2019

# An examination of the systematic error in three common outcome measures of cognitive behavioral therapy for chronic pain using generalizability theory

Justin M. Hughes

# AN EXAMINATION OF THE SYSTEMATIC ERROR IN THREE COMMON OUTCOME MEASURES OF COGNITIVE BEHAVIORAL THERAPY FOR CHRONIC PAIN USING GENERALIZABILITY THEORY

by

JUSTIN M. HUGHES

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts
in
The Department of Psychology
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA

2019

In presenting this thesis in partial fulfillment of the requirements for a master's degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Chair of the Department or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this thesis.

(student signature)

3/11/19

(date)

## THESIS APPROVAL FORM

Submitted by Justin M. Hughes in partial fulfillment of the requirements for the degree of Master of Arts in Psychology and accepted on behalf of the Faculty of the School of Graduate Studies by the thesis committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate on the work described in this thesis. We further certify that we have reviewed the thesis manuscript and approve it in partial fulfillment of the requirements for the degree of Master of Arts in Psychology.

3/11/19
Committee Chair
(Date)

3/11/19

3/12/19

Department Chair

3/12/19 College Dean

4/9/19 Graduate Dean

**ABSTRACT**
The School of Graduate Studies
The University of Alabama in Huntsville

Degree Master of Arts College/Dept Arts, Humanities, and Social Science/ Psychology

Name of Candidate Justin Hughes

Title An Examination of the Systematic Error in Three Common Outcome Measures of Cognitive Behavioral Therapy for Chronic Pain using Generalizability Theory

CBT for chronic pain is an approved therapy for chronic pain which has been shown to improve patient functioning. In order to adequately assess patient improvement, progress must be tracked across treatment. Examination of recent meta-analyses demonstrated currently used outcome measures can be categorized into the three outcome domains of Overall Pain Ratings, Physical Functioning, and Quality of Life. The most commonly used outcome measure within each domain was used to analyze the variability in improvement scores across a 12-week CBT for chronic pain using generalizability theory. Outcome measures included the ADL, NRS, and WHOQOL. The WHOQOL accounted for the greatest degree of variation in improvement scores. Results were compared to the degree of difficulty in use of each of the three outcome measures. The results indicated that the WHOQOL was much more clinically efficient than either the ADL or NRS for research purposes. Limitations and future research are discussed.

Abstract Approval: Committee Chair

Department Chair

Graduate Dean

## ACKNOWLEDGMENTS

The work described in this thesis would not have been possible without the assistance of a number of people who deserve special mention. I would like to thank Dr. Eric Seemann for his suggestion of the research topic and for his guidance throughout all the stages of this and other works throughout my undergraduate and graduate career. Next, I would like to thank Dr. Aurora Torres and Dr. Jeffrey Neuschatz for their support with comments and suggestions. I would also like to thank Dr. Dean Willis and Dr. Michael George of the Covenant Pain Therapies Center for allowing me access to their practice to conduct this research.

In addition, I would like to thank my family and friends who encouraged me to begin work on this degree. My wife, Laura, and my children, Aaron, Caleb, and Maggie, have been my inspiration and my source of strength and energy as I pursued this degree. My parents and grandparents have offered advice, encouragement, and financial support willingly and freely as I follow this ambitious dream of a higher education in psychology. I love and thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

# LIST OF ABBREVIATIONS

QoL ..... Quality of life (overall concept)

CBT ..... Cognitive Behavioral Therapy

APA ..... American Psychiatric Association

RCT ..... Randomized Controlled Trial

IMMPACT .. Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials

MPT ..... Multidisciplinary Pain Therapy

ACT ..... Acceptance and Commitment Therapy

TAU ..... Treatment as usual

MBSR ..... Mindfulness-Based Stress Reduction

MBCT ..... Mindfulness-Based Cognitive Therapy

RA ..... Rheumatoid Arthritis

PAIN ..... Overall Pain Ratings

PF ..... Physical Functioning

QOL ..... Quality of Life (outcome domain)

NRS ..... Numeric Rating Scale

ADL ..... Activities of Daily Living

WHOQOL ..... World Health Organization Quality of Life – BREF

ANOVA ..... Analysis of Variance

EVC ..... Expected Variance Components

LIST OF SYMBOLS

$M$ ..... Mean (arithmetic average)

$SD$ ..... Standard Deviation

H ..... Hours

$X^2$ ..... Chi Squared Goodness of Fit

$Mo$ ..... Mode (middle value in a range)

Cronbach's α ..... Estimate of Reliability

$E\sigma_p^2$ ..... Estimated variance attributable to person

$MS_p$ ..... Mean Square for person

$\sigma_e^2$ ..... Error variance

$n_s$ ..... Number of scales

$F$ ..... ANOVA value

$MSE$ ..... Mean square of the error

$P$ ..... Significance value

$\eta^2_p$ ..... ANOVA effect size

$EVC$ ..... Estimated Variance Component

$T$ ..... T-test value

$d$ ..... Cohen's d (t-test effect size)

# CHAPTER I

# INTRODUCTION

## A. Chronic Pain

Chronic pain can be defined as a subjective experience which lasts longer than six months with little or no relief from medical intervention and that has a debilitating effect on the quality of life (QoL) and activities of daily living of the affected individual (Ehde, Dillworth & Turner, 2014; Jonsdottir, Aspelund, Jonsdottir, & Gunnarsdottir, 2014; de Figueiredo & Griffith, 2016). While the subjective experience of chronic pain may differ from person to person, the responses to these experiences are often similar. Per Ojala et al. (2015), the typical long-term psychological response to chronic pain often includes distress, anxiety, fear, sorrow, despair, emotional lability, depression, exhaustion, and uncertainty over the future, as well as life changes such as loss of friends or loss of the ability to work, the formation of a new 'pain identity', and the addition of pain to the definition of normal life. These responses may often be more distressing than the chronic pain itself.

## B. Cognitive Behavioral Therapy

The psychological symptoms engendered by chronic pain and the accompanying distress can be addressed through psychological intervention. One APA-approved non-

medical intervention for chronic pain is Cognitive Behavioral Therapy (CBT). CBT works on the premise that faulty cognitive processes cause distorted interpretations of reality, thereby compromising the biopsychosocial health of the patient (Thorn, 2004; Castro, Daltro, Kraychete, & Lopes, 2012; Ehde et al., 2014). CBT is designed to remedy those faulty cognitive processes by helping patients identify maladaptive thoughts, attitudes, beliefs, and behaviors that can skew the experience of pain (Castro et al.). Ojala et al. (2015) suggest that pain education, a major component of CBT for the chronic pain patient, is essential to correct maladaptive thoughts or beliefs about chronic pain. CBT has thus been shown to be an effective psychological therapy for chronic pain.

The efficacy of CBT for alleviating the experience of chronic pain and pain interference has been well established in existing literature. Psychological intervention for chronic pain is based on Neuromatrix Theory (Knoerl, Smith, & Weisberg, 2016; de Figueiredo & Griffith, 2016) which suggests that neural pathways are modified by chronic pain and result in cognitive processes that generate dysfunctional thoughts and behaviors such as pain catastrophizing cognitions to ultimately compromise the biopsychosocial health of the chronic pain patient (Castro et al., 2012). CBT has been the subject of many randomized controlled trials (RCT) and has been found to be an effective therapy for alleviating the experience of chronic pain, both at the time of intervention and at 6-month follow-up (Knoerl et al.). Further, a recent study found that CBT was effective at increasing QoL of chronic pain patients over the course of a 12-session CBT (Hughes, Seemann, George, & Willis, 2018). Another study has also shown the efficacy of CBT at reducing not only overall pain, but also pain-related anxiety and depression

(Ehde et al., 2014). Several studies have also demonstrated efficacy of CBT to improve participant function beginning with the first session, and have established that changes in participant function in response to CBT occur in a predictable manner (Hughes et al.; Kleinstauber, Lambert, & Hiller, 2017; Sachser, Keller, & Goldbeck, 2017; Zinzow et al., 2017; Ehde et al.; Farrer, Griffiths, Christensen, Mackinnon, & Batterham, 2014; Watts et al., 2013; Lynch, Berry, & Sirey, 2011). Predictably, CBT has become the gold standard psychological intervention for chronic pain.

C. Patient Progress

For any treatment to be considered successful, patient progress must be adequately measured. Progress in chronic pain treatment is measured using one or more of several established outcome measures. A report by Dworkin et al. (2008) outlined four outcome domains recommended by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) which include pain intensity, physical functioning, emotional functioning, and participant ratings of overall improvement. The IMMPACT report recommended employing at least two of the four outcome measures for any given study and that the outcome measures used in any given study should be the measures most appropriate to the study. IMMPACT also provided that other outcome measures may be used when appropriate.

D. Meta-Analyses

In light of these recommendations, I conducted a review of recent meta-analyses to examine the utilization of these and any other outcome domains to ascertain the

relevance of these recommendations specifically to research concerning psychological intervention for chronic pain. The decision to examine meta-analyses rather than a direct examination of individual studies was made based on the nature of meta-analyses, in that meta-analyses directly assess data reported in available studies to consequently provide an adequate sample of the available literature. Meta-analyses selected for review examined outcomes of clinical research concerning improvement in chronic pain involving CBT or similar therapies regardless of medical treatment in adult participants published after the IMMPACT report. The results of the search revealed four meta-analyses which met those criteria. One of these meta-analyses, Gotink et al. (2015), examined previous meta-analyses and systematic reviews of RCTs of secular mindfulness-based therapies concerning patients with pain of any duration or intensity. This article was excluded as a result of the degree of separation between the final meta-analysis and the initial source material which were almost exclusively published prior to the IMMPACT report. The other three meta-analyses and their relevance to this research are detailed in the following paragraphs, as well as the most highly cited article in PsychInfo and MedLine from each meta-analysis as an example of the articles included in each meta-analysis.

