concluded that CBT via TVCP was effective for reducing obsessive-compulsive symptoms for these three patients. They reasoned that the use of a MBD helped maintain internal validity of results, but also stated that results should be viewed with caution because of the small sample size. In actuality, the MBD design used by the authors did not meet the requirements for strong internal validity because the concurrent baseline period critical to an MBD requires a minimum of three baseline points to establish the stability of pre-intervention functioning (Kratochwill et al., 2010). The authors did not meet this basic criterion in two of the three phase contrasts; thus, no causal relation could be inferred from their results. Furthermore, the authors did not use single-case data analyses to evaluate the magnitude and statistical significance of the observed treatment responses. Overall, this study was a step-forward from the basic case-study design, but it was not

the best example of a single-case experimental design because it lacked many important features critical to establish causality.

Taken together, the TVCP outcome literature of rural populations is limited, but the extant studies support its effectiveness. Most studies used group comparison designs, but varied with regards to types of treatment interventions and psychological problems assessed. In addition, the majority of studies had small sample sizes typically ranging between eight and fifteen patients undergoing TVCP, with the largest sample being 68 patients (McCord et al., 2011). Also of note, the majority of studies did not present the ethnic composition of their samples. This greatly limits generalizability and applicability of results for clinicians desiring to know if treatments are effective for particular patients. The literature review suggests the need for further effectiveness research. In this context, it is important to review relevant methodological issues in psychotherapy outcome research and the kinds of information that common designs can contribute to the growing literature base in this field.

Methodological Issues in Psychotherapy Research

Each research design has its strengths and its limitations, but each has value in its own right (APA, 2006). This is an important point to remember because as psychology evolved and new designs emerged in psychotherapy research, various zeitgeists contributed to the acceptance and use of certain designs over others. The challenge for the field of psychotherapy research has been and continues to be a struggle to find a design which appeals simultaneously to scientists and practitioners. Below is a review of prominent methodological issues that have arisen as psychotherapy research evolved

from its origin – the case study design. Additionally, a review of recent recommendations for combining single-case and group designs is provided because this type of mixed-methods design is proposed by some researchers to be a potential remedy for bridging the scientist-practitioner divide (Barlow & Nock, 2009).

Case Studies

The case study emanated as a natural by-product of the evolution of the field of psychology and was almost entirely the sole methodology of clinical investigations through the first half of the 20th century (Barlow et al., 2009). Many early psychologists championed this observational design (Robinson, 2011). For example, the case study was famously exemplified by Breuer and Freud (1957) in their documentation of the "talking cure" and hypnosis for hysteria treatment in their patient Anna O.

Case studies focus on an individual or individual unit, such as a small group of people or a classroom of students (Nock, Michel, & Photos, 2007). This design is qualitative in nature, describing specific contexts, history, and unique aspects of the case. Hypotheses about influential factors are derived retrospectively from anecdotal evidence (Nock et al., 2007). The absence of experimental controls or systematic measurements in case study designs allow for more freedom and flexibility to be incorporated into clinical work (Nock et al., 2007). The case study is also very useful for documenting novel clinical experiences that can help derive new insights which can later be tested experimentally (APA, 2006). In addition, aggregation of systematic case studies can provide good information about clinical utility for patients with similar characteristics (APA, 2006).

Although they are clinically useful and can help create hypotheses to be tested under experimental conditions, case studies are limited in their ability to produce results with sufficient internal and external validity (Nock et al., 2007). As a result, the uncontrolled case study was strongly rejected by scientists in the 1950s (Barlow et al., 2009). The consequential rejection of these uncontrolled designs by the majority of scientists in the field led to mainstream adoption of group comparison research.

Group Research

At the beginning of the 20th century, the study and measurement of individual differences led to the proposed discovery that "nature strove to produce the average man but, due to various reasons, failed, resulting in errors or variances in traits that grouped around the average" (Barlow et al., 2009, p. 5). It was assumed that "Where nature failed…man could pick up the pieces, account for the errors, and estimate the average man through statistical techniques" (Barlow et al., 2009, p. 5). Galton, Pearson, Fisher, and others were strongly influenced by this presumption and developed many sophisticated descriptive statistics (i.e., correlation, analysis of variance, statistical inference testing) that advanced the field of measurement and psychological research (Barlow et al., 2009). As psychology evolved into a science, the lure of generalizing one's findings by using novel and powerful group-based statistics led to the popularity of methodological designs that could produce such findings (Barlow et al., 2009). Thus, as the intensive study of the individual fell out of favor, the popularity of group experimental designs rose and gained prominence within psychological science (Barlow et al., 2009).

The most common group research designs utilized in psychology are between-groups, factorial, and within-subjects group experimental designs (Heppner, Wampold, & Kivlighan, 2008). There are a variety of ways to execute these three designs, each with its own advantages and disadvantages (Heppner et al., 2008). Probably the most popular between and within-subjects groups designs are the pretest-posttest control group designs (Heppner et al., 2008).

Group research designs can obtain their scientific rigor through controlling threats to internal validity via random assignment of subjects, use of control groups, and manipulation of independent variables between or within groups (Heppner et al., 2008). Standards and procedures for evaluation and reporting exist for RCTs (Schulz, Altman, Moher, & CONSORT Group, 2010) and other group designs (Des Jarlais, Lyles, Crepaz, & TREND Group, 2004; Shernoff & Kratochwill, 2003; Kratochwill & Stoiber, 2002) to help promote scientific foundations that improve causal inferences of results. Although RCTs are considered by some to be the gold-standard methodological design, they are not always feasible for ethical, financial, or logistic reasons.

Ethical concerns are one of many limitations of RCTs cited in the psychotherapy outcome literature. Many clinicians feel that withholding treatment by forming a no-treatment control group goes against ethical standards of the profession (Barlow et al., 2009). Although this practice continues in the zeitgeist for ESTs, there are objections by many in the profession (Barlow et al., 2009).

Another limitation of group research in general results from the practical issue of recruiting large numbers of participants. Recruiting large numbers of participants poses

financial obstacles in large group studies which compensate participants for their involvement because of the costs of obtaining sufficient sample sizes to attain adequate levels of statistical power. In addition, some group studies (e.g., RCTs) will face challenges in recruitment of sufficient homogeneous participants for a targeted disorder when the study design requires exclusion of participants with comorbid disorders (Barlow et al., 2009; Westen et al., 2004).

Group comparison designs are also limited because they typically only assess each participant once prior to and after treatment. Inherent in this type of design is the lack of attention given to intersubject variability (Barlow et al., 2009). The clinical course of a specific patient during treatment (within-subject variability) is ignored, leading to a loss of change patterns found in time-series designs (Barlow et al., 2009). This lack of attention to change over time is not helpful to clinicians interested in understanding how patients progress through treatment (Barlow et al., 2009)

Probably the most significant limitation of group research involves the use of statistical averaging of group data to infer whether a treatment is effective for a particular individual. When the answer to the question "Is psychotherapy effective?" is approached from the group vantage point, important findings are obscured (Barlow et al., 2009). Jung (1957) noted that "The statistical method shows the facts in the light of the ideal average but does not give us a picture of their empirical reality. While reflecting an indisputable aspect of reality, it can falsify the actual truth in a most misleading way" (p. 6).

There are three primary reasons that statistical averaging of group results can be potentially misleading and lead to inaccurate inferences about individuals in the group. First, the mean statistic is “one number representing the entire dataset, and may not be any individual’s scored score in the data” (Thompson, 2006, p. 40). For example, if five people each score a ten on a measure and five other people each score a zero, the mean is equal to five although no one person had that score. In such an example, any inferences about treatment effectiveness for the average individual based on group (mean) data would be inaccurate because (a) the average person portrayed by group results does not exist in the group and (b) the group is composed of extreme scores with a mean located at the center.

The statistical averaging of group results is also problematic because “the mean can be highly influenced by relatively few anomalous scores in the data, even when the number of scores is quite large” (Thompson, 2006, p. 44). In other words, even if the majority of people in a group scored very similarly, a few people in a group with extreme scores could pull the group mean away from the majority of scores toward the extreme scores. In this case, the group mean would be misleading if interpreted as the performance of any individual in the group because the mean may not be close to the majority of scores in a sample with extreme outliers.

Lastly, statistical averaging of group results is problematic because “researchers tend to lump together all subjects with the same manifest problem, regardless of the etiological, mediational, contextual, and maintenance facts that underlie and act to perpetuate the maladaptive pattern" (Fishman, 1981, p. 244). The assumption in group

research is that the individuals in the group are very similar because they have the same diagnosis or that statistical analyses can account for individual differences. However, this is a problematic assumption because people are inherently unique and statistical analyses cannot control for all unique factors in personality, context, or other relevant individual characteristics. The average results from group studies only represent any individual in a group to a certain degree, and may not represent a specific individual at all. These characteristics of statistical averaging of group results create a problem for clinicians wishing to apply such findings to their individual patients.

Limitations of group research have prompted several clinical investigators (e.g., Barlow, 1980; Kazdin, 1981; Lazarus & Davison, 1971) to propose the need for development of "…an alternate research paradigm for building and testing an effective approach to psychotherapy, one that both emerges from therapist-patient interactions and individualizes the intervention for the particular case at hand" (Goldfried & Wolfe, 1996, p. 1013). Although the traditional clinical case study was not considered the best option because it was uncontrolled, researchers realized it could play an important role in isolating therapeutic mechanisms of change if experimental conditions could be added to this design (Barlow et al., 2009). In this way, the single-case experimental design was born and a return to the individual as a focus for psychotherapy research occurred.

Single-Case Experimental Studies

Single-case experimental studies (SCES) have several essential characteristics that distinguish them from other research methods and serve to enhance the scientific rigor of this approach. Much like the case study of early psychotherapy research, an

individual case or cluster of cases is the unit intervention and analysis in SCES (Kratochwill et al., 2010). However, in juxtaposition with the traditional case study, SCES attempt to utilize various control mechanisms in an effort to establish a causal relation between an intervention (independent variable) and changes observed in the outcome of interest (dependent variable; Kratochwill et al., 2010).

Perhaps the most fundamental property of SCES which facilitates its experimental framework is the *repeated measurement* of a dependent variable(s) over time across all phases of the study (Nock et al., 2007). Repeated measurement of the dependent variable(s) should occur early, frequently, and consistently using reliable assessment instruments or methods (e.g., rater judgments) which can be determined as statistically reliable (Nock et al., 2007). If conducted in this way, the early introduction and repeated measurement of the dependent variable(s) creates a *baseline phase*, or pre-intervention level of functioning (labeled Phase A), which is then compared to changes that occur during the intervention phase (labeled Phase B; Nock et al., 2007). This AB contrast is the most basic element of all single-case designs (Kratochwill et al., 2010).

As the participant progresses from baseline to experimental phase(s), comparisons between these phases can be made because repeated measurements create large numbers of data points in each phase that essentially serve the function of data obtained for comparison by group studies (Nock et al., 2007). However, instead of comparing differences between groups, repeated measurement allows for evaluation of *intra-subject variation* on the dependent variable(s) as an individual progresses from no-treatment (Phase A) to treatment (Phase B) across time (Hilliard, 1993). Comparison of

AB phase differences can involve *single case quantitative analysis* of an individual's treatment response because of the resulting time series data produced by repeated measurements (Hilliard, 1993).

However, with a simple AB design it is difficult to validly draw causal inferences about results produced from quantitative analyses because alternative explanations for the observed intervention effect cannot be ruled out (Kratochwill et al., 2010). Threats to internal validity include ambiguous temporal precedence (did the IV come before the DV), participant selection effects, history effects, maturation effects, statistical regression (toward the mean), attrition, and testing effects (Kratochwill et al., 2010). To minimize these threats, SCES require *intentional manipulation of the independent variable* with an expected change in the outcome variable along with important design considerations (Kratochwill et al., 2010; Nock et al., 2007). Some single-case researchers suggest internal validity is best obtained with a more complex and structured design than the simple AB contrast (Kratochwill et al., 2010). Standards have been outlined for evaluating methodological soundness and credibility of results from SCES (Kratochwill et al., 2010; Shernoff & Kratochwill, 2003).

Some researchers (Kratochwill et al., 2010; Shernoff & Kratochwill, 2003) have recommended standards and procedures for evaluating if designs meet criteria for methodological soundness and internal validity. For instance, in addition to basic SCES requirements (i.e., intentional manipulation of the IV with repeated measurement of DV), the What Works Clearinghouse (WWC) standards (Kratochwill et al., 2010) and other similar standards (Shernoff & Kratochwill, 2003) require that SCES meet a

minimum of three data points per phase and at least three demonstrations of an intervention effect across three different phase changes. Each design is judged according the following categories: (a) *Meet Evidence Standards*, (b) *Meet Evidence Standards with Reservations*, or (c) *Do Not Meet Evidence Standards* (Kratochwill et al., 2010). For instance, a multiple baseline design must have a minimum of six phases with at least five data points per phase to *Meet Standards*, or a minimum of six phases with at least three data points per phase to *Meet Standards with Reservations* (Kratochwill et al., 2010). The purpose of these standards is to provide a framework to evaluate SCES in a uniform fashion rather than in an idiosyncratic manner.

Despite stringent standards and procedures recommended for SCES, these designs are extremely flexible and adaptable to the needs of clinicians and researchers (Kratochwill et al., 2010). If an intervention is not having the desired effect on the patient, the clinician can adjust or manipulate the intervention accordingly while continuing measurement of the outcome variable (Horner et al., 2005). In many experimental designs, this type of mid-intervention change would be prohibited because the researchers would worry about the lack of controls impacting the causal relationship between the independent and dependent variable(s). However, in SCES, as long as repeated measurement of the outcome variable(s) continues, this can become a new baseline period and an adjusted intervention could be implemented after this new baseline period is deemed to be stable (Kratochwill et al., 2010).

Another important advantage of SCES is the representative nature of the data with regards to psychotherapy treatment. SCES produces time-series data which presents

a more realistic portrayal of treatment response than pre- and post-treatment measurements typically obtained in large group studies (Barlow et al., 2009). This is helpful to clinicians and scientists interested in isolating mechanisms of change and understanding factors involved in treatment response (Barlow et al., 2009). This is a strength of SCES which promotes the applicability of the results for clinicians (Barlow et al., 2009; Kratochwill et al., 2010).

Along with many advantages of SCR, there are also limitations. Probably the most obvious criticism recognized in the field of SCR is the issue of the generalizability of findings (Hilliard, 1993). It seems fairly obvious that study results obtained from an N = 1 or even an N = 4 would not allow for valid inferences about the behavior of the larger population. Similar to large group studies, the problem of generalizability in SCR is addressed with future replications (Barlow et al., 2009) and the aggregation of results through meta-analysis (Allison & Gorman, 1993). In this way, replications of SCES serve to establish the evidence-base needed to develop empirically supported treatments (APA, 2006).

Another potential limitation of SCR (and group research) is a phenomenon known as autocorrelation, or serial dependency (Busk and Marascuilo, 1988; Matyas & Greenwood, 1990, 1996). Autocorrelation is a characteristic of time-series data in which proximal data points are correlated. Autocorrelation is problematic in SCR because it may inflate effect sizes, distort inference statistics (p-values), and impair visual ratings by judges, thus leading to more Type I errors with statistical models that assume independence of data (Busk & Marascuilo, 1988; Matyas & Greenwood, 1990, 1996;

Parker & Brossart, 2003). However, the extent to which autocorrelation is a problem is determined by the amount of autocorrelation and the statistic used. Some statistical methods are more impacted by autocorrelation than others (Busk & Marascuilo, 1988; Matyas & Greenwood, 1990, 1996; Parker & Brossart, 2003). Fortunately, various recommendations and statistical procedures have been outlined in the field of SCR to address autocorrelation by providing suggestions about how to control undesirable high levels of autocorrelation (Parker, Cryer, & Byrns, 2006).

A recent resurgence in single-case research methods in the last decade has encouraged some researchers to suggest combining single-case and group methods to address the methodological concerns that arise when only one method is used. Below is a review of this movement and recommendations for a paradigm shift in reporting of psychotherapy outcome research. In addition, a brief discussion of the current study's methodological contribution to the emerging paradigm is presented to provide context about how recommendations were incorporated.

Mixed-Method: Single-Case and Group Designs

Along with Carl Jung, Gordon Allport was one the first psychologists to advocate for the idiographic (individual) approach as a way to confront the "wobbly laws of generality…with the concrete person" (Allport, 1962, p. 407). Instead of an overreliance results produced from group research studies, Allport (1962) believed that psychotherapy research should "start with individual behavior…then seek our generalization…but finally come back to the individual…for a fuller and more accurate

assessment" (p. 407). This sentiment, although not the norm in the field, was espoused by other psychotherapy researchers as well.

As group comparison research dominated the literature, there was an attempt by many in the field to return to the individual as a focus of study (Barlow et al., 2009). In a significant attempt to steer the field of psychotherapy research into the applied domain, Bergin and Strupp (1970) published an article titled "New Directions in Psychotherapy Research". This article proposed the single-case design as a "new paradigm of inquiry…which will move us forward in our understanding of the mechanisms of change" (Bergin & Strupp, 1970, p. 19).

Group Comparison Designs with Case-Based Studies

Strupp (1980a, 1980b, 1980c, 1980d) was one of the first psychologists to utilize single-case research methods to analyze psychotherapy outcomes from a larger group study. He conducted a systematic comparison of case-based studies from the Vanderbilt Psychotherapy Project, a study of 18 male college undergraduate students undergoing psychotherapy twice-weekly. Specifically, patients were divided into high success and low success groups according to various outcomes obtained with a personality inventory, patient-rated relationship scale, and a therapist-rated process scale. Based on these outcomes, the author systematically compared two sets of matched cases with similar issues and treated by the same psychotherapist, but where one patient was judged to be a success and the other deemed a therapeutic failure.

