Methodological Decision-Making in Evaluation: An Examination of Implications of Using a Stage of Change Outcome Variable

Author: Julia Rollison

Persistent link: http://hdl.handle.net/2345/bc-ir:103736

This work is posted on eScholarship@BC, Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2014

Copyright is held by the author, with all rights reserved, unless otherwise noted.
Abstract

METHODOLOGICAL DECISION-MAKING IN EVALUATION: AN EXAMINATION OF IMPLICATIONS OF USING A STAGE OF CHANGE OUTCOME VARIABLE

Julia Rollison, Author
Lauren Saenz, Chair

Methods for classifying and treating an outcome variable are critical to explore in health research and evaluation, given the potential impact the choice of method may have on the findings and subsequent recommendations (Merbitz, Morris, & Grip, 1989). Further, given the prominent application of the Transtheoretical Model in health research, the stages of change construct continues to be a critical outcome measure concept used in various applied evaluation studies (Bridle et al., 2005; Nigg, 2002; Prochaska, DiClemente, & Norcross, 1992). The purpose of this dissertation study was to determine if findings differ depending on the approach to categorizing and analyzing a stage of change outcome variable, and if so, to highlight how these may affect policy and programmatic decision-making.

Using data from a study on evidence-based program adoption decisions, this dissertation examined five approaches to treating and analyzing a single Decision to Adopt outcome variable. These different approaches were compared from both a methodological and pragmatic perspective. Hypothetical stakeholder illustrations were used to highlight differences in decision-making priorities and use of findings based on role, background, and organizational priorities.

In comparing methods for classifying and treating the stage of change outcome variable, findings revealed notable differences in effect size, estimation, implication of
major findings, and limitations of approach. The hypothetical stakeholder illustrations stressed the significance of personal values and preferences as key influential factors in decision-making and use of evaluation results.

This dissertation highlighted how decisions are inextricably linked to the logic model and underlying theory, particularly as it relates to defining evaluation questions, determining how to categorize constructs, and assigning value to codes. Further, it reinforces the significance of contextual considerations in evaluation and how these cannot be ignored in the decision-making process (e.g., budgetary limitations, practical constraints, political factors). The proposed directions for future research seek to continue advancing this understanding of the impact of methodological decisions in different contexts and help improve the utility of evaluations more broadly.
Acknowledgements

I would like to express deep appreciation to my dissertation committee including Dr. Lauren Saenz, my committee chair, and readers Dr. Larry Ludlow, Dr. June Horowitz, and Dr. Jessica Williams. Your input and feedback has been incredibly valuable throughout this process and I have truly enjoyed working with each of you. Thank you so much for your time and insights.

I would like to thank Dr. Gary Hill, an amazing mentor and even better friend. I would not be where I am today without you – both professionally and personally.

Finally, I would like to thank my family and friends, especially Mom, Jay, and Jeremy. It has been a tough few years but your continued interest, support, and patience have been invaluable. I’m looking forward to spending my newfound free time with you!
Dedication

I would like to dedicate this dissertation to my father, the late Dermot Rollison.
Personal Placement

As a former employee of the firm conducting the original research study, I conducted analyses on the Readiness to Implement Motivational Interviewing scale, a component of the initial survey instrument, as well as supported the development of informational packets of select evidence-based practices. Given my familiarity with the study and the research team members, I felt this could be a trusted source of data to use for my dissertation. In addition, the study’s principal investigator, Dr. Jessica Williams, agreed to serve as a committee member for this dissertation which provided advantages from both a historical and subject-matter expert perspective. Finally, the ability to conduct research embedded in an emerging field such as Dissemination and Implementation research combined with investigating linkages with utilization-focused evaluation is interesting to me both personally and professionally.
# Table of Contents

Abstract ........................................................................................................................................... i
Acknowledgements .......................................................................................................................... iii
Dedication .......................................................................................................................................... iv
Personal Placement .......................................................................................................................... v
List of Figures .................................................................................................................................... viii
List of Tables .................................................................................................................................... ix

**Chapter One: Introduction** .......................................................................................................... 1
   Introduction ..................................................................................................................................... 1
   Statement of the Problem ............................................................................................................ 2
   Placement in the Field .................................................................................................................. 4
   Constructs and Variables of Interest ........................................................................................... 5
   Research Questions ..................................................................................................................... 6
   Significance of the Research ....................................................................................................... 7
   Limitations of the Study .............................................................................................................. 12
   Conclusion .................................................................................................................................. 13