The first meta-analysis reviewed was created to address commonly used outcome measures in multidisciplinary pain therapy (MPT). Deckert et al. (2016) conducted a meta-analysis to examine the state of the literature in multidisciplinary pain management for the purposes of the development of a standardized set of core outcome domains and measures for research in MPT. The purpose of this meta-analysis was very similar to that of the current study, with the exception that Deckert et al. focused on MPT whereas the

current study focuses on psychological aspects of chronic pain treatment only. The authors cited the IMMPACT report as a foundational study for the need for core outcome measures, but cited several difficulties specific to MPT in using those outcomes recommended by IMMPACT. Eligibility criteria for inclusion by Deckert et al. (2016) were established prior to a search of the literature and included MPT studied through RCTs and longitudinal non-randomized studies. For studies which met those criteria, final inclusion criteria required chronic pain be present for a minimum of 3 months, that the intervention be performed by at least physical therapists and psychotherapists/psychologist, and that the profession be stated by the authors. Also, outcome measures had to be clearly stated, described, and assessed both at baseline and follow-up in a minimum of 10% of all studies. The authors conducted electronic searches in MedLine, Embase, and AMED for articles published through August 2013. Search strings included terms for chronic pain, MPT, and RCT. Study characteristics and reported outcome domains were extracted in duplicate by all reviewers. Of the 626 full texts which met initial inclusion criteria, 556 were excluded primarily because pain duration was not clearly defined in the source material. A total of 70 reports met all predefined inclusion criteria. Included studies consisted of 20 RCTs and 50 longitudinal non-randomized studies published between 1985 and 2013. Most of these studies were conducted in Western Europe and the USA. Individual studies reported between 1 and 34 different outcomes, resulting in 145 different outcomes.

As a result of this meta-analysis, Deckert and colleagues (2016) identified 12 individual outcome categories based on the frequency with which their constituent outcome measures appeared in the literature and grouped them into Physical Health,

Social Health, and Mental Health domains. Although no specific outcome measure was used by all studies, most of the studies assessed a combination of the three domains. The Physical Health domain was the only domain assessed in each study. Pain intensity, depressive symptoms, and physical functioning were most commonly reported, and QoL and sickness impact were considered superior outcome categories because these constructs combined the three domains. The Physical Health category consisted of pain intensity, pain site, disability, and physical functioning. The most reported outcome in this area was pain intensity and site, used in 61 of 70 studies. Fear, depressive symptoms, psychological distress, coping, self-efficacy, and catastrophizing were the most reported outcome categories regarding the Mental Health domain, reported in 28 of 70 studies. The majority of studies examined depressive symptoms, followed by fear in general, fear of pain, and avoidance of movement. The Social Health domain, which examined work ability, sick leave, and work status, was evaluated less frequently in the included studies.

As an example of the articles included in Deckert et al. (2016), McCracken and Gutierrez-Martinez (2011) conducted a study to investigate a range of treatment processes in Acceptance and Commitment Therapy (ACT) for chronic pain that is more comprehensive in comparison to those investigated in previous studies, such as general psychological acceptance and mindfulness. Participants were patients at a tertiary care pain clinic in southwest England between September 2006 and June 2009. All participants reported persistent pain of 3 months duration or longer and significant levels of pain-related distress and disability. This study included 168 individuals (112 women, 56 men) between the ages of 18 and 77 years ($M = 43.5$, $SD = 13.0$) who completed a

three-or-four-week course of multimodal therapy for chronic pain, as well as the three-month follow-up session.

Participants completed a series of assessment instruments before and after treatment and at the 3-month follow-up. The instruments administered in McCracken and Gutierrez-Martinez (2011) included measures of acceptance of chronic pain, general psychological acceptance, mindfulness, values-based action, depression, pain-related anxiety, sickness impact, medical visits, and pain ratings. Participants received a treatment program that was a form of ACT specifically designed for group therapy in a specialty care setting, and within a multimodal therapy program. Specifically, methods focused on enhancing acceptance of pain and other psychological experiences, contact with the present moment, self-as-observer, cognitive diffusion, values, and committed action.

Treatment was delivered 5 days per week for 6.5 h each day. Each treatment day included approximately 2.25 h of physical conditioning, one 1.5 h of mindfulness training, and 1 h of activity management, with the remainder of the time devoted to other aspects of skills training and health education. All the methods used by the McCracken and Gutierrez-Martinez (2011) were designed not to target pain or other symptoms for removal, but instead to alter patient experience of these symptoms to reduce impact and improve functioning. Immediately following treatment and at 3-month follow-up, participants reported significantly lower levels of depression, pain-related anxiety, physical and psychosocial disability, medical visits and pain intensity in comparison to the start of treatment. Almost all effect sizes relative to treatment onset remained at a

medium or large level at the 3-month follow-up, with the exception of pain intensity and number of medical visits which were of a small size.

Pike, Hearn, and Williams (2016) sought to update the findings of a previous meta-analysis (Williams, Eccleston, & Morley, 2012) for the purpose of examining the current evidence for the effectiveness of psychological treatment for chronic pain excluding headache in terms of physical functioning. The authors cited the previous meta-analysis which compared CBT for chronic pain, excluding headache, with treatment as usual, waitlist, and medical treatment for their primary search method and inclusion criteria. The previous meta-analysis found moderate effect sizes for CBT when compared to the aforementioned control groups, but found that outcome domains and measures were varied and disparate across their reviewed studies. Pike et al. sought to expand on the previous meta-analysis by broadening the psychological treatment from CBT to any validated psychological treatment and by examining the efficacy of available outcome measures of improvements related to healthcare use and work status.

The authors conducted a search similar to Willaims et al. (2012) to include all relevant articles from the original meta-analysis and extended the search to include articles published through January 2015. Search terms included pain and "22 relevant phrases available by contacting the author (Williams et al.)." Articles were retrieved from Medline, Embase, PsychInfo, and the Cochrane Central Register of Controlled Trials. The authors also searched the reference lists of retrieved articles. Inclusion criteria for the Pike et al. (2016) meta-analysis were full publication of an RCT of adult participants reporting non-headache chronic pain of a minimum of 3 consecutive months in a peer-reviewed journal, a design based on an existing psychological model with at

least one trial group consisting of a psychological intervention delivered by a qualified professional compared to a control group. Other requirements included a minimum of 10 participants in each group at the end of treatment, and health care use, medication use, work absence post-treatment, or a combination of the three as outcome measures.

The search resulted in 1,915 potential articles while another 65 articles were identified from the previous meta-analysis and reference list search, totaling 1,980 articles screened for inclusion in Pike et al. (2016). Of those, 1,956 were excluded on abstract review, primarily resulting from absence of an appropriate outcome measure. After full-article review of the remaining 24 articles, 6 were excluded. Two were excluded because the outcomes were reported in a way that precluded reanalysis, two were insufficiently psychological, one did not meet the minimum of 3 months of chronic pain, and one was an internet study. Of the remaining studies, 13 RCTs were included from the Williams et al. (2012) meta-analysis. The updated search found 4 additional reports, and one report was found from reference lists of other studies, totaling 18 trials, 4 of which yielded no usable data. The 14 included articles comprised 2,253 participants (74% female) at the start of treatment and 1,932 at the end. Most participants were between 35 and 60 years of age, with a mean of 46 years. Mean duration of pain was approximately 4 years with the bell-curve skewed to the right, including a report of 50-year pain duration. The majority of studies tested cognitive and behavioral treatments such as ACT or MBSR within MPT. The overall effect showed moderate superiority of intervention over control for health care use, but no significant effect for medication use or work loss. According to Pike et al. (2016), from the viewpoint of improving intervention and long-term well-being, the authors concluded both reducing health care

use and improving return to work to be worthwhile outcomes. From a research viewpoint, the results demonstrated that health care use and return to work may not be effective outcome measures.

As an example of the articles included in Pike et al. (2016), McCracken, Sato, and Taylor (2013) conducted a study to establish the efficacy of ACT for chronic pain in a primary care setting. Primary outcome and treatment process variables included disability, depression, physical functioning, pain, acceptance, emotional functioning, patient global impression of change, and a question about changes in medication. Health care utilization was also tracked through patient report. This study included a randomized trial of group treatment for people with chronic pain recruited from general practice. The treatment included a combination of methods to promote psychological flexibility using experience-based methods, and de-emphasized lecturing and information-giving. After baseline assessment, participants were randomized to ACT plus standard treatment or treatment-as-usual (TAU) alone using random assignment. Inclusion in McCracken et al. (2013) required persistent pain rated greater than 4 out of 10 with longer than 3 months' duration, a pain-related medical visit in the past 6 months, significant pain-related distress and disability, consistent analgesic medications use, ability to communicate in English, and age 18 years or older. The participants ($n = 73$) ranged from 23 to 86 years old ($M = 58.0$, $SD = 12.8$ years), and 27.6% were 65 or older. Most were women (68.5%) and white British (97.3%). Comparisons using independent t-tests showed that the two experimental conditions did not differ significantly in any assessed demographic.

Immediately post treatment, McCracken et al. (2013) found significant group differences in favor of ACT for depression with a moderate effect size, but no significant changes for disability, physical functioning, or pain. There was also a significant post-treatment effect on patient global impression of change, with 53.3% reporting overall improvement compared to 25.0% in the TAU group. There was no significant group difference on emotional functioning. At the 3-month follow-up, no significant group differences were found for physical functioning or pain. The effects at follow-up for disability and depression were medium. There remained roughly twice as many participants who rated themselves as improved at follow-up in the ACT condition compared to the TAU condition, but this effect did not reach statistical significance. Once again, as at post-treatment, there were no significant group differences in the emotional functioning or in pain medication changes.