Strupp (1980a, 1980b, 1980c, 1980d) did not use an experimental design. Instead, he used a systematic case study method in which he integrated quantitative data

(patient outcome scores) as descriptive indicators of treatment response. He focused on process, the relationship, patient personality, and patient-therapist dynamics. Through this qualitative analysis, Strupp extrapolated various clinical insights about the patient-therapist relationship, patient personality characteristics, and experience level of therapists which he reasoned had impacted the outcome of psychotherapy.

More recently, Fishman (2011) reminded the field of psychotherapy of the importance of Strupp's (1980a, 1980b, 1980c, 1980d) work to obtain clinically useful research from group studies. The author proposed the use of what he termed the "Individual Case-Comparison" method as a way to identify therapeutic factors influencing treatment response across cases. Like Strupp (1980a, 1980b, 1980c, 1980d), Fishman (2011) advocated for systematically comparing successful and unsuccessful cases drawn from a successful randomized controlled trial's (RCT) treatment condition. This method is one recommendation from a more general movement intended to contextualize psychotherapy research findings for practical application that is helpful and appealing to clinicians (Dattilio et al., 2010).

Edwards et al. (2004) proposed a general method for combining group comparison analyses with case-based studies. The basis for this type of mixed-methods approach is "grounded in the data of cases, it is testable against new cases, and its generalizations are lawful relationships between operationalized phenomena that have been observed and replicated" (p. 592). The authors suggested that a researcher should first start with documentation of a process in one case. Next, the same process should be demonstrated in similar cases, and then subsequent cases should be used to find evidence

that confirms said process or is different from it in important and well-defined aspects. Using this procedure, the authors reasoned that theory begins to emerge as a structure of principles, hypotheses, and distinctions that are advancing toward a refined formulation that includes the breadth of phenomena observed in all cases. They recommend that along the way, plausible rival hypotheses should be identified and eliminated (or tested further for modification of theory) by examining the implications of these hypotheses for the case in question and other relevant cases.

The authors recognize that given the complexity of human behavior, new data will be regularly collected which may or may not confirm the original hypotheses (Edwards et al., 2004). They reason that the new data obtained in the process of theoretical formulation is beneficial because new evidence will contribute to and extend the current theory. Of course, they also noted that refinement and progression of theory will be dependent upon the quality and extensiveness of the data available at the time. The implication of using this type of theory-building methodology is that science is recognized as a complex *process* which is ever-evolving, testable, changeable, falsifiable, and open to new insight if it advances knowledge.

Group Comparison Designs with Single-Case Experimental Studies

Although several researchers have made important recommendations and described a general approach for conducting a mixed-methods design using single-case and group comparison research, the recommendations are vague and mostly limited to case-based studies. Specific methods for integrating single-case experimental designs with group experimental designs – as recommended by Barlow and Nock (2009) – have

not yet been established. This is a new paradigm shift which raises questions about ways to conduct the design and report results within this design. Best-practices may exist for evaluating designs (Kratochwill et al., 2010; Shernoff & Kratochwill, 2003) and reporting results for individual research methods, but the best way to present the combined results of single-case and group analyses so as to appeal to scientists and practitioners is currently up for debate. To date, this movement is in its infancy, so methods for reporting results must be considered an exercise under examination by the field of scientist-practitioners. The current study's method should be judged in this context.

The present study is one example of combining a group comparison design with a single-case experimental study. Specifically, this study used traditional group and single-case quantitative methods to report results along with a variation of the case-comparison method recommended by some researchers promoting a paradigm shift for reporting psychotherapy treatment results (Dattilio et al., 2010; Edwards et al., 2004; Fishman, 2011). The following general steps were taken to conduct this mixed-methods study.

First, quantitative group analyses were conducted to understand the practical, clinical, and statistical significance of the sample's treatment response on various outcome measures. Next, a summary of the typical characteristics (i.e., demographics and diagnosis) and treatment response of the sample was presented in the context of the average patient described by the group results. Four patients who met the What Works Clearinghouse standards for single-case experimental studies (Kratochwill et al., 2010)

were then analyzed for practical, clinical, and statistical significance on the same measures as the group. The results of the single-case analyses of each patient were presented in a narrative format which described their unique characteristics and treatment responses. Lastly, the characteristics and treatment responses of the four individual patients were compared to those of (a) the average patient and (b) to the other individual patients in the single-case design in order to facilitate clinical insights from the quantitative data.

CHAPTER III

METHOD

Study Methods

Context

Texas has one of the largest rural populations in the country, but it is the state with the greatest proportion of counties designated as MHPSAs (HRSA, 2008; Trust for America's Health, 2008) and it ranks 48th with regards to state funding allocated for mental health services (Texas Health Institute, 2008). One quintessential example of a MHPSA in rural Texas is the Brazos Valley – a seven county region spread out across Central Texas. According to the 2006 Brazos Valley health survey, 62% of adult residents reported that they needed mental health services but were unable to obtain them; more than 50% of the respondents who reported a need for alcohol abuse treatment could not obtain services; and 50% of respondents reported experiencing at least one day of "poor mental health" in the last 30 days (Center for Community Health Development [CCHD], 2006). Similarly, the Brazos Valley community health survey conducted in 2010 indicated that the rates of those with depressive symptoms matching either "other depressive syndrome" or a "major depressive syndrome" were 10.4% for Whites, 24.6% for Blacks, and 12.9% for Hispanics (Brossart et al., 2013).

In consideration of such disparities and a myriad of barriers to mental health service availability and accessibility in rural Central Texas, the Brazos Valley Health Partnership (BVHP) was developed in 2002 by a group of community leaders, stakeholders, and service providers in conjunction with the CCHD at the TAMU Health

Sciences Center to provide sustainable healthcare services in the Brazos Valley. The BVHP established a primary care health facility to serve rural residents in a county that covers 1072 square miles, has a population of approximately 16, 344 residents (Wendel et al., 2011), and an average distance of 45.2 miles to a medical care center (CCHD, 2010). Along with medical care, this facility provides behavioral telehealth services via videoconferencing with doctoral students in the TAMU Counseling Psychology program (McCord et al., 2011; Wendel et al., 2011).

Participants

Participants were 52 patients (women = 40, men = 12) receiving healthcare services from a primary care facility in a rural town in the Brazos Valley (Texas) during a one year period between June 1, 2011 and August 15, 2012. During this period, 11 patients (women = 6, men = 5) dropped out of treatment after only one or two sessions and were therefore excluded from the study. The remaining 41 patients (women = 34, men = 7) who attended three or more sessions were included in the study. Patients attended an average of 11.10 sessions (*SD* = 7.89), with a range of 3-40 sessions for the entire sample. Table 1 presents the demographic information for the sample of 41 patients.

Table 1. Demographics of patients receiving treatment.

Patient Characteristics	Women (n = 34)	Men (n =7)	Total (n = 41)
Age			
M	41.12	40.57	41.02
SD	12.87	12.88	12.71
Range	14-57	15-55	14-57
Ethnicity			
Caucasian	29 (70.50%)	6 (14.50%)	35 (85%)
African-American	2 (5%)	0	2 (5%)
Latino	2 (5%)	0	2 (5%)
Asian	0	1 (2.50%)	1 (2.50%)
Biracial	1 (2.50%)	0	1 (2.50%)
Employment Status			
Employed	3	0	3 (7.30%)
Unemployed	18	5	23 (56.10%)
Disabled (SSI)	9	1	10 (24.40%)
Students	4	1	5 (12.20%)
Diagnoses (i.e. comorbidity)			
Depressive Disorder	28	6	34
Post-Traumatic Stress Disorder	8	1	9
Panic Disorder	9	1	10
Anxiety Disorder	7	1	8
Bipolar Disorder	2	0	2
Schizophrenia	0	2	2
Bereavement	5	3	8
ADHD	0	3	3
Pain [a]	9	2	11
Comorbidity	24 (60%)	6 (15%)	30 (75%)
Taking Psychotropic Medications	32 (78%)	6 (15%)	38 (93%)
Sessions			
M	11.18	10.14	11.10
SD	8.21	5.79	7.79
Range	3-40	3-17	3-40

Note. *M* = mean; *SD* = standard deviation; SSI = social security insurance disability benefits. [a] Pain was not diagnosed as a psychological disorder, but was a presenting concern causing significant psychological distress.

Eighty-five percent of patients were of Caucasian ethnicity, 5% African American, 5% Latino, 2.50% Asian, and 2.50% were biracial. Ages ranged from 14-57 years old with a mean age of 41.02 (*SD* = 12.71). Seventy-five percent of patients were diagnosed with two or more disorders. Fifty-six percent of the study sample reported being unemployed, 24% were receiving social security benefits for a disability, 12% were students, and only 7% of patients were employed. All patients received behavioral telehealth services free of cost. According to standard clinic procedures, all patients are first evaluated by a physician before being referred for behavioral telehealth treatment. As a result, 93% of the sample was treated with psychotropic medications in addition to behavioral telehealth services.

Four patients were selected from the sample for analysis by SCR methods because they met basic WWC standards for a SCES (Kratochwill et al., 2010). The WWC standards are intended to promote methodological soundness of a SCES. Specifically, the WWC selection criteria used for inclusion of the four patients for analysis by SCR methods required that patients had a minimum of three baseline and five intervention data points per phase on the CORE-B measure.

All four patients analyzed by SCR methods were Caucasian women who reported being unemployed. Two women reported they received social security disability benefits, one woman was a student, and the other woman was a stay-at-home mother. The mean age for these four female patients was 37.75 years (*SD* = 9.59). Three of the four women were diagnosed with a severe clinical mood disorder (i.e., PTSD, Bipolar Disorder, MDD) along with another psychological disorder and/or physical pain. All

three women diagnosed with a severe clinical mood disorder also reported a history of trauma (i.e., sexual and/or physical abuse) and were taking psychotropic medications. The other woman analyzed by SCR methods was diagnosed with an Adjustment Disorder with mixed anxiety and depression, but also reported physical pain and was taking medication for her pain symptoms. These four patients participated in an average of 14.75 (*SD* = 1.65) sessions of TVCP.

Treatment

TVCP was conducted weekly by two (one male, one female) Master's level practitioners (mean experience = three years) enrolled in a counseling psychology doctoral program under the supervision of a licensed psychologist. The treatment intervention is considered "treatment-as-usual" because treatment fidelity was not evaluated and treatment was tailored to individual patient needs. Theoretical orientations and techniques included cognitive-behavioral, existential-humanistic, biopsychosocial, and psychodynamic-interpersonal. Each intervention was tailored for the individual patient's needs depending on the practitioner's theoretical style and patient diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders Text Revision (DSM-IV-TR; American Psychiatric Association, 2000). Patients were diagnosed by their therapist after their third session and a treatment plan was established based on the treatment plan, patient goals, presenting concerns, and diagnoses. Weekly supervision and case consultation was provided by a licensed psychologist.

Outcome Measures

Clinical Outcomes in Routine Evaluation – Short Form B (CORE-B; CORE Systems Group, 1998; Evans et al., 2002). The CORE-B is an 18-item questionnaire derived from the longer, 34-item CORE-Outcome Measure (CORE-OM) that was developed in the United Kingdom to inform practice based on evidence. This self-report instrument utilizes a 5-point Likert scale (*0 = not at all, 1 = only occasionally, 2 = sometimes, 3 = often, 4 = most or all of the time*) to assess for *Global Distress* experienced by a patient over the last week. Factor analysis of the CORE-OM and CORE-B suggest that both measures assess four domains of general mental health, including *Symptoms* (anxiety, depression, trauma, and physical symptoms), *Risk* (of harm to self and others), *Well-being* (affective state, self-esteem, and coping) and *Functioning* (general, interpersonal, and coping; Evans et al., 2002). Total scores on the CORE-B range between 0 and 72, with mean scores (between 0 and 4) calculated for each domain (subscale) and the Global Distress scale. Higher scores on all scales indicate greater psychological distress.

Several validity studies of the CORE have been conducted in a variety of primary and secondary care settings with several different types of populations (Barkham, Culverwell, Spindler, Twigg, & Connell, 2005; Barkham, Gilbert, Connell, Marshall, & Twigg, 2005; Connell et al., 2007; Evans, Connell, Barkham, Marshall, & Mellor-Clark, 2003; Mellor-Clark, Connell, Barkham, & Cummins, 2001). Studies of the CORE-OM suggest this measure exhibited convergent validity with the Beck Depression Inventory (r = .85; Evans et al., 2002) and the Beck Anxiety Inventory (r = .65; Evans et al., 2002)

as well as discriminant validity from the Beck Hopelessness Scale (r = .34; Cahill et al., 2006) and the Inventory of Interpersonal Problems – Avoidant (r = .45; Cahill et al., 2006). The CORE-Short Forms have exhibited convergent validity with the Beck Depression Inventory – II (r = .88; Cahill et al., 2006). Test-retest reliability for the CORE-OM has been reported at r = .90 (Barkham et al., 2001). Various investigations have also reported good internal consistency reliability of the CORE-OM with coefficient alphas (α) ranging from .94 (Evans et al., 2002) to .95 (Barkham, Gilbert et al., 2005) for all items, and ranging from .75 to .94 across the four domains (Evans et al., 2002). Internal consistency analyses for this sample produced coefficient α = .89 for the CORE-B obtained during the intake session.

Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer, & Williams, 2001). The PHQ-9 was used to assess patients' depressive symptoms. The nine questions on the PHQ-9 reflect DSM–IV-TR criteria for depressive disorders (Kroenke et al., 2001). The PHQ-9 is a 9-question self-report instrument with a 4-point Likert scale (*0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly everyday*) that asks respondents to rate the frequency of mental health symptoms they experienced over the previous 2-week period. Scores on the PHQ-9 range from 0–27, with scores between 0 and 4 indicating the absence of depression, 5–9 *mild* depression, 10–14 *moderate* depression, 15–19 *moderately severe* depression, and ≥ 20 *severe* depression (Kroenke et al., 2001).

The PHQ-9 has been used with numerous populations (Kroenke, Spitzer, Williams, & Lowe, 2010). Studies of the PHQ-9 suggest this measure has exhibited convergent validity with the Mental Component Summary score of SF-36 Health Survey

(*r* = -.68; Milet, Hudson, Baron, & Thombs, 2010), mental health scale of the SF-20 Health Survey (*r* = .73; Kroenke et al., 2001), and the Beck Depression Inventory – II (*r* = .72; Titov et al., 2011). The PHQ-9 has also exhibited discriminant validity with Physical Health Component Summary score of SF-36 Health Survey (*r* = -.43; Milet et al., 2010) and the physical health scale of SF-20 Health Survey (*r* = .37; Kroenke et al., 2001). The PHQ-9 has exhibited excellent test-retest reliability (*r* = .94; Zuithoff, et al., 2010) and consistently high internal consistency with coefficient alphas ranging from .89-.92 (Kroenke et al., 2001). The internal consistency reliability of the PHQ-9 obtained during the intake session produced a coefficient α = .91 for this sample.

Short Form-12 General Health Survey (SF-12; Ware, Kosinski, & Keller, 1996; Ware, Kosinski, Turner-Bowker, & Gandek, 2002). The SF-12 (version 2) was used to assess patients' physical and mental health. This 12-item Likert-style self-report measure is a short version of the SF-36 General Health Survey (Ware & Sherbourne, 1992) which is commonly used to assess health related quality of life. The SF-12 has been validated in many studies within primary care and medical settings (Lenert, Sherbourne, Sugar, & Wells, 2000; Wells & Sherbourne, 1999) because it provides information about health issues that interfere with daily functioning across various domains. The Mental Component Summary (MCS) score was used as a general indicator of Mental Health and the Physical Component Summary (PCS) score was used as a general indicator of Physical Health. Both the PCS and MCS have scores ranging from 0-100, with a mean of 50 and a standard deviation of 10 in the general population (Ware et al., 1996). Higher

scores on each scale reflect higher quality of life along with less distress and perceived limitations in general life roles.

Although various normative studies have been conducted with the SF-12 and SF-36 for medical patient populations, there are no studies which provide evidence that these measures have sufficient specificity and sensitivity to determine the presence of any particular mental health diagnosis. As such, there are no clearly established cut-off values that would differentiate clinical populations with mental health disorders from non-clinical, or psychologically healthy, populations. Thus, analysis of clinically significant and reliable change via the Jacobson and Truax (1991) method was not used with this measure. However, one large medical outcome study (Ware et al., 1996) showed varying ranges of MCS and PCS mean scores that included a comparison of patient samples with minor medical problems (MCS = 53.82, PCS = 47.42), serious physical health problems only (MCS = 52.51, PCS = 38.75), serious mental health problems only (MCS = 37.03, PCS = 49.32), and serious mental and physical problems combined (MCS = 43.18, PCS = 36.34). These categories and mean scores were used for comparison to the rural sample in this study.

Research examining the convergent and discriminant validity of the SF-12 with psychological measures is limited. However, one study by Milet et al. (2010) found that the MCS of SF-36 Health Survey exhibited convergent validity with the PHQ-9 (r = -.68) while the PCS exhibited discriminant validity with the PHQ-9 (r = -.43). Test-retest reliability coefficients for the PCS and MCS have ranged between r = .86 - .89 and r = .76 - .77, respectively (Ware et al., 1996). Internal consistency reliabilities of the MCS in

past studies have ranged between α = .73 - .79 and α = .78 - .82 for the PCS (Larson, 2002; Larson, Schlundt, Patel, Beard, & Hargreaves, 2008). Internal consistency reliabilities of the SF-12 MCS and PCS subscales (abbreviated hereafter as MCS-12 and PCS-12) obtained during the intake session for this sample were α = .78 and α = .89, respectively.