**Chapter Two: Review of the Literature** ..................................................................................... 14
   Research on Dissemination of Evidence-based Practices ......................................................... 14
   Transtheoretical Model and Stages of Change ......................................................................... 16
      *Stages of Change* .................................................................................................................... 20
      *Other Transtheoretical Model Elements* ............................................................................. 24
   Decision to Adopt ....................................................................................................................... 26
   Levels of Measurement .............................................................................................................. 30
   Ordinal and Interval Data Controversy ...................................................................................... 32
   Measuring Stages of Change ...................................................................................................... 36
      *Algorithm Approach* ............................................................................................................ 36
      *Scale Approach* .................................................................................................................... 41
      *Comparison of Approaches* .................................................................................................. 44
   Stage of Change Variable Treatment ....................................................................................... 46
   Analytical Considerations .......................................................................................................... 51
   Factors Affecting Methodological Decision-Making ................................................................. 56
   Conclusion .................................................................................................................................. 61

**Chapter Three: Methods** .......................................................................................................... 62
   Sample .......................................................................................................................................... 62
   Variables Measured and Instruments Used .............................................................................. 63
   Data Collection Procedures ....................................................................................................... 65
   Data Treatment ............................................................................................................................. 66
   Analyses ........................................................................................................................................ 67
      *Research Question 1* ................................................................................................................. 69
<table>
<thead>
<tr>
<th>Research Question</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 2</td>
<td>69</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>70</td>
</tr>
<tr>
<td>Research Question 4</td>
<td>71</td>
</tr>
<tr>
<td>Research Question 5</td>
<td>71</td>
</tr>
<tr>
<td>Research Question 6</td>
<td>72</td>
</tr>
</tbody>
</table>

Conclusion: 79

Chapter Four: Results 80

Sample Description 80

Research Question 1 80

Research Question 3 84

Research Question 4 86

Research Question 5 87

Research Question 6 90

Methodological Factors 92

Pragmatic Factors 98

Conclusion 101

Chapter Five: Discussion 102

Implications for Stage of Change Variable Treatment and Analysis 102

Summary 105

Methodological Decision-Making Considerations 106

Questions to Consider 113

Chapter Six: Practical Application 115

Defining Program Logic 115

Study Design, Measurement and Analytic Approaches 117

Interpreting and Using Evaluation Findings 118

Summary 124

Chapter Seven: Conclusion and Future Research 125

Conclusion 125

Practical Recommendations 127

Limitations of the Study 129

Directions for Future Research 131

References 133
List of Figures

Figure 2.1. Stages of Change with Evidence-based Practice Example

Figure 3.1. Illustration of Research Question Variable Treatments

Figure 4.1. SPSS Output for Research Question 1

Figure 4.2. SPSS Output for Research Question 2

Figure 4.3. SPSS Output for Research Question 3

Figure 4.4. SPSS Output for Research Question 4

Figure 4.5. SPSS Output for Research Question 5

Figure 4.6. Visual Comparison of Cases Within Groupings
List of Tables

Table 3.1. Response Options and Corresponding Stages of Change
Table 3.2. Research Question 6 Discussion Comparison Format
Table 4.1. Sample Sizes for Group Condition
Table 4.2. Distribution Based on Posttest Score (Research Question 1)
Table 4.3. Distribution Based on Number of Stages (Research Question 2)
Table 4.4. Distribution Based on Dichotomized Action (Research Question 3)
Table 4.5. Distribution Based on Dichotomized Progress (Research Question 4)
Table 4.6. Distribution Based on Ordered Categorical Variable (Research Question 5)
Table 4.7. Parameters from the Ordinal Regression of Group Condition on the Outcome Variable
Table 4.8. Research Question 6 Comparison
Table 4.9. Comparison of Methodological Factors
Table 5.1. Sample Considerations for Methodological Decision-making for a Stage of Change Outcome Variable
CHAPTER ONE: INTRODUCTION

This chapter presents the purpose of this dissertation, specific research questions, and an overview of the study’s placement in the health field, particularly in dissemination and implementation (D&I) research. It also highlights data gaps such as the need for more research at the organizational level and improved measurement and subsequent analysis of findings.

Introduction

This dissertation seeks to examine and better understand the impact of decisions about how to treat outcome variables. This study uses secondary data from a completed research project where the original study team aimed to improve knowledge around how comparative effectiveness research results can be effectively packaged, disseminated, and adopted by community health organizations and their providers (Williams, Williams, et al., 2013). The prospective part of the research study examined decision-making factors that influence an organization’s intent to adopt an evidence-based program.