Finally, Veehof, Trompetter, Bohlmeijer, and Schreurs (2016) conducted a meta-analysis as a follow-up to a previous meta-analysis to compare the effectiveness of CBT to that of several other psychological treatments for chronic pain. In the former meta-analysis (Veehof, Oskam, Schreurs, & Bohlmeijer, 2011), CBT was found to be the most effective of those treatments investigated, with all other treatments showing only small effect sizes. The latter meta-analysis was conducted to include more recent systematic reviews of ACT, Mindfulness-Based Stress Reduction (MBSR), and Mindfulness-Based Cognitive Therapy (MBCT) as a result of the considerable expansion of the representation of those treatment modalities in the available literature.

Veehof and colleagues (2016) included studies from the previous meta-analysis and conducted a search of articles published prior to December 2013 in PubMed,

EMBASE, PsycInfo, and the Cochrane Central Register of Controlled Trials. All databases were searched for English language studies using an extensive list of previously defined search terms. The reference lists of newly included studies were also examined for additional eligible studies. Initial inclusion criteria required studies include ACT, MBSR, or MBCT for chronic pain, that pain be present for a minimum of 6 months, and that studies be executed using a randomized controlled design. The search initially yielded 1,393 titles. After removal of duplicates, two reviewers independently selected potentially eligible studies on the basis of title and abstract. A total of 163 studies were identified as being potentially eligible for inclusion in the study. The final selection was then made by two reviewers. Studies were excluded if the target treatment modality was just one of several modalities provided simultaneously to the treatment group, the intervention consisted of a single treatment session, insufficient data were reported, or absence of common outcome measures, including pain intensity, depression, anxiety, disability, pain interference, or QoL. Further, as the purpose of the meta-analysis was to compare psychological treatment modalities, studies which included a control group other than waitlist, TAU, or education were removed.

Including the eight randomized controlled studies from in the previous meta-analysis, Veehof and colleagues (2016) analyzed a total 30 studies comparing ACT, MBSR, and MBCT with waitlist, TAU, or education only. These 30 studies evaluated a total of 1,285 subjects. The participants were adults with a mean age between 35 and 60 years, and the majority were women. Study size ranged from a small pilot study ($n$ = 14) to a large-scale study ($n$ = 112), with an average of 51 participants. The programs used included a MBSR-based program ($n$ = 11), an ACT-based program ($n$ = 9), a

combination of MBCT and MBSR ($n$ = 1), and a MBCT-based program ($n$ = 1). Most programs consisted of 8 weekly group sessions, each session ranging in length from 1.5 to 2.5 h. Comparison groups included waitlist ($n$ = 10), education/support group ($n$ = 8), and TAU ($n$ = 7) as the control group. Effects at post-treatment on were assessed for pain intensity ($n$ = 22), depression ($n$ = 16), anxiety ($n$ = 6), pain interference ($n$ = 4), disability ($n$ = 10), and QoL ($n$ = 11). Results of the data synthesis revealed a moderate effect size for all psychological modalities and similar effect sizes to CBT. Thus, the authors concluded that acceptance- and mindfulness-based interventions for chronic pain were moderately effective on a number of beneficial outcomes, and comparable in effect to CBT.

As an example of the articles included in Veehof (2016), Zatura et al. (2008) examined whether individuals with rheumatoid arthritis (RA) exhibit greater resilience when given the opportunity to learn new responses to their chronic pain and functional limitations. The authors examined the value of two distinct approaches to psychological treatment of RA, one based on established cognitive behavioral methods emphasizing pain management and the other based on mindfulness meditation and positive social engagement to target emotion regulation. Both active treatments were compared with an established arthritis education curriculum to determine if either produced greater benefits. A total of 144 patients (68.1% women, 31.9% men) participated in the intervention trials. The study used a two-factor mixed design consisting of a treatment group with three levels (MBCT, CBT for pain, and arthritis education) and a two-level variable for recurrent depression, both assessed pre- and post-treatment. Outcome variables included an assessment of depression, daily pain levels, positive and negative affect, efficacy of

pain coping, efficacy of pain control, medical examination, and stimulated interleukin-6 production assays.

The results of Zatura et al. (2008) showed that both cognitive behavioral and affective interventions were useful, but in different ways, and may depend on participant history of depression. Both measures of pain coping efficacy and catastrophizing indicated a consistent pattern. Patients with recurrent depression in the mindfulness group showed a greater shift across treatment in their efficacy expectations for coping successfully with pain and decreasing catastrophization compared to the other groups. Results also revealed that the CBT group yielded better cognitive control and the mindfulness group demonstrated better emotion regulation. Thus, it was concluded that mindfulness-based cognitive therapy may yield similar results to CBT for chronic pain through different but similar methods.

E. Outcome Domains

Reports in these meta-analyses employed a total of 409 individual outcome measures. When redundant outcome measures were removed, the remaining measures revealed 18 distinct outcome categories. These categories were then condensed into three overarching outcome domains. The Overall Pain Ratings (PAIN) domain consisted of only one outcome category, pain rating, cited 85 times in the three meta-analyses. Only one meta-analysis (Veehof et al., 2016) listed the specific instruments used to assess outcomes in this or any domain. Of the 30 articles cited by Veehof et al., 28 cited pain rating as an outcome category as assessed by 7 different instruments. Of those 28 articles, 20 relied either on a direct verbal *0 – 10* rating of current pain or a response to

one pain-related question on one of several questionnaires rating pain on a *0 – 10* Likert-type scale. The other 8 articles involved qualitative descriptions of pain and often included a *0 – 10* pain rating as well. When assessed empirically, one of the seven instruments, the Numeric Rating Scale, was observed to have been used more often than the others, $X^2$ (6, $n$ = 28) = 18.29, $p$ = .01. Individual outcome domains and their categories can be found in Table 1.1.

Table 1.1 Outcome category representation within each domain

| PF | *n* | Pain | *n* | QOL | *n* |
|---|---|---|---|---|---|
| Disability | 40 | Pain Rating | 85 | Depression | 63 |
| Physical Functioning | 35 | | | Fear | 28 |
| Sickness Impact | 14 | Total | 85 | QoL | 26 |
| Pain Interference | 7 | | | Catastrophization | 10 |
| Pain Location | 5 | | | Coping | 10 |
| | | | | Psych. Distress | 10 |
| Total | 101 | | | Self-Efficacy | 9 |
| | | | | Anxiety | 7 |
| | | | | Total | 163 |

The Physical Functioning (PF) domain consisted of 8 distinct outcome categories, including a direct examination of physical functioning by a medical doctor or physical therapist, self-reports of disability, work status, sickness impact, pain interference, pain

location, medication use, and physician consultation. With respect to the results of Pike et al. (2016), outcome measures regarding work status, healthcare use, and medication use were disregarded to result in 5 distinct outcome categories. Across the three meta-analyses, disability was cited most often, with 40 citations, followed by physical function (35 citations), sickness impact (14 citations), pain interference (7 citations), and pain location (5 citations), yielding 101 citations. Of the 30 articles cited by Veehof et al. (2016), 20 cited either the disability or pain interference category as assessed by a total of 9 different instruments. All 20 articles assessed these outcome categories using either questionnaires that assessed their respective outcome category with one or two questions as part of a larger construct, by direct examination by a physician or physical therapist, or by ad hoc questionnaires. When assessed empirically, the disability and physical function categories were observed to have been reported more often than the other categories, $X^2$ (4, $n$ = 101) = 52.22, $p < .01$, but neither of the instruments reported in Veehof et al. within any category were observed to have been used more often than the others, $X^2$ (8, $n$ = 20) = 14.21, $p$ = .07 (see Table 1.1).

The Quality of Life domain (QOL) consisted of 8 distinct outcome categories that assess a broad range of variables related to QoL that could not be classified into either of the other two domains, but were not assessed often enough to be classified into its own domain. Those categories include assessments of either health-related or general QoL (26 citations), mental health variables, such as depression (63 citations), fear of pain or movement (28 citations), catastrophization (10 citations), self-efficacy (9 citations), and anxiety (7 citations), or the general impact of pain on daily life, such as psychological distress (10 citations) and coping (10 citations), for a total of 163 citations.

Categories in this domain were cited a total of 39 times across the 30 articles cited by Veehof et al. (2016). All instruments used to assess these categories were questionnaires that assess those specific symptoms. Those categories were depression (21 citations), QoL (11 citations), and anxiety (7 citations). When assessed empirically, the depression category was observed to have been reported significantly more often than the other categories, $X^2$ (7, $n$ = 163) = 124.53, $p < .01$, but neither of the instruments reported in Veehof et al. were demonstrated to have been reported more often than the other instruments in either category of depression, $X^2$ (8, $n$ = 21) = 12.02, $p$ = .15, anxiety, $X^2$ (2, $n$ = 7) = 2.03, $p$ = .37, or QoL, $X^2$ (5, $n$ = 11) = 7.01, $p$ = .22 (see Table 1.1).

F. Outcome Measures

When these domains were compared to recommendations from the IMMPACT report (Dworkin et al., 2008), several areas of overlap were seen. The PAIN category was analogous to the IMMPACT recommended Pain Intensity category. Pain rating was defined as patient-subjective rating of their pain at that moment. This was typically assessed in the meta-analyses using the IMMPACT-recommended Numeric Rating Scale (NRS), a simple verbal rating of pain on a scale ranging from *0 - 10*.