Procedures

Timeline and Measurement Schedule. Figure 1 displays the study timeline, measurement schedule, and the flow of participants through the study. The clinic was established on March 2009. Upon establishment of the clinic, routine assessment included the PHQ-9 and the SF-12 administered during the intake session and every four sessions thereafter. The current study began on June 1, 2011 and continued to use the established measurement schedule. Beginning September 2011, the current study incorporated the CORE-B into the routine clinical assessment. At that time, additional CORE-B baseline assessments were included prior to the intake session and routine administration of the CORE-B was conducted every two sessions during treatment. All patients treated in the clinic between June 2011 and August 2012 were included in the current study.

Baseline Assessment. During the baseline assessment period (Phase A), patients were screened by the clinic administrator with the CORE-B upon initial contact with the primary care facility. Following a physician's referral for behavioral telehealth services, the counselor attempted to contact the patient to confirm the therapy appointment. When patients were reached by telephone, the counselor confirmed the appointment and

administered the CORE-B verbally. Not all patients could be contacted for initial screening assessments with the CORE-B, but all patients completed the CORE-B in addition to the PHQ-9 and SF-12 prior to beginning the initial intake session. Of the twenty-three patients who completed the CORE-B, seven were contacted for three baseline assessments, eleven completed two baseline assessments, and five were assessed only once during the intake session.

Design

This study utilized a within-subjects group design to evaluate the effectiveness of TVCP for this rural sample at pre- and post-treatment periods. The addition of a planned baseline period before the intake pre-treatment assessment and a repeated measurement schedule allowed for single-case quantitative analyses of four patients with AB contrasts. Based upon the number of phases (eight) and data points per phase (three baseline, five intervention), the single-case design *Meets Standards with Reservations* for methodological soundness (Kratochwill et al., 2010). Single-case and group results were compared as a mixed-methods approach to provide a multidimensional perspective.

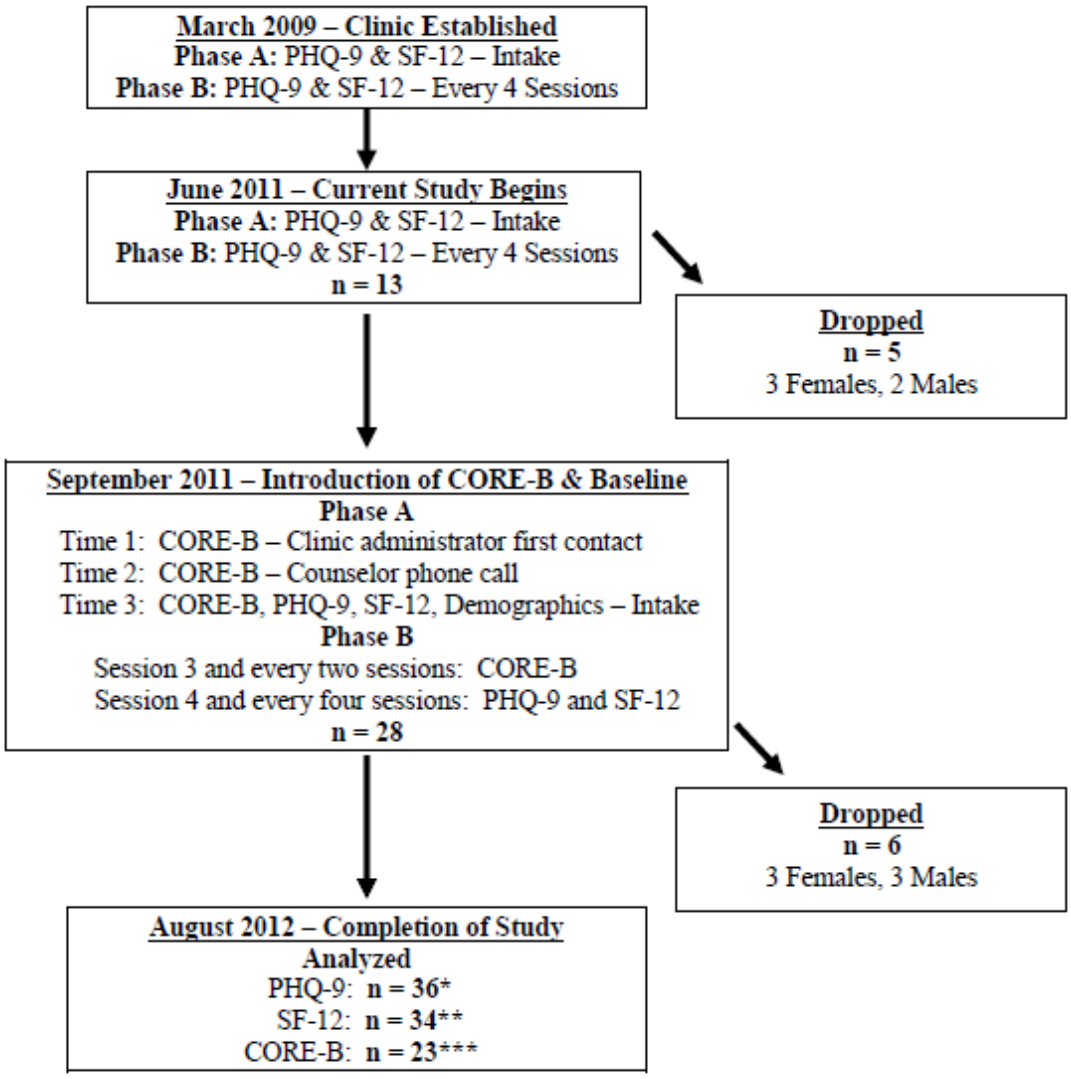

Figure 1. Study timeline, measurement schedule, and flow of participants through the study. Phase A = pre-treatment; Phase B = treatment. * Five patients participated in three treatment sessions and did not complete the PHQ-9. ** Two Spanish speakers did not complete SF-12 or the CORE-B. ***Thirteen patients entered treatment prior to introduction of the CORE-B and did not complete the CORE-B.

Analyses

Group. Group research analyses included paired-samples *t* tests of pre- and post-treatment mean scores on the PHQ-9, MCS-12, PCS-12, and each of the CORE domains

(Global Distress, Symptoms, Well-Being, Functioning, and Risk). Cohen's (1988) *d* effect size was derived from the *t* statistic for the paired-samples using the following formulas from Rosenthal (1991): $d = t / \sqrt{n}$, where *t* is the *t* statistic and *n* equals the number of pairs. Furthermore, bootstrap internal replicability analysis (re-sampling *N* = 1000) was conducted for each paired-samples *t* test to provide greater confidence in the stability of the sample's results (Thompson, 2006).

In addition to the main group analyses of pre- and post-treatment outcome scores, a two-way RMANOVA was conducted to evaluate the potential presence of a therapist effect over time on each of the outcome variables given the difference in therapists. The two-way RMANOVA with therapist (male, female) as the between-subjects effect and time (pre, post) as the within-subjects effect was conducted for each outcome variable. If a statistically significant therapist effect was found to exist, a one-way between-subjects ANOVA was then conducted to examine the difference in the severity of patients' pre-treatment scores in order to assure that the therapist effect was not due to this difference in scores.

Clinically Significant Change. Analysis of reliable improvement and clinically significant change (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991) for all individuals in the sample was conducted on the PHQ-9, CORE-B Global Distress scale, and the CORE-B subscales (Well-Being, Symptoms, Functioning, Risk). Several definitions of clinically significant change exist in the psychotherapy outcome literature (Bauer, Lambert, & Neilsen, 2004; Jacobson et al., 1984; Jacobson & Truax, 1991). The most prevalent method used to evaluate clinical significance of patient change is the

Jacobson et al. (1984) method (Bauer et al., 2004). This method has since been updated (Jacobson & Truax, 1991) and involves the following two steps: 1) identification and use of a normative cut-off value that differentiates between a patient/dysfunctional population and a non-patient/functional population; and 2) determination of a reliable change index (RCI) which signifies a patient's change from pretest to posttest is statistically reliable, and therefore not an artifact of measurement error.

Based on the two criteria (cut-off value and RCI), individuals are classified as *Recovered* (passed both cut-off value and RCI criteria), *Improved* (passed RCI criterion but not the cut-off), *Unchanged* (passed neither criteria), or *Deteriorated* (passed RCI criterion but worsened; Jacobson & Truax, 1991). Using the Jacobson and Truax (1991) method, RCI was calculated as 1.96 times the standard error (*SE*) of the difference according to the formula $SE = SD\sqrt{1-\alpha}$, and utilizing the internal consistency coefficient alphas (α) and standard deviations (*SD*) obtained during initial intake administration of the PHQ-9 (α = .92; *SD* = 6.23) and with the CORE-B Global Distress scale (α = .89; *SD* = .81), Well-Being (α = .89; *SD* = .85), Symptoms (α = .89; *SD* = 1.06), Functioning (α = .89; *SD* = 1.00), and Risk (α = .89; *SD* = .82) subscales.

The method of establishing clinically significant change involved the use of normative cut-off values which distinguished clinical and non-clinical samples. These clinical and non-clinical cut-off values were derived from samples used in normative studies during the development of the PHQ-9 and CORE-B outcome measures. These clinical cut-off values were provided in the PHQ-9 official scoring algorithm (Kroenke et al., 200l; PHQ, 2012) and CORE manual (CORE Systems Group, 1998; Evans et al.,

2002). Normative cut-off values used were as follows: PHQ-9 (≤ 4), CORE Global Distress scale (male = 1.19, female = 1.29), CORE Well-Being subscale (male = 1.37, female = 1.77), CORE Symptoms (male = 1.44, female = 1.62), CORE Functioning (male = 1.29, female = 1.30), and CORE Risk subscale (male = .43, female = .31). Patients were considered to make clinically significant change (Recovered) if they were in the clinical/dysfunctional range at any point during Phase A and moved to the non-clinical/functional range upon their final assessment.

Single-Case. Single-case quantitative analyses were conducted for four individual cases selected to meet the WWC standards of three baseline and five intervention data points for single-case research methods prescribed by Kratochwill et al. (2010). Analyses of the four cases included (a) graphing of each patient's CORE-B Global Distress scale scores over time by phase, (b) visual analysis of graphed data by two raters, (c) and simple mean shift regression of all CORE-B scale scores. In addition, analysis of direction and magnitude of autocorrelation (r_{auto}) or serial dependency for the entire series was conducted to understand the influence of time-series data on *p* values for statistical tests on the CORE-B Global Distress scale. Significant positive autocorrelation (r_{auto}*) is troublesome because it suggests that the *p* values are too low (optimistic) and the confidence intervals are too narrow, thus overestimating Type I error (Busk & Marascuilo, 1988; Matyas & Greenwood, 1990, 1996; Parker & Brossart, 2003).

Visual analysis of the four patients' CORE-B Global Distress scale graphed data was conducted by two raters using four categories of effect sizes, namely *none*, *small*,

moderate, and *large*. Visual analysis is an important method for judging the effectiveness of SCR time-series data which can be combined with statistical techniques to improve reliability of judgments (Brossart, Parker, Olson, & Mahadevan, 2006). Inter-rater reliability for the two raters was calculated at $\alpha = 1.00$ using classical test theory with ratings for the four patients scaled as item-level data.

Simple mean shift regression was used with patients' CORE-B Global Distress scale and CORE-B subscale scores as dependent variables and phase as the independent variable, coded as zero for the baseline and one for the treatment phase. This statistical model compares the mean level difference between the two phases (A vs. B) for each patient without including trend (Cohen & Cohen, 1983). This model produces a *t* statistic which was used to derive the effect size Cohen's *d* by using the formula from Rosenthal (1991): $d = 2t / \sqrt{df_{error}}$, where *t* is the *t* statistic and *df* equals the degrees of freedom (error). This statistical analysis was applied to the CORE-B Global Distress scale and subscales for the four individual patients analyzed with single-case research methods.

Effect sizes were provided with a more detailed description of each patient's treatment response that included markers of reliable and clinically significant change (Recovered, Improved, Unchanged, or Deteriorated) for all CORE-B scales and the PHQ-9. In addition, pre- and post-treatment scores on the MCS-12 and PCS-12 were also included as additional indicators of treatment response. Lastly, relevant patient and therapy characteristics (i.e., demographics, diagnoses, presenting problems, number of sessions, and goals) were also integrated to give the reader context about each patient and the treatment process. Together, quantitative results and a qualitative description of

treatment factors and patient context can be used by the reader to judge the overall effectiveness of TVCP for each patient.

CHAPTER IV

RESULTS

Missing Data

Overall, total missing data for all measures was minimal given that a large portion of assessment was therapist-administered. Missing data was most likely the result of patients overlooking items when completing the measures on their own after session. On the PHQ-9, 1% of total PHQ-9 data was missing. Total scores for the PHQ-9 were summed despite missing item data. Consequences of dealing with missing data in this fashion for individual PHQ-9 scores would result in lower depression scores. On the SF-12, 0.1% of total SF-12 data was missing. Total scores on the SF-12 were summed despite missing item-level data and would result in slight improvements in the scores of perceived quality of mental (MCS) and physical (PCS) health. On the CORE-B, 2% of total data was missing. Mean scale scores on the CORE-B Global Distress and subscales were calculated to adjust for individual missing items. Depending on the scale and missing item, CORE-B scores may suggest over or under corrections to their score. However, obtaining a mean scale score mitigates the impact of missing items.

Group Results

Table 2 presents the results for paired-samples *t* test analyses of patients' pre- and post-treatment mean scores on the PHQ-9, CORE-B Global Distress scale, MCS-12, and PCS-12. Figure 2 presents a graphical display of these pre- and post-treatment results. Table 3 presents the results for paired-samples *t* test analyses of patients' pre- and post-treatment mean scores on the four CORE-B subscales (Well-Being, Symptoms,

Functioning, Risk). A graphical display of these pre- and post-treatment results is presented in Figure 3. Included in both tables are mean difference statistics with 95% confidence intervals (*CI*), inference statistics (significance set at $p = .05$), and Cohen's *d* effect sizes.

Bootstrap analyses were conducted for all paired-samples *t* tests using a re-sampling of $N = 1000$. Results of the bootstrap analyses for all outcomes (PHQ-9, CORE-B scales, MCS-12, and PCS-12) suggested confidence that the mean difference statistics and standard errors were stable, indicating that results were not overly influenced by sampling error variance. All *p* values for the bootstrap paired-samples *t* tests were below the established $p = .05$ significance level and all confidence intervals were comparable to those of the original paired-samples *t* tests. These results suggested confidence that paired-samples *t* test results would replicate with future similar samples.

Table 2. Paired-samples *t* test results for PHQ-9, CORE-B Global Distress scale, MCS, and PCS.

	PHQ-9	**CORE-B**	**MCS**	**PCS**
N	36	23	34	34
M_1 (*SD*)	17.42 (6.23)	2.40 (0.72)	31.08 (13.18)	45.47 (15.94)
M_2 (*SD*)	9.58 (6.52)	1.61 (0.92)	41.39 (15.28)	41.73 (14.82)
M_{diff} (*SE*)	7.84 (6.07)	0.78 (0.75)	-10.31 (16.18)	3.74 (9.74)
M_{diff} 95% *CI*	5.78, 9.89	0.46, 1.11	-15.95, -4.66	0.34, 7.14
t (*df*)	7.75 (35) ***	5.02 (22) ***	-3.72 (33) *	2.24 (33) *
D	1.29	1.05	-0.64	0.38

Note. The variation in sample size resulted from (a) the initiation of the study with the PHQ-9 prior to introducing the CORE-B, and (b) Spanish speakers who completed the PHQ-9 but not the CORE-B or SF-12. M_1 = pre-treatment mean; M_2 = post-treatment mean; *SD* = standard deviation; $M_{diff.}$ = mean difference; *SE* = standard error of mean difference; *CI* = confidence intervals; *t* = *t* statistic; *df* = degrees of freedom; *d* = Cohen's *d*. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

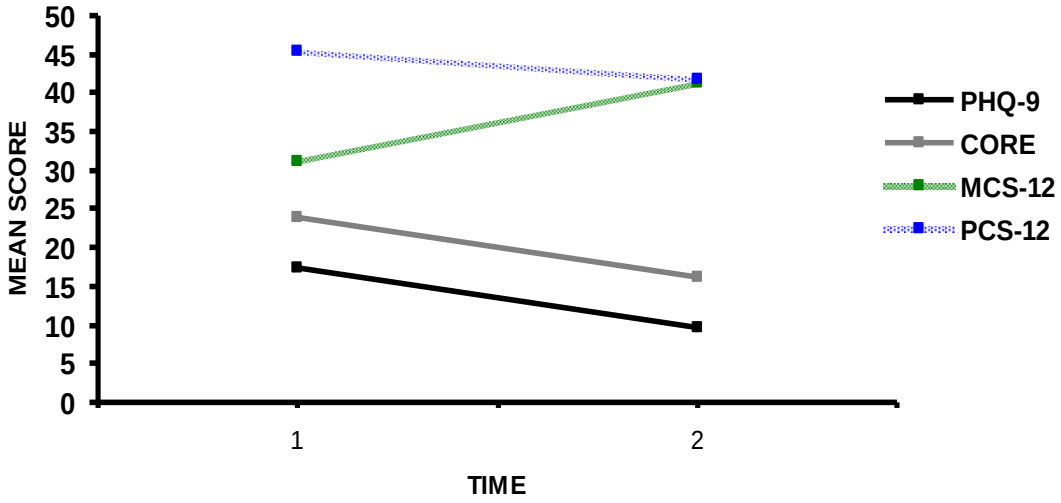

Figure 2. Paired-samples *t* tests results for PHQ-9 (0-27), CORE-B Global Distress scale (0-40), MCS-12 (0-100), and PCS-12 (0-100). Time 1 = Pre-treatment assessment; Time 2 = Post-treatment assessment.

Table 3. Paired-samples *t* tests results for CORE-B subscales.