The PF domain was specifically recommended by IMMPACT. Within the PF domain, the disability and physical functioning categories were reported most often in the meta-analyses, and were thus assumed to represent the outcome categories most often used to assess outcomes across psychological treatment for chronic pain. The disability category was almost exclusively assessed using direct questions or ad hoc questionnaires, and thus do not lend themselves to analysis for research purposes. The only standardized

assessment found within the depression and physical functioning categories was a direct examination by a medical doctor or physical therapist using a prescribed rating scale, such as the Activities of Daily Living scale (ADL; Cress, Petrella, Moore, & Schenkman, 2005). The ADL is a standardized rating of overall physical function as assessed by a physical therapist based on the participant's ability to complete certain activities for a specific length of time. The ADL is only one of several instruments available for the direct examination of physical function by a medical professional. Although no one instrument has been accepted as the industry standard, most of these instruments display adequate reliability and validity for research purposes.

The QOL outcome domain expands on the Emotional Functioning IMMPACT recommendation by including other categories of non-physical function described above. The QOL outcome domain was recommended by Decker et al. (2016) as a superior outcome domain, as it takes into account categories from the other two domains, thus ostensibly setting apart QOL as a more preferred outcome domain in that it may deliver adequate data for the other two outcome domains without directly examining those domains. QoL can be affected by chronic pain as a result of mobility limitations (Stubbs, Schofield, & Patchay, 2016), depression and anxiety (Inoue et al., 2016), pain frequency and intensity (Jonsdottir et al., 2014), demoralization, disruption of emotion and thought patterns through attentional demand from chronic pain (de Figueiredo & Griffith, 2016), pain catastrophizing, and a desire to escape from pain, often including suicidal ideation (Trinanes, González-Villar, Gómez-Perretta, & Carrillo-de-la-Peña, 2016). Although depression was observed to have been reported most often in this domain, no instrument exists to specifically assess depression along with the other

categories within this domain; however, several instruments exist to examine QoL that includes an assessment of depressive symptoms, as well as other categories within this domain. One such instrument which has been assessed for validity and reliability within chronic pain populations, as well as other medical populations and healthy controls is the World Health Organization Quality of Life – BREF (WHOQOL) (Shawver et al., 2016; Skevington, Lofty, & O'Connell, 2004). The WHOQOL is a 26-item questionnaire that assesses 6 of the 8 categories within this domain and all 3 categories reported by Veehof et al. (2016), as well as several categories from the other two domains.

G. Clinical Efficiency

When considering these instruments and their application, it can easily be observed that vast differences exist in the administration of these instruments and in the utility of the data gathered by these instruments for research purposes. Regarding administration, data from the NRS can be acquired by simply asking the participant one question (Dworkin et al., 2008), and data from the WHOQOL is obtained by having the participant answer 26 questions on a *1 – 5* Likert-type scale which must be calculated and converted into a standardized score (Shawver et al., 2016; Skevington et al., 2004). Both of these assessments can be completed during a regular office visit to a physician or psychologist, typically in less than 5 minutes. In contrast, the ADL requires participants to make an appointment with a physical therapist, thus incurring an added investment of time and money, and then complete a series of activities that may or may not be strenuous for the participant (Cress et al., 2005).

Regarding the usefulness of the data for research purposes, limitations exist within each instrument. Limitations specific to the NRS and other similar assessments include the applicability of NRS data to CBT for chronic pain and validity of the data. As CBT for chronic pain is designed to ease the experience of pain, rather than decrease the pain itself (Castro et al. 2012), PAIN data may be less than adequate as an outcome measure for psychological intervention for chronic pain. Further, resulting from the subjectivity of pain ratings, NRS data can fluctuate based on experiences, mood, fatigue, or time of day (Schneider et al., 2012; Tupper, Rosenberg, Pahwa, & Stinson, 2013; Bartley, Robinson, & Staud, 2017). Also, a greater degree of variation within NRS data has been found in patients with depressive symptoms than in those with fewer depressive symptoms (Zakoscielna & Parmelee, 2013). Moreover, in a study of pain variability in two separate samples of patients with a chronic painful disease, Schneider et al. found that day-to-day fluctuations in pain ratings averaged 13% in one sample and 17% in the other. Per the IMMPACT report (Dworkin et al., 2008), changes in NRS data between 10 – 30% are considered 'minimally significant', thus creating an inherent discrepancy in the study of pain ratings. Limitations specific to the ADL primarily focus on the scope of the data, in that it only assesses physical function, without regard to pain ratings or variables assessed in the QOL domain (Cress et al., 2005). In addition, as it relates to the study of chronic pain, performance on both the ADL and WHOQOL may be affected by true fluctuations in chronic pain (Cress et al., 2005; Ehde et al., 2014; Jonsdottir, et al., 2014; de Figueiredo & Griffith, 2016).

One final limitation to each of these instruments, as with all instruments, concerns the accuracy of the data to reflect the true score of the individual. This concept is known

as error variance, the accuracy of a score obtained by an instrument compared to the true score of an individual (Coaley, 2014). This deviation between the true and acquired scores can vary for many reasons, and may include the relationship with test administrators, the measures being used, the procedures involved in testing, the surrounding environment, and the context in which the test is taken. Sources of error variance specific to this research, known as systematic error, has a predictable effect on scores by introducing a consistently measurable bias across administrations and populations. This can include factors in test design, including standardization and ambiguity in print materials, and response styles adopted by participants, such as giving socially desirable responses rather than a more accurate response which may be less socially desirable.

One major contributor to the systematic error for any test involves the degree of difficulty in completing the test. That is, the performance of any one participant on any given test may vary based on the degree of difficulty in completing that test (Coaley, 2014). Thus the question of clinical efficiency, the ability to produce the most accurate data with the least degree of difficulty, presents itself. As was stated previously, the degree of difficulty in completing the three outcome measures of interest is readily observable. The NRS can be completed by verbally responding to one question (Dworkin et al., 2008). The WHOQOL can be completed typically within 5 – 10 minutes by responding to 26 Likert-type questions, followed by a series of computations conducted by the test administrator to produce a score on each of four domains of the WHOQOL (Shawver et al., 2016; Skevington et al., 2004). By contrast, the ADL requires approximately one hour of physical activity which may or may not be painful or

strenuous depending on medical conditions and current pain ratings (Cress et al., 2005). Accordingly, if systematic error within each instrument were statistically equal, the relative clinical efficiency of these instruments would be ordered NRS, WHOQOL, and ADL, in order from greatest to least clinical efficiency, respectively.

H. Hypothesis

After examining the meta-analyses (Deckert et al., 2016; Pike et al., 2016; Veehof et al., 2016) and the outcome categories and outcome measures within each domain, sufficient evidence was found to justify using the NRS, ADL, and WHOQOL, as outcome measures representative of the larger outcome domains of PAIN, PF, and QOL, respectively. Although these outcome measures were selected because of their representation in the literature (NRS, ADL) or because of the established reliability and validity of the instrument in the absence of an established representative in the domain (WHOQOL), the question remains of the relative clinical efficiency of the three. The clinical efficiency of each outcome measure can be established by determining the usefulness of the data acquired by each instrument compared to the degree of difficulty in obtaining that useful data (Coaley, 2014). An examination of the relative clinical efficiency of these three outcome measures, and by extension the outcome domains, can establish a reliable and valid method of tracking patient progress across and following CBT for chronic pain. This can also reduce the degree of physical and financial burden on the patient for both research and treatment purposes.

The results of the meta-analyses (Deckert et al., 2016; Pike et al., 2016; Veehof et al., 2016) effectively demonstrated the use of the PAIN, PF, and QOL outcome domains

in chronic pain research. Notably, little information is available concerning the relative clinical efficiency of commonly used assessments within these outcome categories. Thus, the purpose of this study was to examine the within-subjects variability in each of the assessments described within each outcome domain as a measure of systematic error similar to Shavelson and Webb (1991) and Briesch, Swaminathan, Wels, and Chafouleas (2014). A comparison of these instruments revealed a significant difference in the relative clinical efficiency, i.e., the reliability of the data as compared to the degree of difficulty in obtaining said data, of these measures to track the progress of chronic pain patients across and following CBT. In light of this, I expected to find that the WHOQOL would demonstrate less systematic error attributable to the instrument by controlling for the greatest degree of variance across multiple assessments of individual participants, followed by the ADL and the NRS, respectively. Thus, I hypothesized that the QOL domain would show greater clinical efficiency than the other domains, and that the PAIN domain would show the least clinical efficiency relative to the other domains.

Considering the impact of chronic pain on the functionality and daily lives of individuals with these painful conditions, the benefits of this research is clear. The establishment of an outcome measure or core group of measures that provide consistently reliable data without adding undue burden to the patient would benefit both the patient and the research. The results of this research can be used to demonstrate which of the three outcome measures is most efficient for research purposes. This allows for the establishment of a standard outcome measure for research into CBT for chronic pain, and may have applications for other treatment modalities as well.

# CHAPTER II

# METHODOLOGY

## A. Participants

As this study was conducted as a retrospective archival study, no interaction occurred between the researcher and patients. Data were gathered from an existing archive compiled from the medical records of patients ($N$ = 1,883) who were treated for chronic pain at a mid-sized tertiary multidisciplinary chronic pain treatment facility in Huntsville, Alabama between the dates of June 2013 and August 2018 inclusive. Of those 1,883 patients, 1,567 patient files were missing at least one necessary data point, resulting in a final sample size of 316. To ensure that the 316 participants constituted a representative sample of the archive, mean scores were compared between the archive and the sample. No significant differences were observed. This demonstrates that the sample was representative of the larger archive. Participant demographics for this study revealed the sample to be comprised of 211 women (65%) and 277 Caucasians (86%), and age ranged from 19.1 – 79.9, ($M$ = 47.3, $SD$ = 10.8, $Mo$ = 52.2). Consent was obtained from all patients to allow their health data to be used for research purposes prior to the first session of CBT with no incentives for participating and no ramifications for opting out. All HIPAA, APA, AMA, and Human Subjects standards were followed. Descriptive statistics of the sample are available in Table 2.1.

Table 2.1 Patient demographic data

| Category | Frequency | | % |
|---|---|---|---|
| Female | 208 | | 65.8 |
| Male | 108 | | 34.2 |
| African-American | 42 | | 13.3 |
| Caucasian | 273 | | 86.4 |
| Other | 1 | | 0.3 |
| Age Range | *M* | *SD* | *Mo* |
| 19.1 – 79.9 | 47.29 | 10.83 | 52.284 |

B. Design

This study employed a within-subjects repeated measures design to examine the relative clinical efficiency of the three outcome domains of PAIN as measured using the NRS, PF as measured using the ADL, and QOL as assessed using the WHOQOL, which were identified through a review of meta-analyses, as they relate to tracking patient progress across CBT for chronic pain. For the purpose of this study, clinical efficiency was defined as the systematic error attributable to the assessment instrument within its respective outcome domain compared to the effort and time required to obtain the intended data from that instrument. The purpose of this study was not to debunk or

endorse any one outcome measure, but simply to determine which, if any, of these three measures revealed the most useful information when gathered from in-practice patients, as opposed to a laboratory study, with the least time and effort on the part of the patient for research purposes. Consequently, I make no argument for the use of one outcome measure over another in terms of treatment planning.

The construct of interest to this study is the systematic error attributable specifically to one instrument within each of the three outcome domains of QOL, PF, and PAIN. Commonly used outcome measures within each outcome domain served as dependent variables. The NRS was assessed as a subjective measure of PAIN. Physical Therapist-rated ADL was examined as a measure of PF. The WHOQOL was examined as a measure of QOL. Data from all three instruments recorded from assessments immediately prior to the beginning of and immediately following the completion of a 12-session group CBT for chronic pain were analyzed to determine the systematic error attributable specifically to the instruments using generalizability theory similar to Shavelson and Webb (1991) and Briesch et al. (2014). An analysis of this type allows for an accurate examination of the systematic error attributable to the instrument, and further allows for a replication with expansion study to determine which, if any, of these outcome measures and domains reveal the greatest clinical efficiency for research purposes.

## C. Materials

As this study was conducted as a retrospective archival study, the primary materials for this study consist of Microsoft Excel for data collection and IBM SPSS

version 24 for data analysis. Other materials used in this study include the questionnaires and rating systems which comprise the dependent variables. A brief description of each follows.

1. Activities of Daily Living

The ADL is a rating made by a physical therapist based on the ability of the participant to complete certain activities common in daily life for a specific length of time (Cress et al., 2005). The assessment considers the participant's general overall physical condition, vital signs, current diagnoses and disabilities, current pain medications, and the need for assistance to complete an activity, e.g., a cane for walking. The maximum score for the assessment is 50 points. An ADL score of 45 or greater is considered minimal interference with daily activities. As a tool for treatment planning, the ADL is a clinically useful instrument which has been standardized for several specific diagnoses, and has an established test-retest reliability ranging .93 - .98. The ADL is a proprietary material and thus cannot be included in an appendix.

2. Numeric Rating Scale of Overall Pain

The NRS is a verbal rating of overall pain on a ratio scale of *0 – 10*. The NRS score is assessed by a registered nurse during a regular office visit. Patients are instructed to rate their overall pain at that particular moment on a scale of *0 – 10*, with *0* meaning no pain and *10* meaning worst pain imaginable. Some patients chose to give their pain rating as a range. In those instances, the NRS rating was recorded as the mean rounded to the nearest whole number within that specified range, e.g., a range of *3 – 5* was recorded as *4*; a range of *3 – 6* was

recorded as *5*. As pain is a purely subjective experience, an objective analysis of the precision of an instrument designed to estimate momentary pain level is almost impossible. Several studies have been conducted to examine the within-subjects variation within the NRS, with typical results revealing between 13 – 18% variation across repeated assessments (Schneider et al., 2012). The NRS is a 1-question verbal test and therefore cannot be included in an appendix.

3. World Health Organization Quality of Life – BREF

The WHOQOL is a 26-question instrument designed to assess QoL across physical (7 questions; Chronbach's $\alpha = 0.82$), psychological (6 questions; Chronbach's $\alpha = 0.81$), social (3 questions; Chronbach's $\alpha = 0.68$), and environmental (8 questions; Chronbach's $\alpha = 0.80$) scales (Shawver et al., 2016; Skevington et al., 2004). The WHOQOL has been examined extensively for discriminant validity between medical populations and healthy controls with significant results in all four domains of the WHOQOL at the $\alpha = .01$ level. The instrument was developed using a cross-sectional design of over 10,000 patients in 23 countries across all populated continents. Scores range from *0 – 100* on a ratio scale. The WHOQOL is a proprietary material and thus cannot be included in an appendix.

D. Procedure

The data for this study were collected by staff at the host facility for research purposes. Informed consent was obtained from all individual patients prior to data collection with no benefits for participating or repercussions for declining to participate.

All procedures performed in this study were in accordance with the ethical standards of the UAH Institutional Review Board of Human Subjects Committee (see APPENDIX A). As this study included no patient contact, and as consent to allow medical and psychological data to be used for research purposes was given by each patient prior to inclusion, this study was approved to proceed without obtaining direct consent from each patient. Patients in this study engaged in CBT in accordance with prescribed standards (Thorn, 2004). The CBT program consisted of an initial intake session, at which time patients completed the first set of assessments, and 12 group-CBT content sessions, followed by a final assessment session. Patients attended one group session on a weekly or biweekly basis. The mean completion time was 24.16 weeks (range = 12 – 26 weeks). Following the final session, each assessment was scored and entered into electronic medical records by staff at the host facility. Scores from these assessments were retrieved by the primary investigator and entered into an Excel spreadsheet. No personally identifiable information was recorded.

E. Statistics

Prior to any statistical analysis, data from the NRS and ADL were converted to a *100*-point scale to correspond with the WHOQOL to allow for an accurate comparison of variance components between each instrument. A repeated measures ANOVA was used for each outcome measure dependent variable at initial and final assessment to establish that changes in response to CBT within this participant set did occur for each dependent variable, and that those changes were similar to changes reported in previous studies. Next, scores from the initial assessment of each scale were subtracted from the scores at

the final assessment of each scale within an individual to give an overall improvement score. Systematic error components were then analyzed within each outcome measure dependent variable by conducting a variance components analysis using the improvement score as the dependent variable and using patients and each scale within the three outcome measures as random factors. Expected variance components (EVC) were then calculated similar to Shavelson and Webb (1991) and Briesch et al. (2014) to examine the percentage of total variance attributable to each source (person, scale, error). Per Shavelson and Webb, EVC were calculated using the following formula:

$$E\sigma_p^2 = \frac{(MS_p - \sigma_e^2)}{n_s}, \tag{2.1}$$

where $E\sigma_p^2$ is the EVC attributable to person, $MS_p$ is the mean square for person obtained from the ANOVA output, $\sigma_e^2$ is the EVC for the error term, i.e., equal to the mean square for the error term obtained from the ANOVA output, and $n_s$ equals the total number of items in the scale, i.e., the combined 6 scales from the three outcome measures. The formula for calculating EVC for scales is similar to Equation (2.1), with the exception of using the mean square for scales and the total number of participants.

Significant main effects were further analyzed using paired-samples t-tests in accordance with the stated a priori hypothesis of the source of the variance within the three outcome measures. Finally, a variance components analysis and EVC was conducted to examine the four scales of the WHOQOL without the other two outcome measures, and a third variance components analysis and EVC was conducted to examine

the Physical scale of the WHOQOL in contrast with the other three scales of the WHOQOL.

# CHAPTER III

## RESULTS

The results of the repeated measures ANOVA revealed that changes did occur in all three dependent variables as a result of CBT. The results revealed significant changes occurred both within each scale, $F(1, 1890) = 356.01$, $MSE = 213.49$, $p < .001$, $\eta^2_p = .16$, and between the scales, $F(5, 1890) = 90.67$, $MSE = 454.63$, $p < .001$, $\eta^2_p = .44$. This suggests that the data for each of the three dependent variables were similar to data reported in other studies which demonstrate patient progress as a result of CBT for chronic pain (Hughes et al., 2018; Kleinstauber et al., 2017; Sachser et al., 2017; Zinzow et al., 2017; Ehde et al., 2014; Farrer et al., 2014; Watts et al., 2013; Lynch et al., 2011). Therefore, the results of this study can be generalized to all CBT for chronic pain treatments.

Next, systematic error was examined to determine the sources of variance in each dependent variable outcome measure using an SPSS Variance Components analysis. The results of the Variance Components analysis indicated that a significant main effect was present both for scale, $F(5, 1575) = 157.74$, $MSE = 186.56$, $p < .001$, $\eta^2_p = .33$, and for person, $F(315, 1575) = 2.29$, $MSE = 186.56$, $p < .001$, $\eta^2_p = .31$, with moderate effect sizes observed for both main effects. These results were then converted into EVC as indicated in Shavelson and Webb (1991) which demonstrated that the degree of variance

attributable to the scale was greater than the degree of variance attributable to the person, $EVC_{person}$ = 40.07 (12.55%), $EVC_{scale}$ = 92.53 (28.99%), $EVC_{error}$ = 186.56 (58.45%) (see Figure 3.1).