	WELL-BEING	**SYMPTOMS**	**FUNCTIONING**	**RISK**
N	23	23	23	13
M_1 (*SD*)	2.77 (0.83)	2.39 (0.83)	2.41 (0.88)	1.15 (0.80)
M_2 (*SD*)	1.75 (1.11)	1.90 (1.20)	1.57 (1.11)	0.42 (0.70)
M_{diff} (*SE*)	1.02 (0.98)	0.49 (1.06)	0.84 (0.98)	0.73 (0.92)
M_{diff} 95% *CI*	0.60, 1.45	0.03, 0.95	0.41, 1.26	0.17, 1.29
t (*df*)	4.98 (22) ***	2.22 (22) *	4.09 (22) ***	2.84 (12) *
D	1.04	0.46	0.85	0.79

Note. The variation in sample size is due to clients not endorsing risk items at pre-treatment. M_1 = pre-treatment mean; M_2 = post-treatment mean; *SD* = standard deviation; M_{diff}. = mean difference; *SE* = standard error of mean difference; *CI* = confidence intervals; *t* = *t* statistic; *df* = degrees of freedom; *d* = Cohen's *d*.
* = $p < .05$, ** = $p < .01$, *** = $p < .001$.

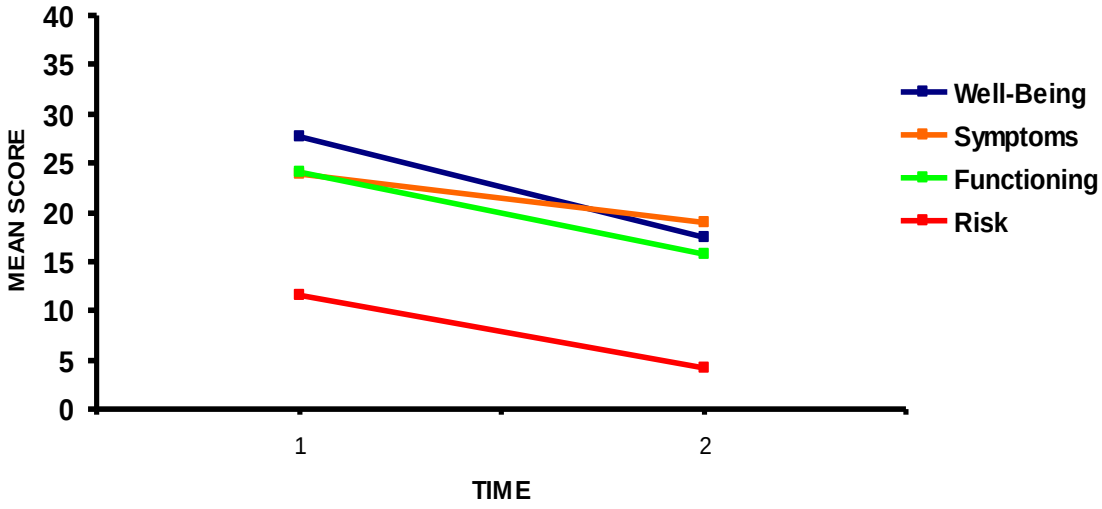

Figure 3. Paired-samples *t* tests results for CORE-B subscales (0-40): Well-Being, Symptoms, Functioning, and Risk. Time 1 = Pre-treatment assessment; Time 2 = Post-treatment assessment.

PHQ-9: The PHQ-9 group mean for this sample (n = 36) was in the moderately severe range of depression symptoms at pre-treatment. Paired-samples *t* tests results of patients' PHQ-9 scores indicated a statistically significant and large reduction in depression symptoms after treatment, t (35) = 7.75, $p < .001$, d = 1.29. This mean reduction in depression symptoms resulted in group improvement into the mild range of depression symptoms.

Table 4 presents results for patients who Recovered and Improved according to the PHQ-9. Of the 36 patients who completed the PHQ-9 over the course of treatment, one did not initially meet the clinical criteria necessary to calculate clinical and reliable

change. Thus, of the 35 remaining patients, 23% Recovered, 46% Improved, 31% remained Unchanged, and 0% Deteriorated.

Table 4. Clinically Significant and Reliable Change – PHQ-9

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	8	0	8
Not Sufficient Change	16	11	27
Not in Clinical Range at Onset	0	1*	1
Total	24	12	36

Classifications	
% Recovered	(8/35) = 23%
% Improved	(16/35) = 46%
% Unchanged	(11/35) = 31%
% Deteriorated	(0/0) = 0%

Note. *One patient was not included in the total count for either clinically significant or reliable change because they did not score in the clinical range at onset of treatment.

CORE-B Global Distress: Paired-samples *t* tests results of patients' (n = 23) mean CORE-B Global Distress scale scores indicated a statistically significant reduction in overall psychological distress after treatment, t (22) = 5.02, $p < .001$, d = 1.05. The results of patients who Recovered and Improved according to their mean scores on the CORE-B Global Distress scale are presented in Table 5. Of the 23 patients who completed the CORE-B over the course of treatment, one patient did not initially meet the clinical criteria necessary to calculate clinical and reliable change. Thus, of the 22

remaining patients, 32% Recovered, 27% Improved, 41% remained Unchanged, and 0% Deteriorated.

Table 5. Clinically Significant and Reliable Change – CORE-B Global Distress scale.

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	7	0	7
Not Sufficient Change	6	9	15
Not in Clinical Range at Onset	0	1*	1
Total	13	10	23

Classifications	
% Recovered	(7/22) = 32%
% Improved	(6/22) = 27%
% Unchanged	(9/22) = 41%
% Deteriorated	(0/0) = 0%

Note. *One patient was not included in the total count for either clinically significant or reliable change because they did not score in the clinical range at onset of treatment.

CORE-B Risk: Twenty-three patients completed the CORE-B over the course of treatment, but 10 patients never endorsed risk during pretreatment assessment. Paired-samples *t* tests results of patients' (n = 13) mean CORE-B Risk subscale scores indicated a statistically significant reduction in level of risk after treatment, $t\ (12) = 2.84$, $p < .05$, $d = 0.79$. Table 6 presents the number of patients who Recovered and Improved according to their mean scores on the CORE-B Risk. Eleven patients did not initially meet the clinical criteria necessary to calculate clinical and reliable change on this scale.

Thus, of the 12 remaining patients, 58.33% Recovered, 8.33% Improved, 33.33% remained Unchanged, and 0% Deteriorated.

Table 6. Clinically Significant and Reliable Change – CORE-B Risk subscale.

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	7	0	7
Not Sufficient Change	1	4	5
Not in Clinical Range at Onset	0	11*	11
Total	8	15	23

Classifications	
% Recovered	(7/12) = 58.33%
% Improved	(1/12) = 8.33%
% Unchanged	(4/12) = 33.33%
% Deteriorated	(0/0) = 0%

Note. *Eleven patients were not included in the total count for either clinically significant or reliable change because they did not score in the clinical range at onset of treatment.

CORE-B Well-Being: Paired-samples *t* tests results of patients' (n = 23) mean CORE-B Well-Being subscale scores indicated a statistically significant improvement in emotional well-being after treatment, t (22) = 4.98, $p < .001$, d = 1.04. Furthermore, results of patients who Recovered and Improved according to their mean scores on the CORE Well-Being subscale are presented in Table 7. All 23 patients who completed the CORE-B over the course of treatment met the cut-off criteria necessary to calculate

clinical and reliable change on this subscale. Of these 23 patients, 52% Recovered, 13% Improved, 35% remained Unchanged, and 0% Deteriorated.

Table 7. Clinically Significant and Reliable Change – CORE-B Well-Being subscale.

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	12	0	12
Not Sufficient Change	3	8	11
Not in Clinical Range at Onset	0	0	0
Total	15	8	23

Classifications	
% Recovered	(12/23) = 52%
% Improved	(3/23) = 13%
% Unchanged	(8/23) = 35%
% Deteriorated	(0/0) = 0%

CORE-B Symptoms: Paired-samples *t* tests results of patients' (n = 23) mean CORE-B Symptoms subscale scores indicated a statistically significant reduction in symptoms after treatment, t (22) = 2.23, $p < .05$, $d = 0.46$. The results of patients who Recovered and Improved according to their mean scores on the CORE-B Symptoms is presented in Table 8. Of the 23 patients who completed the CORE-B over the course of treatment, two patients did not initially meet the clinical cut-off criteria necessary to calculate clinical and reliable change for the Symptoms subscale. Thus, of the 21

remaining patients, 29% Recovered, 33% Improved, 38% remained Unchanged, and 0% Deteriorated.

Table 8. Clinically Significant and Reliable Change – CORE-B Symptoms subscale.

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	6	0	6
Not Sufficient Change	7	8	15
Not in Clinical Range at Onset	0	2*	2
Total	13	10	23

Classifications	
% Recovered	(6/21) = 29%
% Improved	(7/21) = 33%
% Unchanged	(8/21) = 38%
% Deteriorated	(0/0) = 0%

Note. *Two patients were not included in the total count for either clinically significant or reliable change because they did not score in the clinical range at onset of treatment.

CORE-B Functioning: Paired-samples *t* tests results of patients' (n = 23) mean CORE-B Functioning subscale scores indicated a statistically significant improvement in patients' functioning after treatment, t (22) = 4.09, $p < .001$, d = 0.85. The results of patients who Recovered and Improved according to their mean scores on the CORE-B Functioning subscale are presented in Table 9. Of the 23 patients who completed the CORE-B over the course of treatment, two patients did not initially meet the clinical cut-off criteria necessary to calculate clinical and reliable change for the Functioning

subscale. Therefore, of the 21 remaining patients, 33% Recovered, 24% Improved, 43% remained Unchanged, and 0% Deteriorated.

Table 9. Clinically Significant and Reliable Change – CORE-B Functioning subscale.

	Reliable Change		
Clinically Significant Change	*Yes*	*No*	*Total*
Change to Normal Population	7	0	7
Not Sufficient Change	5	9	14
Not in Clinical Range at Onset	0	2*	2
Total	12	11	23

Classifications	
% Recovered	(7/21) = 33%
% Improved	(5/21) = 24%
% Unchanged	(9/21) = 43%
% Deteriorated	(0/0) = 0%

Note. *Two patients were not included in the total count for either clinically significant or reliable change because they did not score in the clinical range at onset of treatment.

SF-12: At pre-treatment, the group mean MCS-12 was almost two standard deviations below the mean, in a range much lower than patient samples with serious mental health problems (e.g., 37.03; Ware et al., 1996). Paired- samples *t* tests results of patients' (n = 34) mean MCS-12 indicated a statistically significant improvement in perceived mental health quality after treatment, t (33) = -3.72, $p < .05$, d = -0.64. On the PCS-12, the group pre-treatment mean was considered to be comparable to patient samples with minor medical problems (e.g., 47.42; Ware et al., 1996). Paired-samples *t*

tests results of patients' (n = 34) mean PCS-12 indicated a statistically significant decline in perceived physical health quality after treatment, t (33) = 2.24, $p < .05$, $d = 0.38$.

Therapist Effect: A two-way RMANOVA was conducted to evaluate the potential presence of a therapist effect over time on each outcome variable. The two-way RMANOVA with therapist (male, female) as the between-subjects effect and time (pre, post) as the within-subjects effect revealed a statistically significant therapist effect over time on the PHQ-9, F (1, 34) = 7.51, $p = .01$, $\eta_p^2 = .18$. This effect indicated that patients made larger mean reductions in depression symptoms with the male therapist ($M_{diff.}$ = 9.86, SE = 6.46) than with the female therapist ($M_{diff.}$ = 4.64, SE = 6.44) over the course of treatment. A one-way between-subjects ANOVA was conducted to ensure that the therapist effect was not due to significant differences in the severity of patients' pre-treatment PHQ-9 scores between therapists. The one-way between subjects ANOVA resulted in a statistically non-significant difference in severity of pre-treatment PHQ-9 scores between therapists, F (1, 34) = 0.04, $p = .84$, suggesting that the therapist effect was present on this outcome measure.

Further analyses indicated that the therapist effect was not present for the other outcome variables. The two-way RMANOVA revealed a statistically non-significant therapist effect over time on the MCS-12 (F (1, 32) = 2.35, $p = .14$, $\eta_p^2 = .07$), PCS-12 (F (1, 32) = 0.03, $p = .86$, $\eta_p^2 = .001$), and CORE-B Global Distress scale (F (1, 21) = 0.28, $p = .60$, $\eta_p^2 = .01$). The two-way RMANOVA also revealed the following statistically non-significant therapist effects over time on the all subscales of the CORE-B: Well-Being (F (1, 21) = 2.79, $p = .11$, $\eta_p^2 = .12$), Symptoms (F (1, 21) = 1.09, p =

.31, η_p^2 = .05), Functioning (*F* (1, 21) = 0.22, *p* = .64, η_p^2 = .01), and Risk (*F* (1, 21) = 1.04, *p* = .33, η_p^2 = .09).

Summary of Group Results: "The Average Patient"

The following is a description of the characteristics and treatment response of the average patient in this sample derived from the group results for comparison with single-case results. If this average patient existed in this sample, the group results portray a Caucasian (85%) woman (83%) diagnosed with a moderately severe (*M* = 17.42, *SD* = 6.23) depressive disorder (83%) and a co-occurring psychological disorder or physical pain (75% comorbidity). She would have minor physical health issues (PCS: *M* = 45.47, *SD* = 15.94), but very poor mental health which interferes with her quality of life (MCS: *M* = 31.08, *SD* = 13.18). She would have participated in approximately 11 sessions of TVCP. Her mental health improvements after TVCP would result in statistically significant reductions in depression (PHQ-9: *d* = 1.29, *p* < .001), symptom distress (CORE-B Symptoms: *d* = 0.46, *p* < .05), and level of risk (CORE-B Risk: *d* = 0.79, *p* < .05), as well as improvements in emotional well-being (CORE-B Well-Being: *d* = 1.04, *p* < .001), and functioning (CORE-B Functioning: *d* = 0.85, *p* < .001). At post-treatment, she would have felt less distressed (CORE-B Global Distress: *d* = 1.05, *p* < .001) and impaired by her mental health quality (MCS: *d* = -0.64, *p* < .05), but would report experiencing a slight decline in perceived physical health quality (PCS: *d* = 0.38, *p* < .05).

Single-Case Results

The single-case results are presented below in a narrative format with a more detailed description of each patient's characteristics (i.e., gender, ethnicity, and diagnosis), situational context (i.e., presenting problem, historical data), and a brief description of treatment. Each patient's treatment response is presented with various effect sizes and indictors of change. Effect sizes are derived from (a) visual analysis of judges' ratings (none, small, medium, or large) of CORE-B Global Distress scale graphs, and (b) simple mean shift regression (Cohen's *d*) of Phase A vs. Phase B mean differences of all CORE-B scales. Indicators of change include (a) pre- and post-treatment PHQ-9, MCS, and PCS scores, and (b) clinically significant and reliable change (Recovered, Improved, Unchanged, or Deteriorated) on all CORE-B scales and the PHQ-9 derived from pre- and post-treatment scores. Together, the effect sizes and indicators of change can be used by the reader to judge the overall effectiveness of treatment for each patient.

However, it is important to remind the reader that the following results are based upon statistical (*p* value), practical (visual ratings and simple mean shift regression), and clinical significance (pre- vs. post- mean cut-off scores) methods. Each of the analytic methods for assessing these three types of significance is valid and each has its own value, but each method differs from the others and should be interpreted in its proper context to understand the unique treatment response of each patient. For example, a patient may exhibit large improvements in mean scores between phases (practical significance = Cohen's *d)*, but this change may not be statistically significant ($p > .05$)

because statistical significance is influenced by sample size (i.e., number of data points). Such a change may or may not meet criteria for reliable improvement or clinically significant change because the Jacobson and Truax (1991) method is based solely on the pre- and post- treatment cut-off scores. In fact, even if a patient makes large improvements in mean scores between phases which result in statistically significant results, the pre- and post- treatment scores may indicate that a patient did not make reliable improvement or clinically significant change if their final score indicates a relapse in symptoms (the opposite of this statement is also true). Thus, the reader is cautioned to (a) read the single-case results in the context of the method used to assess significant change on each outcome, and (b) interpret findings from each method as valid and compatible with other methods of calculating significance.

Visual analysis was conducted by two raters on the graphed CORE-B Global Distress scale data of the four Caucasian female patients presented in Figure 4. This figure displays the CORE-B Global Distress scale mean scores (range = 0 - 4) for four patients with all scores multiplied by 10 (range = 0 - 40) to allow for better visual display (multiplication does not affect statistical parameters). The results of visual analysis (practical significance) of the CORE-B Global Distress scale by two raters are presented in the narrative below.

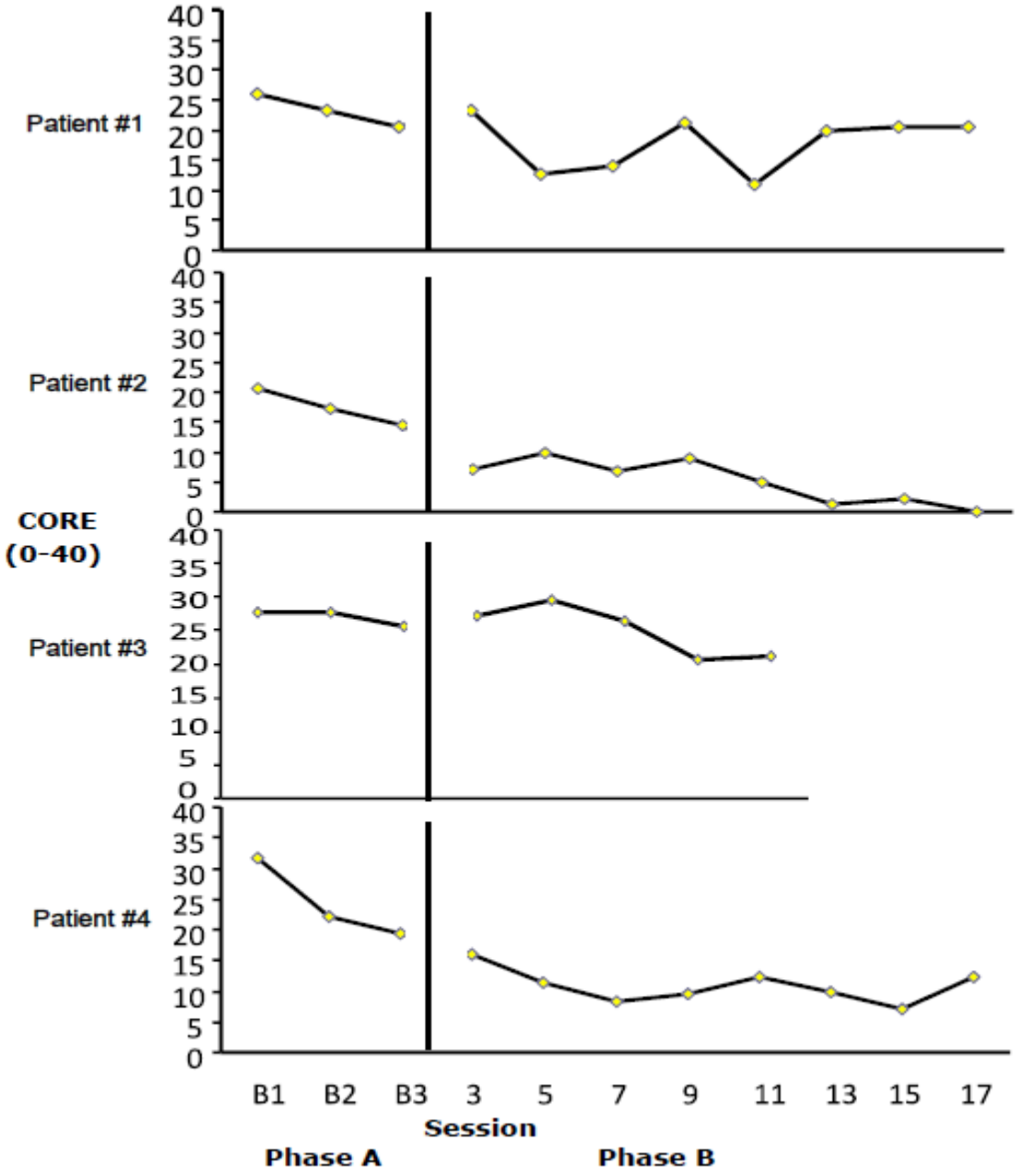

Figure 4. Single-case design graphs of CORE-B Global Distress scale (0-40) for four female patients. B1-B3 = Baseline points 1-3; B3 = intake session; Phase A = Pre-treatment; Phase B = Treatment.

Patient #1: Patient #1 was a stay-at-home mother of five children diagnosed with PTSD and MDD who participated in 17 treatment sessions. Her primary presenting concerns were anxiety, panic attacks, and depression that resulted from childhood sexual abuse and on-going domestic (physical and emotional) abuse from her husband that continued throughout the course of therapy. She was taking anti-depressant and anti-anxiety medication prior to beginning psychotherapy, but her symptoms continued at elevated levels. Her goals for psychotherapy included processing of childhood sexual abuse, better communication with her children, and coping with symptoms of anxiety. Treatment of Patient #1 involved learning coping skills for anxiety, emotional regulation, and self-care (e.g., relaxation exercises, time alone, sleep hygiene) as well as improving assertive communication skills with her children. In addition, the therapist provided psychoeducation about the effects of trauma and helped her process relationship issues.

Over the course of treatment, Patient #1 was judged by raters to have made small therapeutic gains in overall psychological distress according to visual analysis of the graphed CORE-B Global Distress scale scores. Although her score improvement on this scale approached statistical significance, the simple mean shift regression results of the CORE-B Global Distress scale indicated a non-significant mean difference between phases, t (1, 9) = 1.89, p = .09, d = 1.25, r_{auto} = .09. Simple mean shift regression results of her CORE-B subscale scores indicated that Patient #1 experienced statistically significant improvements in Well-Being (t (1, 9) = 3.20, p = .01, d = 2.13) and Risk (t (1, 9) = 3.13, p < .01, d = 2.09) along with statistically non-significant improvements in

Functioning (t (1, 9) = 1.70, p = .12, d = 1.13) and a decline in Symptoms (t (1, 9) = - 0.91, p = .39, d = - 0.61).

After treatment, Patient#1 was classified as Recovered according to the CORE-B Risk subscale (pre = 1.00, post = 0) and the PHQ-9 (pre = 18, post = 4), Improved on the CORE-B Functioning (pre = 2.67, post = 1.67) and Well-Being (pre = 4.00, post = 2.50) subscales, but remained Unchanged on the CORE-B Global Distress (pre = 2.61, post = 2.06) and Symptoms (pre = 2.50, post = 2.83) scales. Despite a slight decline in perceived physical health (PCS: pre = 58.40, post = 56.79), Patient #1 experienced large improvements in perceived mental health quality (MCS: pre = 26.10, post = 35.18). However, it is important to note that although Patient #1 improved on the MCS-12, she exhibited a post-treatment mean score on this measure in a range comparable to that of patient samples with serious mental health problems (e.g., 37.03; Ware et al., 1996). Overall, her results suggested that Patient #1 made partial improvements on some outcomes, but remained unchanged on others.

Patient #2: Patient #2 was a single woman diagnosed with an Adjustment Disorder with mixed anxiety and depressed mood who was also experiencing physical pain issues. Her primary presenting concerns were anxiety and depressive symptoms related to familial conflict and adjustment to the divorce of her parents. She was not prescribed psychotropic medication for her symptoms. Her goals for treatment included improving strained family relationships, adjusting to the stress of her parents' divorce, and improving her self-esteem. Treatment with Patient #2 utilized interpersonal and existential theoretical orientations.

After 15 sessions, Patient #2 was judged by raters to have made large therapeutic gains in overall psychological distress according to visual analysis of the graphed CORE-B Global Distress scale scores. The simple mean shift regression results suggested that Patient #2 made large statistically significant improvements between phases in overall psychological distress according to the CORE-B Global Distress scale, t (1, 9) = 5.11, $p < .001$, $d = 3.41$; $r_{auto} = .60$*. It is important to note that although gains on this scale are large, there was significant positive autocorrelation (*) of data for this patient that could have inflated p values. Simple mean shift regression results of the CORE-B subscale scores indicated that Patient #2 experienced statistically significant improvements in Well-Being (t (1, 9) = 3.00, $p = .01$, $d = 2.00$), Functioning (t (1, 9) = 6.14, $p < .001$, $d = 4.10$) and Risk (t (1, 9) = 3.13, $p = .01$, $d = 2.09$) along with a statistically non-significant improvement in Symptoms (t (1, 9) = 1.78, $p = .11$, d = 1.18).

After treatment, Patient #2 Recovered on the PHQ-9 (pre = 6, post = 1), CORE-B Global Distress scale (pre = 2.06, post = 0), and on all subscales of the CORE-B (Well-Being: pre = 3.00, post = 0; Symptoms: pre = 1.33, post = 0; Functioning: pre = 2.33, post = 0; Risk: pre = 2.00, post = 0). Furthermore, Patient #2 experienced improvements in perceived physical health (PCS: pre = 43.04, post = 47.92) and mental health (MCS: pre = 44.90, post = 57.48) quality after treatment. Taken together, her results suggested that Patient #2 made large therapeutic gains on all outcomes.

Patient #3: Patient #3 was a physically disabled, single woman diagnosed with MDD who was experiencing significant physical health issues (i.e., multiple bypass

surgeries, diabetes) along with pain and anxiety. This patient had limited social support and a history of sexual abuse. This patient was prescribed anti-depressant medication prior to beginning psychotherapy. She participated in 10 treatment sessions. Her goals for treatment included reducing symptoms of depression and anxiety while increasing her self-esteem. Treatment of Patient #3 included cognitive-behavioral and interpersonal therapy techniques to help her improve her self-esteem, maladaptive cognitions, and strained relationships.

Over the course of treatment, Patient #3 was judged by raters to have made small therapeutic gains in overall psychological distress according to visual analysis of the graphed CORE-B Global Distress scale scores. The simple mean shift regression results suggested that she made statistically non-significant improvements between phases in overall psychological distress according to the CORE-B Global Distress scale, t (1, 6) = 0.87, p = .42, d = 0.71; r_{auto} = .44. Simple mean shift regression results of the CORE-B subscale scores indicated that Patient #3 made statistically non-significant improvements in Well-Being (t (1, 6) = 0.19, p = .86, d = 0.16) and Functioning (t (1, 6) = 0.64, p = .55, d = 0.52) in addition to declines in Symptoms (t (1, 6) = -0.09, p = .93, d = -0.07) and Risk (t (1, 6) = -0.24, p = .82, d = -0.20).

After treatment, Patient #3 Improved on the Well-Being subscale (pre = 3.50, post = 2.25), but remained Unchanged on the CORE-B Global Distress scale (pre = 2.78, post = 2.11) and CORE-B Risk (pre = 0.50, post = 1.00), Functioning (pre = 3.17, post = 3.25), and Symptoms (pre = 3.17, post = 2.33) subscales. Patient #3 also experienced a slight decline in depression symptoms on the PHQ-9 (pre = 17, post = 21), but reported

very small improvements in perceived mental (MCS: pre = 33.55, post = 36.95) and physical health quality (PCS: pre = 19.83, post = 20.92). It is important to note that Patient #3 scored significantly worse on the PCS-12 than patient samples diagnosed with serious medical problems (PCS = 38.75; Ware et al., 1996), suggesting physical pain and pain interference of daily living could have negatively influenced treatment outcomes. Overall, her results suggested that Patient #3 made small therapeutic gains on only one outcome.

Patient #4: Patient #4 was a divorced woman diagnosed with Bipolar Disorder, PTSD, and a Substance Abuse Disorder who was also experiencing physical pain. Her primary presenting problems included anxiety, anger, and depression symptoms that resulted from a history of childhood sexual abuse and a long history of domestic abuse as an adult. She was taking a mood stabilizer for Bipolar Disorder. Her goals for treatment included reducing anxiety symptoms and anger. She was also interested in understanding relationship issues and patterns that contributed to domestic abuse and further anger. Treatment involved learning coping skills for emotional regulation, cognitive-behavioral techniques to improve assertive communication skills, and interpersonal counseling intended to promote insight about unhelpful relationship patterns.

After 17 counseling sessions, Patient #4 was judged by raters to have made large therapeutic gains in overall psychological distress according to visual analysis of the graphed CORE-B Global Distress scale scores. The simple mean shift regression results suggested that she made large statistically significant improvements between phases in

overall psychological distress as measured by the CORE-B Global Distress scale, t (1, 9) = 5.15, $p < .001$, $d = 3.34$; $r_{auto} = .55$. Simple mean shift regression results of the CORE-B subscale scores indicated that Patient #4 experienced statistically significant improvements in Well-Being (t (1, 9) = 5.15, $p < .001$, $d = 3.34$), Functioning (t (1, 9) = 7.39, $p < .001$, $d = 4.93$), and Risk (t (1, 9) = 4.22, $p < .01$, $d = 2.81$) along with improvements in Symptoms (t (1, 9) = 1.84, $p < .10$, $d = 1.23$) that approached statistical significance.

After treatment, Patient #4 Recovered on the CORE-B Global Distress scale (pre = 3.17, post =1.24) and on the CORE-B Well-Being (pre = 3.00, post = 1.67) and Functioning subscales (pre = 2.67, post = 0.67). In addition, Patient #4 Improved on the CORE-B Symptoms (pre = 3.33, post = 1.83) and Risk (pre = 2.50, post = 0.50) subscales. Despite a slight decline in perceived physical health quality (PCS: pre = 58.40, post = 56.79) and depression symptoms (PHQ-9: pre = 13, post =15), Patient #4 experienced large gains in perceived mental health quality (MCS: pre = 36.26, post = 51.89). Overall, her results suggested that Patient #4 made large therapeutic gains on the majority of mental health outcomes.

Comparison of Single-Case and Group Results

The characteristics (i.e., demographics, diagnostic profile, symptom severity, and situational context) and treatment responses (on outcome measures) of the four individual patients were compared to each other and the group. This comparison was intended to facilitate the discovery of clinical insights potentially unavailable with only one type of design. The comparison was improved by the fact that all four patients

analyzed by single-case research methods were Caucasian women, and thus highly representative of the larger sample with regards to ethnicity (85% Caucasian) and gender (83% female). Figure 5 was provided to enhance the clarity of how therapeutic changes were experienced by the four patients on each of the CORE-B scales, and how each patient's treatment response (Cohen's *d*) differed from each other and the average patient on these scales.

It is important to note that, while interpreting Figure 5, the two statistical models used for the group (paired-samples *t* tests) and single-case (simple mean shift regression) analyses are distinct, so results should not be directly compared to each other with regards to magnitude of effect. The magnitude of effect produced from group analyses should only be compared with other similar group studies that utilize similar statistical models to produce a comparable effect size. The magnitude of effect produced from single-case analyses can be compared between the four patients and with other single-case studies that use similar statistical models and measures.

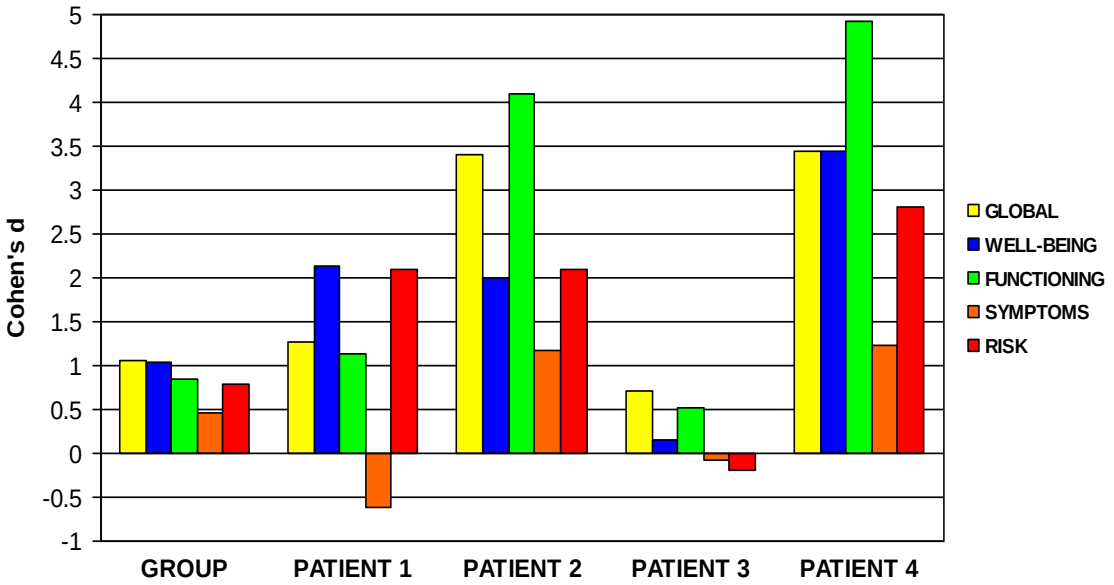

Figure 5. Comparison of treatment response for the group and four patients on CORE-B scales as standardized mean difference effect size, Cohen's *d*. The group effect size was derived from paired-samples *t* tests (pre vs. post-treatment means). Effect sizes for four patients were derived from simple mean shift regressions (Phase A vs. Phase B means).

Clinical Insight #1: Severity of symptoms and complexity of clinical presentation seemed to influence the degree of treatment response. One important observation to note among the subsample of four women and the overall group is the severity and complexity of clinical presentation, including types of diagnoses, presenting problems, and comorbidity. Much like the larger group sample (75% comorbidity), three of the four women were diagnosed with a severe clinical mood disorder along with another psychological disorder and/or physical pain. The severity of mental health issues appears to have influenced the degree of each patient's therapeutic progress.

For instance, Patient #2 made the largest and most consistent gains of the four cases. She Recovered on all scales of the CORE-B and on the PHQ-9 and experienced

large gains on the MCS-12. Although she experienced co-occurring physical pain (like three of the four cases) and depressive symptoms, her pre and post-treatment PCS-12 scores (pre = 43.04, post = 47.92) were consistent with patient samples who had minor medical problems (e.g., 47.42; Ware et al., 1996). Her PCS-12 scores suggested that pain may not have interfered with her daily functioning in a significant way. Furthermore, this patient had the lowest initial endorsement of depression symptoms (PHQ-9: pre = 6) and was the only case that did not report any history of sexual abuse. Thus, the fact that this patient only had mild depression and pain symptoms likely influenced her ability to make large, positive therapeutic gains on all mental health outcomes.

Clinical Insight # 2: Physical health problems seemed to negatively influence treatment response. In contrast to Patient#2, Patient #3 experienced the worst overall outcomes of the four patients. Like the group results, three of the four patients showed declines on the PCS-12. However, only Patient #3 endorsed significant physical health problems and pain which interfered with her quality of life and daily functioning. Her pre- and post-treatment PCS-12 scores (pre = 19.83, post = 20.92) were significantly lower than those of the group (PCS: pre = 45.47, post = 41.73) and patient samples with serious medical problems (e.g., 38.75; Ware et al., 1996), suggesting pain and physical disability were severely distressing. Despite physical health problems indicated on the PCS-12, Patient #3 Improved on the CORE-B Well-Being subscale.