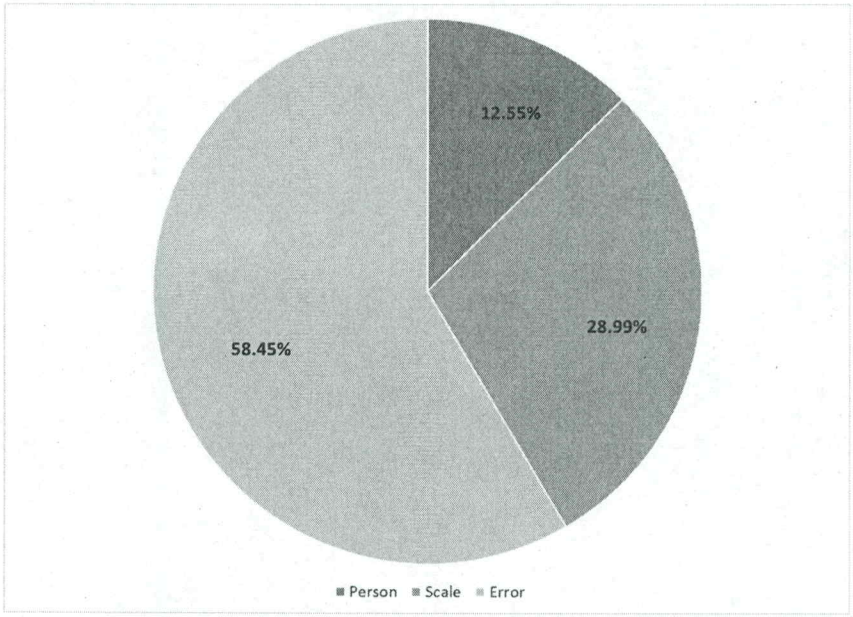


Figure 3.1 Percentage of variance using all six scales

These results were further analyzed using paired-samples t-tests to analyze the a priori hypothesis of significant differences in the degree of variance attributable to each outcome measure. Significant changes were observed across all six scales within the three outcome measures, but the degree of variance attributable to each scale varied, as indicated by effect size. The WHOQOL accounted for the greatest degree of variance as indicated by the large effect size on the Physical scale, $t(316) = 15.60$, $p < .001$, $d = 1.21$, the Psychological scale, $t(315) = 11.59$, $p < .001$, $d = .94$, and the Environmental scale, $t(315) = 9.85$, $p < .001$, $d = .79$, and by the moderate effect size observed in the Social scale, $t(315) = 7.38$, $p < .001$, $d = .59$. The NRS accounted for a small-to-moderate

degree of variance, $t(315) = -5.65$, $p < .001$, $d = .45$, and the ADL accounted for the least degree of variance as evidenced by the small effect size, $t(315) = 4.26$, $p < .001$, $d = .34$.

Next, systematic error was examined to determine the sources of variance in each scale of the WHOQOL using an SPSS Variance Components analysis. Results indicated a moderate main effect for scale, $F(3, 945) = 129.63$, $MSE = 203.63$, $p < .001$, $\eta^2_p = .29$, and a large effect for person, $F(315, 945) = 2.99$, $MSE = 203.63$, $p < .001$, $\eta^2_p = .50$. These results were then converted into EVC as with the previous analysis per Shavelson and Webb (1991). The results of the EVC analysis demonstrated the degree of variance attributable to person was greater than the total degree of variance attributable to the scales of the WHOQOL, $EVC_{person} = 101.31$ (26.16%), $EVC_{scale} = 82.30$ (21.25%), $EVC_{error} = 203.63$ (52.58%) (see Figure 3.2).

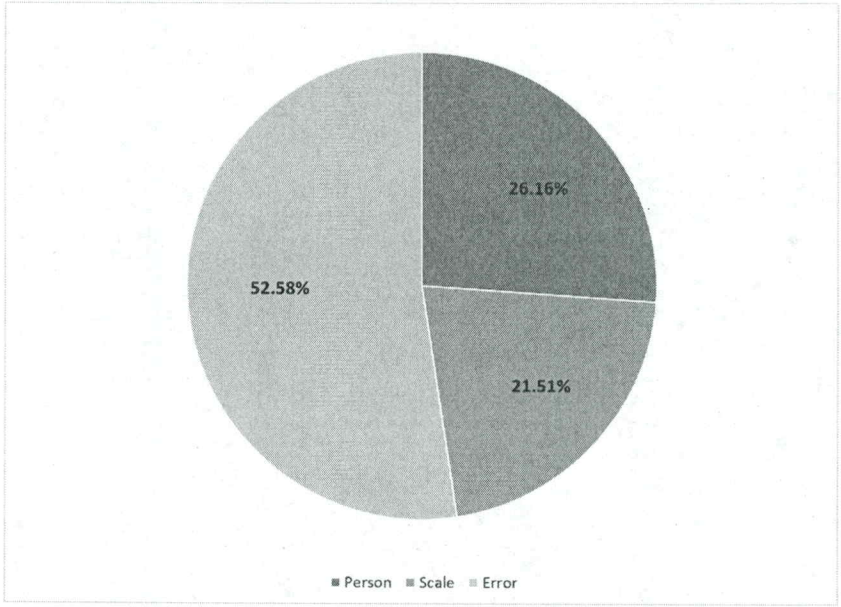


Figure 3.2 Percentage of variance using only the WHOQOL

Finally, as the Physical scale of the WHOQOL accounted for the greatest degree of variance within the six scales, systematic error was examined to determine the degree of variance attributable to the Physical scale of the WHOQOL using an SPSS Variance Components analysis. Results indicated a moderate main effect both for the Physical scale of the WHOQOL, $F(1, 632) = 711.57$, $MSE = 108.03$, $p < .001$, $\eta^2_p = .52$, as well as for person, $F(315, 632) = 5.64$, $MSE = 142.17$, $p < .001$, $\eta^2_p = .74$. These results were then converted into EVC which demonstrated that the degree of variance attributable to person was approximately half that attributable to scale, but still greater than the degree of variance attributable to person in the initial analysis when all six scales of the outcome measures were included, $EVC_{person} = 103.69$ (13.88%), $EVC_{scale} = 242.00$ (32.39%), $EVC_{error} = 401.50$ (53.73%) (see Figure 3.3).

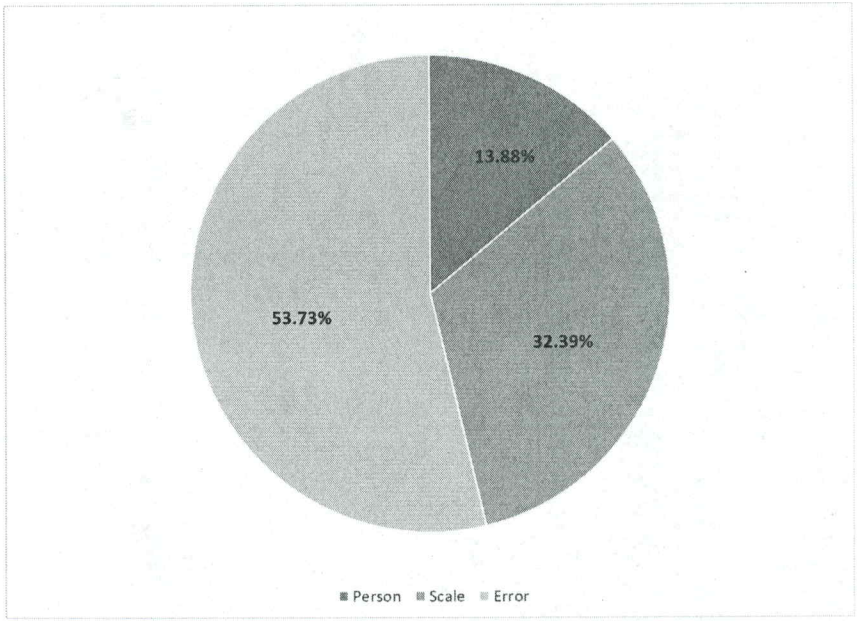


Figure 3.3 Percentage of variance using the Physical scale only

# CHAPTER IV

# DISCUSSION

## A. Summary of Results

The results of the initial data analysis indicated that the data for each of the three dependent variables were typical of patients undergoing CBT for chronic pain. Consequently, the results of this study can be generalized to all CBT for chronic pain treatments. The results of the initial variance components analysis indicated a significant difference between the variance in score attributable to the individual being assessed and the variance attributable to the outcome measure instrument itself. These results indicated that, while CBT for chronic pain has been shown to improve overall patient functioning (Castro et al., 2012; Ehde et al., 2014; Hughes et al., 2018), the observed change in scores attributable to changes in the individual is almost completely obscured by the systematic error found within these three outcome measures. This precludes the determination of any functional improvement resulting from CBT for chronic pain, as nearly all of the variance is attributable to the six scales within the three outcome measure instruments. In essence, the three outcome measures do not provide the same indications of differences in functional improvement resulting from CBT for chronic pain, and this variance within the scales of the outcome measure battery essentially eclipses any benefit to the patient resulting from CBT for chronic pain.