Clinical Insight #3: Partial improvements were exhibited on some outcomes for all four patients. The group results show a statistically significant decline on the PCS-12 with improvements on all mental health outcomes. Similarly, the single-case results

showed that three of the four patients declined on the PCS-12, but all four patients exhibited partial improvements on some mental health outcomes. In contrast to the group treatment response, three of the four patients exhibited declines or miniscule gains on certain mental health outcomes. However, TVCP was effective for improving some mental health outcomes for all four patients.

Clinical Insight #4: Patients' diagnoses and situational contexts seemed to influence the outcomes on which they responded. Further observation suggests the partial improvements and specific outcomes on which patients responded were likely influenced by their diagnoses, presenting problems, and situational contexts. This clinical insight is best exemplified by comparing Patient #1 and Patient #4. Both patients were diagnosed with PTSD as a result of childhood sexual abuse, and both were victims of physical and emotional domestic abuse as adults.

Patient #1 Improved on the CORE-B Functioning and Well-Being subscales and Recovered on the PHQ-9 and CORE-B Risk subscale. She also experienced improvement on the MCS-12 after treatment (pre = 26.10, post = 35.18). However, her post-treatment MCS-12 was comparable to that of patient samples with serious mental health problems (e.g., 37.03; Ware et al., 1996). Considering that her pre- and post-treatment PCS-12 scores were in the normal range, but she experienced a decline on the CORE-B Symptoms subscale, it may be inferred that her anxiety symptoms may have persisted and caused continued emotional distress. Given that Patient #1 was not treated specifically for PTSD with an EST and was in an abusive domestic partnership during treatment, it was also possible that her situational context impacted her therapeutic

progress such that she experienced gains on some mental health outcomes (i.e., PHQ-9, Functioning, Well-Being, Risk), but continued to endorse distress on others (e.g., Symptoms) at post-treatment.

Patient #4 exhibited a similar pattern of partial therapeutic progress, but with different indicators. Like Patient #1, this patient had a history of childhood sexual abuse and adult physical and emotional domestic abuse, but unlike Patient #1 she was not involved in an abusive relationship during treatment. Also like Patient #1, her pre- and post-treatment PCS-12 were in the normal range, suggesting her distress was primarily psychological. In this case, Patient #4 Improved on the CORE-B Symptoms and Risk subscales and Recovered on the CORE-B Well-Being, Functioning, and Global Distress scales. She also experienced a prominent change in her MCS-12 (pre = 36.26, post = 51.89) which started in a range comparable to that of patient samples with serious mental health problems (e.g., 37.03; Ware et al., 1996) and ended in the normal range at post-treatment. Despite such improvements, she exhibited only a slight decline in depression symptoms (PHQ-9: pre = 13, post = 15). Given that she was diagnosed with Bipolar Disorder, it was inferred that the persistent, cyclical pattern of mood change caused by this diagnosis was responsible for continued symptoms of depression despite her improvements on all other mental health outcomes.

Much like the group, the single-case results of the four female patients suggested that despite the severity of their symptoms and the complexity of their clinical presentation, TVCP was generally effective for improving psychological functioning. The degree and domains on which the individual patients responded to treatment

depended on characteristics particular to the patient (i.e., diagnosis, presenting problem) and their situational context. Overall, the treatment response of the four patients examined using a single-case design was similar to the treatment response of the average patient represented by the group results. However, the single-case results provided a greater differentiation of findings and additional clinical insights beyond those obtained from group analyses.

CHAPTER V

CONCLUSIONS

Summary of Results

This study combined single-case and group quantitative research methods to (a) evaluate the effectiveness of TVCP for a rural sample in a primary care setting in Texas; and (b) illustrate the benefits of using a mixed-methods design to appeal to researchers and clinicians. Overall, results of this mixed-methods study indicated that TVCP was effective for improving the mental health of this rural sample – 75% of which were diagnosed with two or more co-occurring psychiatric disorders. Research questions one and two – regarding the effectiveness of TVCP – were addressed via a within-subjects group design. Research question three – regarding the treatment response of four individual patients – was addressed via the single-case experimental design. Finally, research questions four and five – regarding the clinical utility of combining research methods – were addressed by comparing the single-case and group results to derive important clinical insights about treatment response. The following is a discussion of how each research question was addressed in this mixed-methods study.

Research Question #1 asked: Will TVCP produce statistically significant group improvements in mean depression symptoms, global psychological distress, risk level, symptoms distress, emotional well-being, functioning, perceived mental health quality, and perceived physical health quality as measured by the PHQ-9, CORE-B scales (Global Distress, Risk, Symptoms, Well-Being, and Functioning), and the SF-12 (Mental

and Physical Health Component Summary scores)? What percentage of patients will make reliable and/or clinically significant change on the PHQ-9 and CORE-B scales?

Group results indicated that TVCP contributed to large, statistically significant reductions in depression symptoms for this sample as measured by the PHQ-9. In addition to the magnitude of effect produced by TVCP, reliable and clinically significant reductions in depression symptoms were experienced by a large majority of the sample. On the PHQ-9, 46% of patients had Improved depressive symptoms, 23% Recovered from depression, 31% remained Unchanged, and 0% Deteriorated. These findings suggest TVCP made a significant impact in this sample because 83% of patients had a depressive disorder prior to treatment. Such large scale reductions in depression are important to acknowledge given the high prevalence of depressive disorders that present in primary care (Hays, Wells, Sherbourne, Rogers, & Spritzer, 1995; Sherbourne et al., 2001; Wells & Sherbourne, 1999) and rural settings (Brossart et al., 2013), and the deleterious effects that depression has on physical and mental health (Hays et al., 1995).

The group also exhibited large, statistically significant reductions of overall psychological distress on the CORE-B Global Distress scale. On the CORE-B Global Distress scale, 27% of patients Improved, 32% Recovered, 41% remained Unchanged, and 0% Deteriorated. Analyses of CORE-B subscales resulted in similar levels of improvements, suggesting that TVCP was effective for patients on a variety of psychological domains.

One of the most important indicators impacted by TVCP was the level of risk. A large, statistically significant reduction in level of risk of harm to self or others was

reported by a sizeable proportion of the sample that began treatment with elevated risk scores. Specifically, 58% of patients who endorsed risk at pre-treatment Recovered, 8% Improved, 33% remained Unchanged, and 0% Deteriorated according to the CORE-B Risk subscale. Given that the rates of suicide for rural populations are often greater than in urban populations (Eberhardt et al., 2001; Eberhardt & Pamuk, 2004), these findings highlight the importance of how TVCP can help meet the mental health treatment needs of rural populations.

TVCP was also effective for producing a statistically significant reduction on the CORE-B Symptoms subscale, with 29% of patients Recovered, 33% Improved, 38% Unchanged, and 0% Deteriorated according to this measure. In addition, TVCP also proved effective for enriching the emotional well-being for this sample. Large, statistically significant improvements on the CORE-B Well-Being subscale were found, suggesting a shift toward a healthier affective state for the majority of patients. Fifty-two percent of patients Recovered, 13% Improved, 35% remained Unchanged, and 0% Deteriorated according to this subscale. Gains on the CORE-B Functioning subscale were also statistically significant for the sample, suggesting the enhancement of social relationships, general life roles, and coping abilities. Thirty-three percent of patients Recovered, 24% Improved, 43% remained Unchanged, and 0% Deteriorated according to the CORE-B Functioning subscale.

In addition, the group exhibited large, statistically significant improvements in overall perceived mental health quality according to the MCS-12. The group mean began in a range comparable to patient samples with serious mental health issues, but improved

by one standard deviation on the MCS-12 after TVCP. This large improvement suggests greater mental health quality and less interference of mental health symptoms in daily functioning.

In contrast to the overall improvements exhibited in mental health outcomes, the group results indicated a statistically significant decline on the PCS-12. This finding suggests that the sample perceived worsening physical health and interference of physical health issues in daily activities after treatment. Similarly, there were large percentages of patients that remained Unchanged according to the PHQ-9 (31%), CORE-B Global Distress scale (41%), CORE-B Risk subscale (33%), CORE-B Well-Being subscale (35%), CORE-B Symptoms subscale (38%), and CORE-B Functioning subscale (43%).

The reasons for the group decline on the PCS-12 are unclear, and it is difficult to interpret why a large proportion of patients remained Unchanged on various outcome measures. In a multivariate study like this one, multiple outcome variables complicate the ability to draw inferences about reasons for unsuccessful treatment outcomes because patients may fall into multiple categories on multiple outcomes (e.g., Recovered on Well-Being, Improved on Depression, but Unchanged on Risk). Even if follow-up analyses were conducted to compare subgroups of treatment response categories (e.g., Recovered, Improved, Unchanged), results would remain at the group level. Thus, information about a particular patient's treatment response or the degree to which group results might reflect any particular patient's treatment response would remain unclear.

The therapist effect is another finding that is difficult to interpret. Patients treated by the male therapist were found to make significantly larger reductions on the PHQ-9 as compared with patients treated by the female therapist. The therapist effect was not present on any other outcome variables and it was not expected as an a priori hypothesis or controlled for in the study. Therefore, the reasons for this effect remain unclear. Future studies should investigate if this effect replicates and consider controlling for differences in therapist orientations, techniques, personalities, or other relevant variables which may be hypothesized to contribute to a difference in treatment outcomes.

Research Question #2 asked: What are the typical characteristics (i.e., demographics, diagnostic profile, and symptom severity) and treatment response of the average patient as described by group results on the PHQ-9, CORE-B, and SF-12?

The group results were summarized to understand the typical characteristics and treatment response of the average patient. The group results portrayed a Caucasian woman diagnosed with a moderately severe depressive disorder and a co-occurring psychological disorder or physical pain. The group results suggested this average patient would have minor physical health issues (PCS-12) and very poor mental health (MCS-12). After 11 sessions of TVCP, she would have experienced statistically significant reductions in depression (PHQ-9), symptom distress (CORE-B Symptoms subscale), and level of risk (CORE-B Risk subscale), as well as improvements in emotional well-being (CORE-B Well-Being subscale) and functioning (CORE-B Functioning subscale). After treatment, she would have felt less distressed (CORE-B Global Distress scale) and

perceived an improved mental health quality (MCS-12), but she would report a slight decline in perceived physical health quality (PCS-12).

Research Question #3 asked: What are the unique characteristics (i.e., demographics, diagnostic profile, and symptom severity) and treatment responses of four patients as described by single-case quantitative results on the PHQ-9, CORE-B, and SF-12?

The best way to obtain clinically useful data from group results was to take closer look at the individuals from the group and describe their unique characteristics and treatment responses. This step was accomplished with the narrative description of each patient presented in the single-case results. The four individual cases analyzed by single-case methods were demographically representative of the average patient with regards to ethnicity, gender, and severity of mental health issues. Three of the four patients improved on many outcome measures, however, one patient only improved on one mental health outcome. The implications of the single-case results for the four patients are discussed below.

Research Questions #4 and #5 asked: How are the characteristics and treatment response of the average patient similar and/or different than those of the individual patients analyzed by single-case quantitative methods? What clinical insights emerge by comparing and contrasting the characteristics and treatment response of four individual patients with each other and to the average patient with a mixed-methods design?

The comparison of (a) the group results to the single-case results of four patients and (b) the single-case results of four patients to each other was helpful for finding

clinical insights not evident in the group results. The first clinical insight which emerged suggested that the patients in this sample suffered from severe and complicated mental health issues. This clinical insight was derived from the similarities of the four patients to the group with regards to diagnostic profile, prevalence of comorbidity, and types of presenting problems. The poor mental health quality of rural patients has been consistently cited in the literature (Gamm et al., 2010; Smalley et al., 2010), and it presents treatment challenges for this population.

A second clinical insight emerged, suggesting that treatment response would likely result in partial improvements on some mental health outcomes instead of on all mental health outcomes. This clinical insight was derived from the slightly different picture of treatment response that emerged from the single-case results. The group results suggested large, across-the-board improvements in mental health outcomes for the average patient. However, this was not the norm for the four patients. Instead of each patient experiencing the average treatment response with large reliable change on all mental health outcomes, there were two patients who displayed large improvements on most mental health outcomes and two patients who experienced small gains on some outcomes and no gains on others. Of the two patients who made large gains over the course of treatment, one patient (#4) matched the group with regards to severe mental health (i.e., Bipolar Disorder, Substance Abuse, childhood sexual abuse, domestic abuse) while another patient (#2) experienced mild symptoms (i.e., Adjustment Disorder) not representative of the average patient in this sample. This clinical insight suggests confidence that (a) TVCP is likely effective for helping patients with mild to moderate

mental health symptoms on the majority of outcomes, and (b) TVCP will likely be effective in producing partial improvements on some mental health outcomes for patients with severe mental health issues.

In many ways, this conclusion is promising for clinicians working with similar populations who struggle to treat patients with severe and complicated mental health presentations because it suggests that even when a patient is not making large, consistent gains on all expected outcomes, there may be benefits experienced by the patient on some outcomes. This statement is especially promising when the partial improvements are on outcomes which make a meaningful impact. For example, Patient #1 continued to experience psychological distress and symptoms of anxiety, but she Improved on the CORE-B Functioning and Well-Being subscales and Recovered on the CORE-B Risk subscale and PHQ-9. Risk and depression are arguably the two most important markers for judging meaningful change, given the potential lethality resulting from the presence of either factor and the exponentially deleterious effects of their co-occurrence (Henriksson et al., 1993).

The context provided by single-case results about patient characteristics allowed for the observation of another clinical insight about treatment response. The magnitude and domain of outcomes on which patients responded were likely influenced by the unique characteristics of the patients' situational factors, type of diagnosis, and severity of symptoms. For example, Patient #1 continued to experience symptoms of anxiety that were likely the result of on-going domestic abuse, untreated PTSD, and the responsibility of caring for five children. Given her situation and diagnosis,

improvements on the PHQ-9, and CORE-B Risk, Well-Being, and Functioning subscales may have been the most realistic expectation of treatment response for Patient #1. As another illustration, Patient #3 made the fewest and smallest improvements of the four patients. At the same time, she reported severe physical health issues (i.e., multiple cardiac surgeries, diabetes, and physical pain) in addition to minimal social support and a history of trauma (sexual abuse). Her situational context and endorsement of severe physical health problems likely negatively influenced the therapeutic gains for Patient #3.

The negative influence of severe pain and physical health problems on treatment response was another clinical insight derived through comparison of the single-case and group results. The group and three of the four patients in the SCES exhibited declines on the PCS-12. However, unlike the statistically significant declines experienced by the group on the PCS-12, the single-case results suggested that only one of the four patients (Patient #3) was severely distressed and limited by physical health issues. This is an important finding because Patient #3 was considered the least responsive to TVCP. She had severe mental health issues in addition to severe physical health problems. The average patient described by group results had severe mental health issues, but only minor physical health issues. This suggested that severe physical health issues likely played a significant negative role in the therapeutic progress of Patient #3.

These types of clinical insights – which show how unique patient characteristics influence treatment response – help provide another perspective besides that offered by the group results. Comparing the single-case results of four patients to each other and to

those of the group allowed for a more differentiated explanation of treatment response. The clinical insights derived by considering both perspectives are considerably more useful to clinicians because they tell a story of actual patients rather than a myth of the average patient.

At the individual patient level, these kinds of clinical insights are invaluable to practitioners because evidence is concrete, reliable, and useful to clinicians wishing to understand their patient's specific progress or decline on important indicators. At the group level, evidence of treatment effectiveness is valuable because results provide probability estimates (i.e., inferential statistics) that average gains are likely to generalize to a particular type of sample. In essence, both group and single-case studies have their place in psychotherapy research, but seem to provide unique insights when used together.

This study demonstrated one approach to the mixed-methods design recommended by prominent scholars promoting a paradigm shift in the reporting of psychotherapy research (Barlow & Nock, 2009; Dattilio et al., 2010; Edwards et al., 2004; Fishman, 2011). Group analyses evaluated the probability that the sample would benefit from TVCP when measured on various psychological variables and single-case analyses examined the unique context and treatment responses of four patients in the group. This mixed-methods approach provided a contextualized view of change for actual patients undergoing treatment (from the single-case design) in juxtaposition to that of the statistical average patient described by group results. By comparing characteristics and treatment responses of four patients with each other and to the

average patient, clinical insights emerged and produced a multidimensional picture of change that neither the single case nor group designs could provide on their own. In this way, the mixed-methods design communicated results to scientists and practitioners in an effort to help bridge the long-standing divide in the field of psychotherapy research.

Design Considerations and Limitations

Although the group and single-case design did not meet all the standards for methodological soundness required to demonstrate strong causal inferences in this study (Kratochwill et al., 2010; Shernoff & Kratochwill, 2003), several important design features promoted a balance of internal and external validity. For the group design, there was moderate evidence of internal validity resulting from pre- and post-measurements of multiple outcome variables, implementation of valid and reliable of measures, comparison of outcomes to normative data (clinically significant change), and internal replicability (bootstrap) analysis of outcomes. Strong internal validity was not possible because the group design lacked an experimental control group, random assignment, inclusionary/exclusionary criteria (e.g., comorbidity and medication), manualized protocol administration, and treatment fidelity checks.

The single-case design met several fundamental WWC standards with the four cases selected for analysis. Specifically, the single-case design met the minimum number of baseline points per phase, it had the minimum number of phases, it used repeated measurement over the course of treatment, and it assessed multiple sources of outcome data. The single-case design also included intentional manipulation of the independent variable, a description of specific interventions used, reporting of the patient context

(i.e.., problems, diagnosis, gender, ethnicity), and assessment of mean level change using quantitative analysis. Strong internal validity was not possible because this design lacked a concurrent baseline control or withdrawal period, more baseline points needed to establish the degree of stability in the baseline period, monitoring of extraneous variables, and three demonstrations of a clear effect from visual analysis of the graphed outcome variable (CORE-B Global Distress). In addition, internal validity was limited because meaningful change was not assessed for each patient to ascertain if they met their treatment goals and improved on additional important behavioral markers of change (e.g., fewer panic attacks, less nightmares per week, increased social engagement, etc.). Based on standards and procedures for evaluating single-case (Kratochwill et al., 2010) and group designs (Shernoff & Kratochwill, 2003), the overall mixed-method design meets most standards for methodological soundness.