The results of the initial variance components analysis indicated that a significant difference existed between the six scales of the three outcome measures. In support of the first part of my hypothesis, post hoc analysis indicated that the WHOQOL accounted for the greatest degree of variance. This was demonstrated through the large effect sizes of the paired-samples t-tests for the Physical, Psychological, and Environmental scales of the WHOQOL, as well as the moderate effect size for the Social scale of the WHOQOL, all of which were much larger than the effect sizes for the NRS and ADL. These large effect sizes are unsurprising, as the benefits of CBT, i.e., the retrained cognitive processes which allow for more adaptive cognitions, attitudes, and behaviors (Castro et al. 2012), are directly assessed in these scales of the WHOQOL (Hughes et al., 2018). The moderate effect sizes for the Social scale of the WHOQOL could be a result of an interaction with another moderator, such as social support, as demonstrated in Matos, Bernardes, and Goubert (2016). Accordingly, the WHOQOL accounted for the greatest degree of variance of the three outcome measures. Given the previously established ease of utility of this instrument relative to the other two outcome measures, these results establish the WHOQOL as having the greatest clinical efficiency relative to the other two outcome measures.

Contrary to the second part of my stated hypothesis, the ADL accounted for the least degree of variance compared to the other two outcome measures as demonstrated through the effect sizes of post hoc paired-samples t-tests. The relatively similar degree of variance attributable to the ADL and NRS indicate that both are moderately effective at tracking changes in their respective outcome domains across CBT for chronic pain, but that neither are as effective as the least effective scale of the WHOQOL. This diminished

ability to track patient progress across CBT for chronic pain relative to the WHOQOL is likely a result of stated limitations within each scale. The ADL assesses physical functioning using a standardized rating system scored by a trained third-party observer and only assesses physical functioning (Cress et al., 2005) which is unlikely to change as a result of CBT for chronic pain as described by Castro et al. (2012), in that CBT for chronic pain does not directly address either pain or physical functioning. Limitations within the NRS result from the extreme subjectivity of the scale, as NRS data can fluctuate based on minor changes within a number of other variables (Schneider et al., 2012; Tupper et al., 2013; Zakoscielna & Parmelee, 2013; Bartley et al., 2017). Despite the exaggerated sensitivity of the NRS, the ease of utility of the instrument established it as a more clinically efficient outcome measure than the ADL, although both were found to be much less clinically efficient than the WHOQOL. The degree of difficulty in the use of the ADL and the relatively low degree of variance for which it accounted established the ADL as the least clinically efficient outcome measure relative to the other two outcome measures.

The results of the second variance analysis indicated that the four scales of the WHOQOL in absence of the other two measures accounted for approximately twice the amount of variance attributable to person when compared to the EVC for all six scales. These results support the post hoc analyses conducted for the first variance components analysis, which suggested the WHOQOL accounted for the greatest degree of variance, and that the Physical scale accounted for the greatest degree of variance of the four scales of the WHOQOL. As answering the 7 questions on the Physical scale of the WHOQOL is easier than answering the 26 questions from the complete WHOQOL, these results

indicated that the use of the Physical scale of the WHOQOL in absence of the other three scales may be more clinically efficient (Coaley, 2014). The results of the final variance components analysis comparing the Physical scale against the other three scales of the WHOQOL revealed the EVC attributable to both scale and person was approximately equivalent to the respective EVC in the initial variance components analysis. These results suggested the effectiveness of the WHOQOL for tracking patient progress across CBT for chronic pain would be compromised by using only the Physical scale and would be no more effective than using all three outcome measures. Further, the inclusion of all four scales of the WHOQOL would allow for an assessment of all four components of QoL rather than just the physical component, therefore providing a more accurate assessment of the overall QoL of the patient.

B. Limitations and Future Research

Although the results of this study were promising, several limitations presented themselves. Primarily, this study concerned outcome measures for CBT for chronic pain for research purposes only. This report made no claim as to the efficiency of one outcome measure over another concerning treatment planning or patient assessment. Therefore, when considering the results of this report for the purposes of treatment planning, clinicians should interpret these results in light of the overall clinical impression of the patient. Another major limitation to this study concerned the inclusion of only one treatment modality, CBT for chronic pain. While similarities exist between this and other treatment modalities, such as those considered by Deckert et al. (2016), Pike et al. (2016), and Veehof et al. (2016), this report made no claim as to the clinical

efficiency of the WHOQOL, NRS, or ADL when used for other modalities. One further limitation concerned the ability of the three outcome measures to represent their outcome domains. The assumption that these three outcome measures adequately represented their respective domain, and that the domains adequately reflected the current state of the literature, resulted primarily from the findings of the three meta-analyses. Although the methodologies of the three meta-analyses were sound, no study or meta-analysis is without limitation. As a result, the adequacy of these outcome measures in representing their respective domain was a direct reflection of the adequacy of the meta-analyses in reflecting the state of the literature.

One final limitation involved the composition of the subject pool. Although the sample did include patients from both urban and rural settings, as well as patients with a range of backgrounds, including farming, construction, military service, food service, teaching, and professional careers (i.e., engineering, accounting), the sample was comprised primarily of Caucasian female patients from the southeastern United States. As such, this study did not consider regional, cultural, or ethnic differences that may affect the clinical efficiency of the three outcome measures. Further research is warranted to address these issues.

The limitations presented in the previous paragraphs give rise to several interesting areas for future research. To address the limitation concerning the clinical efficiency of these three outcome measures for treatment planning, future research could be conducted to establish a core set of outcome measures for treatment planning similar to Deckert et al. (2016). The establishment of a core set of outcome measures could then be used as a standardized measure of overall patient progress for treatment planning.

Further research could also be conducted to examine the clinical efficiency of these three outcome domains and outcome measures to other similar treatment modalities as in Pike et al. (2016) and Veehof et al. (2016). Similar results across treatment modalities would allow for an accurate comparison of patient progress across these modalities which, in turn, would allow for an accurate comparison of studies regarding patient-specific variables similar to Hughes et al. (2018). A comparison of this sort would then provide clinicians with the necessary information to make the best decision of treatment modality for the individual patient.

Regarding the ability of the outcome measures to represent their respective domains and of the meta-analyses to represent the current state of the literature, future researchers could conduct a meta-analysis of the literature since the publication of the IMMPACT report (Dworkin et al., 2008) for the specific purpose of creating a core set of outcome measures within each domain, followed by a study similar to the current study, with the caveat that the future study would examine the clinical efficiency of outcome measures within a domain rather than across domains. This would allow for the most accurate representation of the state of the literature and of the individual outcome domains. Concerning the final limitation of the composition and demographic information of the subject pool, future research could conduct multiple replication studies in multiple locations to control for regional and cultural differences.

Two further areas of future research present themselves as a natural progression from the current study. The nature of a generalizability study is to examine the sources of variability and provide as much information about that variability as possible (Shavelson & Webb, 1991). The logical progression following a generalizability study

results in a decision study. The purpose of a decision study is to design the best possible application of the results of the generalizability study, similar to conducting a confirmatory factor analysis to support the results of an exploratory factor analysis. Therefore, future researchers should build upon the results of this generalizability study to determine the most efficient way to implement the results of the current study. Finally, as the nature of this study was to determine the clinical efficiency of the three outcome measures with the ultimate goal of improving the utility of these measures for research, future researchers could examine the WHOQOL using factor analysis and generalizability theory to determine which questions and scales specifically apply to CBT for chronic pain. This would allow for the modification of the WHOQOL to reduce the number of questions while maintaining reliability and validity, and ultimately improve the clinical efficiency of the instrument for chronic pain research.

C. Summary

In summation, CBT for chronic pain is an empirically validated psychological therapy for chronic pain which has been shown to improve overall patient functioning. In order to adequately assess patient improvement, patient progress must be tracked across treatment, and although some recommendations have been made, no standard set of outcome measures has been implemented. An examination of recent meta-analyses has demonstrated that currently used outcome measures for psychological treatment for chronic pain can be categorized into the three outcome domains of PF, PAIN, and QOL. The most commonly used outcome measure within each domain was used as a representative of the domain to analyze the variability in improvement scores across a 12-

week CBT for chronic pain. Outcome measures included the ADL (PF), the NRS (PAIN), and the WHOQOL (QOL). Variability was analyzed using generalizability theory as recommended by Shavelson and Webb (1991). The results indicated that the WHOQOL accounts for a large majority of the variation in improvement scores, and far surpasses the degree of variation for which is accounted by the ADL or NRS. These results were then compared to a previously established degree of difficulty in the use of each of the three outcome measures. The results indicated that the WHOQOL was much more clinically efficient than either the ADL or NRS for research purposes. Limitations included the use of CBT as the only treatment modality, the use of these outcome measures for research purposes only rather than for in-practice decision-making, the ability of the meta-analyses to adequately reflect the current state of the literature, and the relative homogeneity of the sample demographics. Future research directives were presented to address these limitations. Future research directives were also given to support or refute the results of this study through replication and to expand and properly implement the results of this generalizability study using a decision study per Shavelson and Webb.

# APPENDIX A

## INSTITUTIONAL REVIEW BOARD APPROVAL FORM

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

| |
|---|
| ☒ Expedited (see pg 2) |
| ☐ Exempted (see pg 3) |
| ☐ Full Review |
| ☐ Extension of Approval |

August 27th 2018

Justin Hughes
Department of Psychology
University of Alabama in Huntsville

Dear Mr. Hughes,

The UAH Institutional Review Board of Human Subjects Committee has reviewed your proposal, *Variability in Chronic Pain Patient Assessment Measures,* and found it meets the necessary criteria for approval. Your proposal seems to be in compliance with this institutions Federal Wide Assurance (FWA) 00019998 and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46).

Please note that this approval is good for one year from the date on this letter. If data collection continues past this period, you are responsible for processing a renewal application a minimum of 60 days prior to the expiration date.

No changes are to be made to the approved protocol without prior review and approval from the UAH IRB. All changes (e.g. a change in procedure, number of subjects, personnel, study locations, new recruitment materials, study instruments, etc) must be prospectively reviewed and approved by the IRB before they are implemented. You should report any unanticipated problems involving risks to the participants or others to the IRB Chair.

If you have any questions regarding the IRB's decision, please contact me.

Sincerely,

Bruce Stallsmith

Bruce Stallsmith
IRB Chair
Professor, Biological Sciences

# REFERENCES


Bartley, E. J., Robinson, M. E., & Staud, R. (2017). Pain and fatigue variability patterns distinguish subgroups of fibromyalgia patients. *The Journal of Pain*, doi:10.1016/j.jpain.2017.11.014

Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology*, *52*(1), 13-35. doi:10.1016/j.jsp.2013.11.008

Castro, M. C., Daltro, C., Kraychete, D. C., & Lopes, J. (2012). The cognitive behavioral therapy causes an improvement in quality of life in patients with chronic musculoskeletal pain. *Arquivos De Neuro-Psiquiatria*, *70*(11), 864-868. doi:10.1590/S0004-282X2012001100008

Coaley, K. (2014). *An introduction to psychological assessment and* psychometrics (2nd Ed.). Thousand Oaks, CA: Sage Publications, Inc.

Cress, M.E., Petrella, J.K., Moore, T.L., & Schenkman, M.L. (2005). Continuous-scale physical functional performance test: Validity, reliability, and sensitivity of data for the short version. *Physical Therapy, 85,* 323 – 335. Retrieved from https://elib.uah.edu/login?url=https://search-proquest-com.elib.uah.edu/docview/67555140?accountid=14476

de Figueiredo, J. M., & Griffith, J. L. (2016). Chronic pain, chronic demoralization, and the role of psychotherapy. *Journal of Contemporary Psychotherapy*, *46*(3), 167-177. doi:10.1007/s10879-016-9331-x

Deckert, S., Kaiser, U., Kopkow, C., Trautmann, F., Sabatowski, R., & Schmitt, J. (2016). A systematic review of the outcomes reported in multimodal pain therapy for chronic pain. *European Journal of Pain*, *20*(1), 51-63. doi:10.1002/ejp.721

Dworkin, R. H., Turk, D. C., Wyrwich, K. W., Beaton, D., Cleeland, C. S., Farrar, J. T., ... & Zavisic, S. (2008). Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *The Journal of Pain, 9*(2), 105-121. doi:10.1016/j.jpain.2007.09.005

Ehde, D. M., Dillworth, T. M., & Turner, J. A. (2014). Cognitive-behavioral therapy for individuals with chronic pain: Efficacy, innovations, and directions for research. *American Psychologist*, *69*(2), 153-166. doi:10.1037/a003574

Farrer, L. M., Griffiths, K. M., Christensen, H., Mackinnon, A. J., & Batterham, P. J. (2014). Predictors of adherence and outcome in internet-based cognitive behavior therapy delivered in a telephone counseling setting. *Cognitive Therapy and Research*, *38*(3), 358-367. doi:10.1007/s10608-013-9589-1

Gotink, R. A., Chu, P., Busschbach, J. V., Benson, H., Fricchione, G. L., & Hunink, M. M. (2015). Standardised mindfulness-based interventions in healthcare: An overview of systematic reviews and meta-analyses of RCTs. *Plos ONE*, *10*(4). doi:10.1371/journal.pone.0124344

Hughes, J. M., Seemann, E. A., George, J. M., & Willis, K. D. (2018). The effects of pre-treatment depressive symptoms on quality of life across cognitive behavioral therapy for chronic pain. *Journal of Clinical Psychology in Medical Settings*. doi:10.1007/s10880-018-9568-5

Inoue, S., Kobayashi, F., Nishihara, M., Arai, Y. P., Ikemoto, T., Kawai, T., Inoue, M., Hasegawa, T., & Ushida, T. (2015). Chronic pain in the Japanese community—Prevalence, characteristics and impact on quality of life. *PLoS ONE*, *10*(6). doi:10.1371/journal. pone.0129262

Jonsdottir, T., Aspelund, T., Jonsdottir, H., & Gunnarsdottir, S. (2014). The relationship between chronic pain pattern, interference with life and health-related quality of life in a nationwide community sample. *Pain Management Nursing*, *15*(3), 641-651. doi:10.1016/j.pmn.2013.07.005

Kleinstauber, M., Lambert, M. J., & Hiller, W. (2017). Early response in cognitive-behavior therapy for syndromes of medically unexplained symptoms. *BMC Psychiatry*, *17*. doi: 10.1186/s12888-017-1351-x

Knoerl, R., Smith, E. L., & Weisberg, J. (2016). Chronic pain and cognitive behavioral therapy: An integrative review. *Western Journal of Nursing Research*, *38*(5), 596-628. doi:10.1177/0193945915615869

Lynch, K., Berry, C., & Sirey, J. (2011). A group-oriented inpatient CBT programme: A pilot study. *The Cognitive Behaviour Therapist*, *4*(1), 38-51. doi:10.1017/S1754470X10000152

Matos, M., Bernardes, S. F., & Goubert, L. (2016). The relationship between perceived promotion of autonomy/dependence and pain-related disability in older adults with chronic pain: The mediating role of self-reported physical functioning. *Journal of Behavioral Medicine*, *39*(4), 704-715. doi:10.1007/s10865-016-9726-x

Ojala, T., Häkkinen, A., Karppinen, J., Sipilä, K., Suutama, T., & Piirainen, A. (2015). Revising the negative meaning of chronic pain—A phenomenological study. *Chronic Illness*, *11*(2), 156-167. doi:10.1177/1742395314555236

Pike, A., Hearn, L., & C. de C. Williams, A. (2016). Effectiveness of psychological interventions for chronic pain on health care use and work absence: Systematic review and meta-analysis. *Pain*, *157*(4), 777-785. doi:10.1097/j.pain.0000000000000434

Sachser, C., Keller, F., & Goldbeck, L. (2017). Complex PTSD as proposed for ICD-11: Validation of a new disorder in children and adolescents and their response to Trauma-Focused Cognitive Behavioral Therapy. *Journal of Child Psychology and Psychiatry*, *58*(2), 160-168. doi:10.1111/jcpp.12640

Schneider, S., Junghaenel, D. U., Keefe, F. J., Schwartz, J. E., Stone, A. A., & Broderick, J. E. (2012). Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: Associations with psychological variables. *Pain*, *153*(4), 813-822. doi:10.1016/j.pain.2012.01.001

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage Publications, Inc.

Shawver, Z., Griffith, J. D., Adams, L. T., Evans, J. V., Benchoff, B., & Sargent, R. (2016). An examination of the WHOQOL-BREF using four popular data collection methods. *Computers in Human Behavior*, *51*(Part A), 446-454. doi:10.1016/j.chb.2015.09.030

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A Report from the WHOQOL Group. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, *13*(2), 299-310. doi:10.1023/B:QURE.0000018486.91360.00

Stubbs, B., Schofield, P., & Patchay, S. (2016). Mobility limitations and fall-related factors contribute to the reduced health-related quality of life in older adults with chronic musculoskeletal pain. *Pain Practice*, *16*(1), 80-89. doi:10.1111/papr.12264

Thorn, B.E. (2004). *Cognitive therapy for chronic pain.* New York: The Guilford Press.

Trinanes, Y., González-Villar, A., Gómez-Perretta, C., & Carrillo-de-la-Peña, M. T. (2015). Suicidality in chronic pain: Predictors of suicidal ideation in Fibromyalgia. *Pain Practice*, *15*(4), 323-332. doi:10.1111/papr.12186

Tupper, S. M., Rosenberg, A. M., Pahwa, P., & Stinson, J. N. (2013). Pain intensity variability and its relationship with quality of life in youths with juvenile idiopathic arthritis. *Arthritis Care & Research*, *65*(4), 563-570. doi:10.1002/acr.21850

Veehof, M. M., Trompetter, H. R., Bohlmeijer, E. T., & Schreurs, K. M. G. (2016). Acceptance- and mindfulness-based interventions for the treatment of chronic pain: A meta-analytic review. *Cognitive Behaviour Therapy*, *45*(1), 5–31. doi:10.1080/16506073.2015.1098724

Watts, S., Mackenzie, A., Thomas, C., Griskaitis, A., Mewton, L., Williams, A., & Andrews, G. (2013). CBT for depression: A pilot RCT comparing mobile phone vs. computer. *BMC Psychiatry*, *13.* doi:10.1186/1471-244X-13-49

Zakoscielna, K. M., & Parmelee, P. A. (2013). Pain variability and its predictors in older adults: Depression, cognition, functional status, health, and pain. *Journal of Aging and Health*, *25*(8), 1329-1339. doi:10.1177/0898264313504457

Zautra, A. J., Davis, M. C., Reich, J. W., Nicassario, P., Tennen, H., Finan, P., …& Irwin, M. R. (2008). Comparison of cognitive behavioral and mindfulness meditation interventions on adaptation to rheumatoid arthritis for patients with and without history of recurrent depression. *Journal of Consulting and Clinical Psychology, 76*, 408–421. doi:10.1037/0022-006X.76.3.408

Zinzow, H. M., Brooks, J. O., Rosopa, P. J., Jeffirs, S., Jenkins, C., Seeanner, J., ...& Hodges, L. F. (2017). Virtual reality and cognitive-behavioral therapy for driving anxiety and aggression in veterans: A pilot study. *Cognitive and Behavioral Practice*. doi:10.1016/j.cbpra.2017.09.002