Several other limitations are acknowledged in this study. One important limitation is the relatively small sample size that was predominantly composed of Caucasian patients, most of which were women. A small, homogenous sample size will indubitably constrain the external validity of the study's results. This sample was also largely unemployed and approximately one quarter of patients reported receiving social security disability benefits. As such, findings should mainly be generalized to other similar rural, female Caucasian samples of lower socioeconomic status.

Another important consideration is the lack of treatment fidelity with regards to a consistent type of psychotherapy delivered via TVCP. Treatment fidelity can improve the internal validity of results and aid in the replication of future studies. Psychotherapy

researchers in the zeitgeist for ESTs would prescribe a manualized treatment administered to a client for a specific disorder with regular fidelity checks and controls (Chambless & Hollon, 1998).

However, as a result of cultural considerations and the preference for a field study which could address patients' complex clinical presentations, treatment fidelity was not a priority in this study. APA principles urge practitioners to provide culturally competent treatment which requires flexibility and alternative methods to meet the individual patient's needs (APA, 1993, 1995). This is especially important for rural patients who have a different set of needs than those of urban individuals and for which the acceptability of treatment is a real concern (HRSA, 2005). There is evidence to suggest that some minority groups respond poorly to ESTs (Lau, 2006), and that acceptability of an intervention can easily influence patient engagement and lead to drop out (Bernal et al., 2009). It is also important to recognize that a "treatment is more effective when compatible with client culture patterns" (Tharp, 1991, p. 802). As such, effectiveness studies in field settings – like this study – can make important contributions to the TVCP literature with rural populations despite a lack of treatment fidelity.

Another important limitation of this design is related to the experience level of psychotherapists. Doctoral level students conducted treatment-as-usual for this rural sample while under the supervision of licensed psychologists. Some research suggests that treatment response is directly mediated by therapist experience level (Stein & Lambert, 1995). Although therapist experience level is a recognized limitation of this

study, the collaboration of the BVHP with the TAMU doctoral counseling psychology program provided one solution to the larger problem of limited accessibility to mental health professionals faced by rural populations across the country (Gamm et al., 2010; Smalley et al., 2010). Thus, this limitation is considered a significant improvement over the national norm of no mental health care for rural populations (National Advisory Committee on Rural Health, 1993).

External validity for the overall study was strong for a variety of reasons. First, the mixed-methods design promotes dissemination of research to scientists and practitioners through the reporting of results that appeal to both parties. The single-case design produced quantitative results in an effect size commonly found in group and single-case research. The comparison of single-case and group results revealed several clinical insights which served to promote external validity of results by providing a more detailed, differentiated explanation of findings with clinical applications.

Another important factor which helps promote external validity was the reporting of patient, therapist, and treatment characteristics. Reporting study characteristics enhances the generalizability of findings to other similar populations. In this study, the homogeneity of the sample would suggest TVCP would be most effective for Caucasian woman in southern rural communities. The results also suggest that TVCP would be helpful for rural patients with a variety of severe and complicated mental health issues.

Implications

The results of this study suggest that TVCP offered through a primary care setting is an effective treatment modality that has the potential for reducing mental

health disparities in rural populations. The BVHP established medical and psychological healthcare for rural residents that otherwise had little or no access to treatment, and in doing so, provided a model for other communities across the nation to follow (McCord et al., 2011; Wendel et al., 2011). The impact of this study is particularly meaningful at a local community level because previous investigations have found high prevalence of depression rates in the Brazos Valley of Texas. Specifically, Brossart et al. (2013) found that depression rates were significantly higher for rural women than their urban counterparts, and that income level and depression scores were inversely related. The current study was overwhelmingly composed of Caucasian women of lower socioeconomic status that resided in the Brazos Valley. The large majority of this study's sample experienced very large reductions in depression symptoms and significant reductions in risk. On a local level, the impact of TVCP makes a tangible difference for this small, rural community by improving access and availability of mental health care.

Hopefully, this research also makes a broader impact on rural healthcare policy by emphasizing the effectiveness, clinical utility, and feasibility of integrating TVCP in rural primary care settings in order to reduce mental health disparities. The integration of behavioral telehealth services into primary care settings in rural areas is needed (Office of the Surgeon General, 1999; National Association of State Medicaid Directors, 2008; Sawyer, Gale, & Lambert, 2006; Smalley et al., 2010). Although there has been significant attention to rural issues and policy to improve rural mental health care access, there is a need to continue to improve the integration of mental health care in primary

care settings for rural residents (Bird, Lambert, Hartley, Beeson, & Coburn, 1998). Medicaid reimburses only about half of states in the U.S. for behavioral telehealth services (Kautz, Mauch, & Smith, 2008). In addition to the need for more states to provide Medicaid reimbursement to mental health providers for TVCP, there is a demand for greater parity in reimbursement for providers as well as a need for increased access to rural patients through primary care settings (Kautz et al., 2008). This study provides evidence that TVCP can be integrated in primary care settings and produce impactful results for rural communities.

This study also makes an important contribution to the currently scant literature regarding the effectiveness of TVCP to reduce mental health disparities for rural residents. Future investigations of TVCP and psychotherapy research in general may benefit by combining single-case experimental studies with group comparison research methods to improve dissemination in the field. By combining these methods, there is an increased capacity to promote a balance of internal and external validity. In addition, the presentation of results will likely be a more accurate and comprehensive evaluation of treatment response than can be found by either research method alone. Furthermore, this type of mixed-methods design enhances mutual communication of psychotherapy treatment results for consumption by scientists and practitioners alike, thus helping to bridge the long-standing gap between science and practice in the field.

REFERENCES

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy, 31,* 621-631. doi: 10.1016/0005-7967(93)90115-B

Allport, G.D. (1962). The general and the unique in psychological science. *Journal of Personality, 30,* 405–422. doi: 10.1111/j.1467-6494.1962.tb02313.x

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychological Association. (1993). Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations. *American Psychologist, 48,* 45–48. doi: 10.1037//0003-066x.48.1.45

American Psychological Association. (1995). *Caring for the rural community: An interdisciplinary curriculum*. Washington, DC: APA Office of Rural Health.

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57,* 1052–1059. doi: 10.1037/0003-066x.57.12.1052

American Psychological Association. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271-285. doi: 10.1037/0003-066X.61.4.271

Backhaus, A., Agha, Z., Maglione, M. L., Repp, A.,…Thorp, . S. R. (2012). Videoconferencing psychotherapy: A systematic review. *Psychological Services, 9,* 111-131. doi: 10.1037/a0027924

Barkham, M., Culverwell, A., Spindler, K., Twigg, E., & Connell, J. (2005). The CORE-

OM in an older adult population: Psychometric status, acceptability, and feasibility. *Aging and Mental Health, 9*, 235-245. doi: 10.1080/13607860500090052

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry, 186,* 239-246. doi: 10.1192/bjp.186.3.239

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,* 184–196. doi: 10.1037/0022-006x.69.2.184

Barlow, D. H. (1980). Behavior therapy: The next decade. *Behavior Therapy, 11*, 315-328. doi: 10.1016/s0005-7894(80)80049-9

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4,* 19-21. doi: 10.1111/j.1745-6924.2009.01088.x

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change (3rd ed.).* Boston, MA: Allyn & Bacon.

Bashur, R. (1997). Critical issues in telemedicine. *Telemedicine Journal, 3,* 113-126.

Bauer, S., Lambert, M. J., & Nielsen, S. L. (2004). Clinical significance methods: A

comparison of statistical techniques. *Journal of Personality Assessment, 82*, 60-70. doi: 10.1207/s15327752jpa8201_11

Benschoter, R. (1967). Multipurpose television. *Annals of the New York Academy of Sciences, 142,* 471-478. doi: 10.1111/j.1749-6632.1967.tb14360.x

Benschoter, R. A., Witson, C. L., & Ingham, C. G. (1965). Teaching and consultation by television. *Mental Hospital, 16,* 99-104.

Bergin, A. E., & Strupp, H. H. (1970). New directions in psychotherapy research. *Journal of Abnormal Psychology, 76,* 13-26. doi: 10.1037/h0029634

Bernal, G., Jimenez-Chafey, M. I., & Rodriguez, M. M. D. (2009). Cultural adaptation of treatments: A resource for considering culture in evidence-based practice. *Professional Psychology: Research and Practice, 40,* 361-368. doi: 10.1037/a0016401

Bird, D. C., Dempsey, P., & Hartley, D. (2001). *Addressing mental health workforce needs in underserved rural areas: Accomplishments and challenges*. Working Paper No. 23. Portland, ME: Main Rural Health Research Center. Retrieved from http://muskie.usm.maine.edu/Publications/rural/wp23.pdf

Bird, D.C., Lambert, D., Hartley, D., Beeson, P.G. & Coburn, A.F. (1998). Rural models for integrating primary care and mental health services. *Administration & Policy in Mental Health*, 25, 287-308. doi: 10.1023/a:1022291306283

Bouchard, S., Paquin, B., Payeur, R., Allard, M., Rivard, V., Fournier, T., . . . Lapierre, J. (2004). Delivering cognitive-behavior therapy for panic disorder with agoraphobia in videoconference. *Telemedicine Journal and e-Health, 10,* 13–25.

doi:10.1089/153056204773644535

Bouchard, S., Payeur, R., Rivard, V., Allard, M., Paquin, B., Renaud, P., & Goyer, L. (2000). Cognitive behavior therapy for panic disorder with agoraphobia in videoconference: Preliminary results. *Cyberpsychology and Behavior, 3,* 894-895. doi:10.1089/109493100452264

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical Practice as a natural laboratory for psychotherapy research: A guide to case-based time series analysis. *American Psychologist, 63*, 77-95. doi: 10.1037/0003-066X.63.2.77

Brecht, R. (1997). Correctional telemedicine. In G. Segre (Ed.), *Telemedicine sourcebook: 1998* (pp. 146-151). New York: Faulkner & Gray.

Breuer, J. & Freud, S. (1957). *Studies on hysteria*. New York, NY: Basic Books.

Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*, 531-563. doi: 10.1177/0145445503261167

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*, 252-263. doi: 10.1002/jclp.21949

Bushy, A. (1998). Health issues of women in rural environments: An overview. *Journal of the American Medical Women's Association, 53*, 53–56.

Busk, P. L., & Marascuilo, L. A. (1988). Autocorrelation in single-subject research: A

counter argument to the myth of no autocorrelation. *Behavioral Assessment, 10*, 229-242.

Cahill, J., Barkham, M., Stiles, W. B., Twigg, W., Rees, A., Hardy, G. E., & Evans, C. (2006). Convergent validity of the CORE measures with measures of depression for clients in cognitive therapy for depression. *Journal of Counseling Psychology, 53*, 253-259. doi: 10.1037/0022-0167.53.2.253

Center for Community Health Development. (2006). *Brazos Valley health status assessment: Executive report.* College Station, TX: School of Rural Public Health.

Cellucci, T., & Vik, P. (2001). Training for substance abuse treatment among psychologists in a rural state. *Professional Psychology: Research and Practice, 32,* 248–252. doi:10.1037/0735-7028.32.3.248

Chambless, D. L, Baker, M., Baucom, D. H., Beutler, L. E., Calhoun, K.. S.,…Woody, S.R. (1998). Update on empirically validated therapies, II. *Clinical Psychology, 51*, 3–16.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7-18. doi: 10.1037//0022-006X.66.1.7

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. V. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points, and comparison with the CIS-R. *British Journal of Psychiatry, 190*, 69–74. doi: 10.1192/bjp.bp.105.017657

Conrad, K. (1998). Making telehealth a viable component of our national health care system. *Professional Psychology: Research and Practice, 29*, 525-526. doi: 10.1037/h0092809

CORE System Group (1998). CORE System (Information Management) Handbook. Leeds, Core System Group. Retrieved at: http://www.wider-mind.com/research/core_system_user_manual.pdf

Cowain, T. (2001). Cognitive-behavioural therapy via videoconferencing to a rural area. *Australian and New Zealand Journal of Psychiatry, 35,* 62–64. doi: 10.1046/j.1440-1614.2001.00853.x

Cully, J. A., Jameson, J. P., Phillips, L. L., Kunik, M. E., & Fortney, J. C. (2010). Use of psychotherapy by rural and urban Veterans. *Journal of Rural Health, 26*, 225-233. doi: 10.1111/j.1748-0361.2010.00294.x

Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research, 15*, 1-17. doi: 10.2196/jmir.2227

Dattilio, F. M., Edwards, D. J. A., & Fishman, D. B. (2010). Case studies within a mixed

methods paradigm: Toward a resolution of the alienation between researcher and practitioner in psychotherapy research. *Psychotherapy: Theory, Research, Practice, Training, 47*, 427-441. doi: 10.1037/a0021181

DeLeon, P. H., Wakefield, M., Schultz, A. J., Williams, J., & VandenBos, G. R. (1989). Rural America: Unique opportunities for health care delivery and health services research. *American Psychology, 44*, 1298–1306. doi: 10.1037/0003-066x.44.10.1298

Des Jarlais, D.C., Lyles, C., Crepaz, N., & the TREND Group (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health, 94*, 361-366.

Dwyer, T. (1973). Telepsychiatry: Psychiatric consultation by interactive television. *American Journal of Psychiatry, 13*, 865-869.

Eberhardt, M. S., Ingram, D. D., & Makuc, D. M. (2001). *Urban and rural health chartbook: Health, United States, 2001*. Hyattsville, Maryland: National Center for Health Statistics. Retrieved at http://www.cdc.gov/nchs/data/hus/hus01cht.pdf

Eberhardt, M.S., & Pamuk, E.R. (2004). The importance of place of residence: Examining health in rural and nonrural areas. *American Journal of Public Health, 94*, 1682–1686. doi: 10.2105/AJPH.94.10.1682

Edwards, J. (1997). Military contributions to digital radiology and networking. In G. Segre (Ed.), *Telemedicine source book: 1998* (pp. 212-216). New York: Faulkner & Gray.

Edwards, D. J. A., Dattilio, F. M., & Bromley, D. B. (2004). Developing evidence-based practice: The role of case-based research. *Professional Psychology: Research and Practice, 3*5, 589-597. doi: 10.1037/0735-7028.35.6.589

Elliott, T. R., Brossart, D. F., Berry, J. W., & Fine, P. R. (2008). Problem-solving training via videoconferencing for family caregivers of persons with spinal cord injuries: A randomized controlled trial. *Behaviour Research and Therapy, 46,* 1220–1229. doi:10.1016/j.brat.2008.08.004

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Toward a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*, 51-60. doi: 10.1192/bjp.180.1.51

Evans, C., Connell, J., Barkham, M., Marshall, C., & Mellor-Clark, J. (2003). Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy*, *10*, 374-388. doi: 10.1002/cpp.384

Fischer, E. P., Owen, R. R., & Cuffel, B. J. (1996). Substance abuse, community service use, and symptom severity of urban and rural residents with schizophrenia. *Psychiatric Services, 47*, 980-984.

Fishman, D. B. (2011). The "Individual-Case-Comparison" method for systematically comparing good-outcome and poor-outcome RCT clients: Editor's introduction. *Pragmatic Case Studies in Psychotherapy, 7, 2*, 242-251.

Fishman, S. T. (1981). Narrowing the generalization gap in clinical research. *Behavioral*

Assessment, 3, 243-248.

Fortney, J. C., Booth, B. M., Blow, F. C., Bunn, J. Y., & Cook, C. A. (1995). The effects of travel barriers and age on the utilization of alcoholism treatment aftercare. *American Journal of Drug and Alcohol Abuse, 21*, 391–406. doi: 10.3109/00952999509002705

Fortney, J. C., Owen, R., & Clothier, J. (1999). Impact of travel distance on the disposition of patients presenting for emergency psychiatric care. *Journal of Behavioral Health Services and Research, 26*, 104-109. doi:10.1007/bf02287798

Gamm, L., Stone, S., & Pittman, S. (2010). Mental health and mental disorders – a rural challenge: A literature review. *Rural Healthy People*, 97-113. Retrieved at: http://srph.tamhsc.edu/centers/rhp2010/08Volume2mentalhealth.pdf

Germain, V., Marchand, A., Bouchard, S., Drouin, M., & Guay, S. (2009). Effectiveness of cognitive behavioural therapy administered by videoconference for posttraumatic stress disorder. *Cognitive Behaviour Therapy, 38*, 42–53. doi:10.1080/16506070802473494

Goldfried, M. R., & Wolfe, B. E. (1996). Psychotherapy practice and research: Repairing a strained alliance. *American Psychologist, 51*, 1007-1016. doi: 10.1037//0003-066x.51.10.1007

Goldfried, M. R., & Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology, 66*, 143-150. doi: 10.1037//0022-006X.66.1.143

Greenberg, L. S., & Newman, F. L. (1996). An approach to psychotherapy change

process research: Introduction to the special section. *Journal of Consulting and Clinical Psychology, 64,* 435-438. doi: 10.1037//0022-006X.64.3.435

Griffiths, L., Blignault, I., & Yellowlees, P. (2006). Telemedicine as a means of delivering cognitive behavior therapy to rural and remote mental health clients. *Journal of Telemedicine and Telecare, 12,* 136-140. doi: 10.1258/135763306776738567

Hart, L. G., Larson, E. H., & Lishner, D. M. (2005). Rural definitions for health policy and research. *American Journal of Public Health, 95,* 1149-1155. doi: 10.2105/ajph.2004.042432

Hays, R. D., Wells, K. B., Sherbourne, C. D., Rogers, W., & Spritzer, K. (1995). Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. *Archives of General Psychiatry, 52,* 11-19. doi: 10.1001/archpsyc.1995.03950130011002

Health Resources and Services Administration. (2005). *Mental health and rural America: 1994*-2005. Rockville, MD: Author.

Health Resources and Services Administration. (2008). *HPSA by state & county.* Retrieved from http://hpsafind.hrsa.gov/HPSASearch.aspx

Health Resources and Services Administration. (2013). *Defining the rural population.* Retrieved from http://www.hrsa.gov/ruralhealth/policy/definition_of_rural.html

Helbok, C.M. (2003). The practice of psychology in rural communities: Potential ethical dilemmas. *Ethics & Behavior, 13,* 367-384. doi: 10.1207/s15327019eb1304_5

Henriksson, M. M., Aro, H. M., Marttunen, M.J., Heikkinen, M.E., Isometsä, E.T.,

Kuoppasalmi, K.I., & Lönnkvist, J.K. (1993). Mental disorders and comorbidity in suicide. *American Journal of Psychiatry, 150*, 935-940.

Heppner, P. P., Wampold, B .E., & Kivlighan, D. M. (2008). *Research design in counseling* (3rd ed.). Belmont, CA: Thomson Brooks/Cole.

Hilliard, R. B. (1993). Single-case methodology in psychotherapy process and outcome research. *Journal of Consulting and Clinical Psychology, 61*, 373-380. doi: 10.1037/0022-006x.61.3.373

Himle, J. A., Fischer, D. J., Muroff, J. R., Van Etten, M. L., Lokers, L. M., Abelson, J. L., & Hanna, G. L. (2006). Videoconferencing-based cognitive behavioral therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy, 44,* 1821-1829. doi:10.1016/j.brat.2005.12.010

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165-179. doi: 10.1177/001440290507100203

Hoyt, D.R., Conger, R.D., Valde, J.G., & Weihs, K. (1997). Psychological distress and help seeking in rural America. *American Journal of Community Psychology, 25*, 449-470. doi: 10.1023/a:1024655521619

H.R. 5300, 103rd Cong., 2nd Sess. (1994).

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

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

Jerome, L. W., DeLeon, P. H., James, L. C., Folen, R., Earles, J., & Gedney, J. J. (2000). The coming of age of telecommunications in psychological research and practice. *American Psychologist, 55*, 407-421. doi: 10.1037//0003-066X.55.4.407

Jung, C. G. (1957). *The undiscovered self*. Princeton, NJ: Princeton University Press.

Kautz, C., Mauch, D., & Smith, S. A. (2008). *Reimbursement of mental health services in primary care settings* (HHS Pub. No. SMA-08-4324). Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49, 2*, 183-192. doi: 10.1037//0022-006x.49.2.183

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*, 110-136. doi: 10.1037/h0022911

Kratochwill, T. R., & Stoiber, K.C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*, 341-389. doi: 10.1521/scpq.17.4.341.20872

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf,

D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

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*, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Lowe, B. (2010). The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry, 32, 4*, 345-359. doi: 10.1016/j.genhosppsych.2010.03.006

Kusmin, L. (2013, November). Rural America at a glance. *Economic Information Bulletin, 24*, 1-6. Retrieved from: http://www.ers.usda.gov/publications/eb-economic-brief/eb24.aspx#.UzhsCYX-qcM

Kusmin, L. & Hertz, T. (2010, September). Rural America at a glance. *Economic Information Bulletin, 68*, 1-6. Retrieved from: http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib68.aspx#.Ur2l37SJLWg

Lambert, D. & Agger, M. (1995). Access of rural Medicaid beneficiaries to mental health services. *Health Care Financing Review*, 17, 133-145.

Larson, C. O. (2002). Use of the SF-12 instrument for measuring the health of homeless persons. *Health Services Research, 37*, 733-750. doi: 10.1111/1475-6773.00046

Larson, C. O., Schlundt, D. Patel, K., Beard, K., & Hargreaves, M. (2008). Validity of

the SF-12 for use in low-income African American community based research initiative (REACH 2010). *Preventing Chronic Disease: Public Health Research, Practice, and Policy, 5*, 1-14.

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology: Science and Practice, 13*, 295–310. doi: 10.1111/j.1468-2850.2006.00042.x

Lazarus, A. A. & Davison, G. C. (1971). Clinical innovation in research and practice. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis* (pp. 196-213). New York: Wiley.

Lenert, L. A., Sherbourne, C. D., Sugar, C., & Wells, K. B. (2000). Estimation of utilities for the effects of depression from the SF-12. *Medical Care, 38*, 763-770. doi: 10.1097/00005650-200007000-00008

Letvak, S. (2002). The importance of social support for rural mental health. *Issues in Mental Health Nursing, 23*, 249-261. doi: 10.1080/016128402753542992

Loberg, M. E. (2006). *The utility of behavioral telehealth for isolated populations* (Master's thesis, School of Professional Psychology). Retrieved from: http://commons.pacificu.edu/spp/88

Manchanda, M., & McLaren, P. (1998). Cognitive behavior therapy via interactive video. *Journal of Telemedicine and Telecare, 4,* 53–55. doi:10.1258/1357633981931452

Matyas, T. A. & Greenwood, K. M. (1990). Visual analysis for single-case time series:

Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23*, 341-351. doi: 10.1901/jaba.1990.23-341

Matyas, T. A., & Greenwood, K. M. (1996). Serial dependency in single-case time series. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 215-243). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Mellor-Clark, J., Connell, J., Barkham, M., & Cummins, P. (2001). Counselling outcomes in primary health care: A CORE system data profile. *European Journal of Psychotherapy, Counselling and Health*, 4, 65-86. doi: 10.1080/13642530110040118

Milet, K., Hudson, M., Baron, M., & Thombs, B. D. (2010). Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: Internal consistency reliability, convergent validity, and clinical correlates. *Rheumatology, 49*, 789-796. doi: 10.1093/rheumatology/kep443

Mitchell, A.J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *The Lancet, 374*, 609-619. doi: 10.1016/S0140-6736(09)60879-5

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

National Association of State Medicaid Directors. (2008). Serving the needs of Medicaid enrollees with integrated behavioral health services in safety net primary care settings. Washington, DC: Author.

Nelson, E., & Bui, T. (2010). Rural telepsychology services for children and adolescents. *Journal of Clinical Psychology: In Session, 66,* 490-501. doi: 10.1002/jclp.20682

Nickelson, D. W. (1998). Telehealth and the evolving healthcare system: Strategic opportunities for professional psychology. *Professional Psychology: Research and Practice, 29,* 527-535. doi: 10.1037//0735-7028.29.6.527

Nock, M. K., Michel, B. D., & Photos, V. I. (2007). Single-case research designs. In D. McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 337-350). Thousand Oaks, CA: Sage Publications.

Office of the Surgeon General. (1999). *Mental health: A report of the surgeon general.* Washington, DC: Department of Health and Human Services.

Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy, 34,* 189-211. doi: 10.1016/S0005-7894(03)80013-8

Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly, 21,* 418-443.

Parker, R. I., & Hagan-Burke, S. (2007). Useful effect size interpretations for single case research. *Behavior Therapy, 38*, 95-105. doi: 10.1016/j.beth.2006.05.002

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single case research: A review of nine nonoverlap techniques. *Behavioral Modification, 35, 4*, 303-322. doi: 10.1177/0145445511399147

Patient Health Questionnaire. (2012). Office coding algorithm. Retrieved from: http://phqscreeners.com/instructions/instructions.pdf

President's New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America*. Final report (DHHS Pub. No SMA-03-3832). Rockville: MD: U.S. Department of Health and Human Services.

Preston, J. (1993). *The telemedicine handbook.* Austin, TX: Telemedical Interactive Services, Inc.

Reese, R. J., Conoley, C. W., & Brossart, D. F. (2006). The attractiveness of telephone counseling: An empirical investigation of client perceptions. *Journal of Counseling and Development, 84*, 54-60. doi: 10.1002/j.1556-6678.2006.tb00379.x

Richardson, L. K., Frueh, B. C., Grubaugh, A., Egede, L., & Elhai, J. D. (2009). Current directions in videoconferencing tele-mental health research. *Clinical Psychology: Science and Practice, 16,* 323-338. doi: 10.1111/j.1468-2850.2009.01170.x

Robinson, O. C. (2011). The idiographic/nomothetic dichotomy: Tracing historical

origins of contemporary confusions. *History & Philosophy of Psychology, 13*, 32-39.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. Ed.). Newbury Park, CA: Sage.

Rost, K., Fortney, J., Fischer, E., & Smith, J. (2002). Use, quality, and outcomes of mental health: The rural perspective. *Medical Care Research and Review, 59*, 231-265. doi: 10.1177/1077558702059003001

Rost, K., Zhang, M., Fortney, J., Smith, J., & Smith Jr., G. (1998). Rural-urban differences in depression treatment and suicidality. *Medical Care, 36*, 1098-1107. doi: 10.1097/00005650-199807000-00015

Rueter, M. A., Holm, K. E., Burzette, R., Kim, K. J., & Conger, R. D. (2007). Mental health of rural young adults: Prevalence of psychiatric disorders, comorbidity, and service utilization. *Community Mental Health Journal, 43*, 229-249. doi: 10.1007/s10597-007-9082-y

S. 2375, 103rd Cong., 2nd Sess. (1994).

Sawyer, D., Gale, J., & Lambert, D. (2006). *Rural and frontier mental and behavioral health care: Barriers, effective policy strategies, best practices*. National Association for Rural Mental Health (NARMH). Retrieved at: http://www.narmh.org/publications/archives/rural_frontier.pdf

Schulz, K. F., Altman, D. G., Moher, & CONSORT Group (2010). CONSORT 2010

Statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine, 152,* 726-732. doi:10.7326/0003-4819-152-11-201006010-00232

Schopp, L., Demeris, G., & Glueckauf, R. L. (2006). Rural backwaters or front-runners? Rural telehealth in the vanguard of psychology practice. *Professional Psychology: Research and Practice, 37,* 165-173. doi: 10.1037/0735-7028.37.2.165

Sears, S. F., Evans, G. D., & Kuper, B. D. (2003). Rural social services systems as behavioral health delivery systems. In B. Stamm (Ed.), *Rural behavioral health care: An interdisciplinary guide* (pp. 109-120). Washington, DC: American Psychological Association Books. doi: 10.1037/10489-008

Shepherd, L., Goldstein, D., Whitford, H., Thewes, B., Brummell, V., & Hicks, M. (2006). The utility of videoconferencing to provide innovative delivery of psychological treatment for rural cancer patients: Results of a pilot study. *Journal of Pain and Symptom Management, 32,* 453- 461. doi:10.1016/j.jpainsymman.2006.05.018

Sherbourne, C. D., Wells, K. B., Duan, N., Miranda, J., Unutzer, J., Jaycox, L., Schoenbaum, M., Meredith, L. S., & Rubenstein, L. V. (2001). Long-term effectiveness of disseminating quality improvement for depression in primary care. *Archives of General Psychiatry, 58,* 696-703. doi: 10.1001/archpsyc.58.7.696

Shernoff, E. S., & Kratochwill, T.R. (2003). *Coding single-participant and group design*

studies: Coding workbook for evidence-based interventions. Unpublished manual. Retrieved at: http://www.indiana.edu/~ebi/EBI-Manual.pdf

Silverman, W. H. (1996). Cookbooks, manuals, and paint-by-numbers: Psychotherapy in the 90's. *Psychotherapy: Theory, Research, Practice, Training, 33*, 207-215. doi: 10.1037/0033-3204.33.2.207

Smalley, K. B., Yancey, C. T., Warren, J. C., Naufel, K., Ryan, R., & Pugh, J. L. (2010). Rural mental health and psychological treatment: A review for practitioners. *Journal of Clinical Psychology, 66*, 479-489. doi: 10.1002/jclp.20688

Stamm, B. H., Piland, N. F., Crouse, B., Boulger, J., Davis, G., Ide, B. A., B . . . Tidwell, K. (2003). Essays from the field. In B. Stamm (Ed.), *Rural behavioral health care: An interdisciplinary guide* (pp. 11–20). Washington, DC: American Psychological Association Books. doi: 10.1037/10489-000

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63*, 182-196. doi: 10.1037/0022-006X.63.2.182

Strupp, H. H. (1980a). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 1). *Archives of General Psychiatry, 37*, 595-603. doi: 10.1001/archpsyc.1980.01780180109014.

Strupp, H. H. (1980b). Success and failure in time-limited psychotherapy: A systematic comparison of two cases (Comparison 2). *Archives of General Psychiatry, 37*, 708-716. doi: 10.1001/archpsyc.1980.01780190106013.

Strupp, H. H. (1980c). Success and failure in time-limited psychotherapy: With special

reference to the performance of a lay counselor (Comparison 3). *Archives of General Psychiatry, 37,* 831-841. doi: 10.1001/archpsyc.1980.01780200109014

Strupp, H. H. (1980d). Success and failure in time-limited psychotherapy: Further evidence (Comparison 4). *Archives of General Psychiatry, 37,* 947-954. doi: 10.1001/archpsyc.1980.01780210105011.

Texas Health Institute. (2008). *Texas key health indicators.* Paper presented by Camille Miller at the Texas Indigent Health Care Association Annual Conference, Austin, TX.

Tharp, R. G. (1991). Cultural diversity and the treatment of children. *Journal of Consulting and Clinical Psychology, 59,* 799-812. doi: 10.1037/0022-006x.59.6.799

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: The Guilford Press.

Thurston-Hicks, A., Paine, S. & Hollifield, M. (1998). Rural psychiatry: Functional impairment associated with psychological distress and medical severity in rural primary care patients. *Psychiatric Services, 49,* 951-955.

Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behavior Therapy, 40,* 126-136. doi: 10.1080/16506073.2010.550059

Trust for America's Health. (2008). *Texas Public Health Data: Key Health Facts.* Retrieved from http://healthyamericans.org/states/?stateid_TX

Tuerk, P. W., Yoder, M., Ruggiero, K. J., Gros, D. F., & Acierno, R. (2010). A pilot study for prolonged exposure therapy for posttraumatic stress disorder delivered via telehealth technology. *Journal of Traumatic Stress, 23,* 116-123. doi: 10.1002/jts.20494

Wagenfeld, M. O., Murray, J. D., Mohatt, D. F., & DeBruyn. (1994). *Mental health and rural America: 1980-1993, An overview and annotated bibliography*. Washington, DC: Health and Human Services Administration, Office of Rural Health Policy, 1-116.

Wagenfeld, M. O. (2003). A snapshot of rural and frontier America. In B. Stamm (Ed.), *Rural behavioral health care: An interdisciplinary guide* (pp. 33-40). Washington, DC: American Psychological Association Books. doi: 10.1037/10489-002

Wagenfeld, M. O., & Buffum, W. E. (1983). Problems in, and prospects for, rural mental health services in the United States. *International Journal of Rural Health, 12*, 89-107.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 629-640. doi: 10.1001/archpsyc.62.6.629

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34,* 220-233. doi:10.1097/00005650- 199603000-00003

Ware, J. E., Kosinski, M., Turner-Bowker, D., & Gandek, B. (2002). *Version 2 of SF-12 Health Survey.* Lincoln, RI: Quality Metric.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30,* 473-483.

Wasem, C. & Puskin, D. (2000). High-tech with the human touch: Using telehealth to reach America's children. *Professional Psychology: Research and Practice, 31,* 3-4. doi: 10.1037/0735-7028.31.1.3

Wells, K. B., & Sherbourne, C. D. (1999). Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. *Archives of General Psychiatry, 56,* 897-904. doi: 10.1001/archpsyc.56.10.897

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

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). Empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130,* 631–663. doi: 10.1037/0033-2909.130.4.631

Zuithoff, N., Vergouwe, Y., King, M., Nazareth, I., van Wezep, M. J., Moons, K. G., &

Geerlings, M. I. (2010). The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: Consequences of current thresholds in a cross-sectional study. *BMC Family Practice, 11, 98*, 1-7. doi:10.1186/1471-2296-11-98

APPENDIX

NOMENCLATURE

ANOVA	Analysis of Variance
APA	American Psychological Association
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
BVHP	Brazos Valley Health Partnership
CBT	Cognitive Behavioral Therapy
CI	Confidence Intervals
CORE-B	Clinical Outcomes in Routine Evaluation – Short Form B
CORE-OM	Clinical Outcomes in Routine Evaluation – Outcome Measure
d	Cohen's d Effect Size
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders (4^{th} ed.)
DV	Dependent Variable
EBPP	Evidence-Based Practice in Psychology
ESTs	Empirically Supported Treatments
HRSA	Health Resource and Service Administration
IV	Independent Variable
M	Mean
M_{diff}	Mean Difference
MBD	Multiple Baseline Design
MCS	Mental Component Summary

MDD	Major Depressive Disorder
MHPSAs	Mental Health Professional Shortage Areas
OCD	Obsessive-Compulsive Disorder
OMB	Office of Management and Budget
ORHP	Office of Rural Health Policy
PCS	Physical Component Summary
PE	Prolonged Exposure
PHQ-9	Patient Health Questionnaire – 9
PTSD	Posttraumatic Stress Disorder
RCI	Reliable Change Index
RCTs	Randomized Controlled Trials
RMANOVA	Repeated Measures Analysis of Variance
SCES	Single-Case Experimental Studies
SCR	Single-Case Research
SD	Standard Deviation
SE	Standard Error
SF-12	Short Form General Health Survey – 12
SF-36	Short Form General Health Survey – 36
SSRIs	Selective Serotonin Reuptake Inhibitors
TAMU	Texas A&M University
TVCP	Telehealth Videoconferencing Psychotherapy
WWC	What Works Clearinghouse