In the original study, the research team treated the outcome (i.e., the decision to adopt an evidence-based program) as a continuous, interval-level variable and used the group mean change score from baseline to follow-up. The team found that the scores for Decision to Adopt the evidence-based program increased for all organizations after the baseline assessment regardless of the intervention condition although the Intervention group experienced a significantly greater increase than the Comparison group (Williams, Williams, et al., 2013).
If the research team opted for an alternative approach such as treating the *Decision to Adopt* outcome as a dichotomous “action” variable, thus using logistic regression approaches, or as strictly ordinal with associated non-parametric procedures, would the results have differed? This dissertation seeks to capitalize on the availability of data from community-based health organization staff to address this specific gap in knowledge. Specifically, research questions focus on if, and how, different treatments of the outcome variable will affect the study findings. The *Decision to Adopt* outcome variable, based on the stages of change construct in the Transtheoretical Model (Prochaska et al., 1992), is characterized using five different approaches and appropriate analyses were performed based on each approach. The results from each analysis were then compared on various methodological and pragmatic factors.

**Statement of the Problem**

Methods for classifying and treating an outcome variable are critical to explore in health research given the potential impact the choice of method may have on the findings and subsequent recommendations. Stakeholders, particularly policy makers, use research findings to shape future thinking and practice even in cases where the methodological approach may be questionable (Merbitz et al., 1989). As more people seek to answer critical research questions that can influence organizational and federal programs and policies, a balance between rigor and practicality is needed to ensure findings are useful and defensible for intended audiences.
This issue is particularly evident with D&I research. In a recent funding opportunity announcement through the National Institutes for Health (NIH), the purpose of D&I research is detailed:

D&I research intends to bridge the gap between public health, clinical research, and everyday practice by building a knowledge base about how health information, interventions, and new clinical practices and policies are transmitted and translated for public health and health care service use in specific settings (NIH, 2013, “Research Terms,” para. 1).

Understanding decisions to adopt evidence-based programs and practices among health care organizations is of utmost importance for translating research to practice. This translation optimizes the benefits of a specific evidence-based intervention ultimately improving patients’ health outcomes (Damschroder et al., 2009).

The purpose of this dissertation is to determine whether findings differ depending on the approach to categorizing and analyzing a stage of change outcome variable, and if so, to highlight how these can affect policy and programmatic decision-making. This dissertation illustrates five approaches to treating and analyzing a single outcome variable using data from a study on evidence-based program adoption decisions among community-based health organizational staff (Williams, Blais, et al., 2013). Second, through a comparative analysis of the five approaches, implications of the choice of approach in terms of significance, magnitude, and relevance of findings are examined and discussed. Finally, the potential use of information by policy makers and other stakeholders is presented in light of the relative rigor and practicality of each approach.
Placement in the Field

The goal of the health care system in the United States is to provide high quality care to all patients and communities (Burns & Grove, 2009). A key component of ensuring quality care is to base practices on the best evidence available from research. Evidence-based practice is defined as “the conscientious integration of best research evidence with clinical expertise and patient values and needs” (Burns & Grove, 2009, p.617). Evidence-based decision-making in public health is viewed as using the best available scientific evidence regarding a program or practice and transferring that evidence into practice while taking into account local resources and goals (Jacobs et al., 2012). This evidence can be quantitative or qualitative in nature and often varies in terms of the perceived value of the evidence (Satterfield et al., 2009).

Effective interventions have been identified in the research setting but effectiveness in practice often changes when “real world” conditions are introduced. This evidence-based program “translation” process from research settings into real-world settings is long and complex (Durlak & DuPre, 2008). At a high level, there are four main phases of diffusion into a real-world setting including dissemination, adoption, implementation, and sustainability (Durlak & DuPre, 2008). Without being adopted and implemented with fidelity in the field, the time and resources that went into developing an intervention may be viewed as wasted (Simpson & Flynn, 2007). Simpson and Flynn (2007) refer to this as the balance between “evidence-based” and “practice.”

Implementation science, also known as translation research, focuses on the effective adoption of evidence-based programs and practices into an organization (Schaffer, Sandau, & Diedrick, 2012). According to the open access journal,
Implementation Science, implementation research is defined as “the scientific study of methods to promote the systematic uptake of proven clinical treatments, practices, organizational, and management interventions into routine practice, and hence to improve health. In this context, it includes the study of influences on patient, healthcare professional, and organizational behavior in either healthcare or population settings” (Implementation Science, 2014, para. 3). This research is intended to address practical questions policy makers and other decision-makers may have around programs such as degree of effectiveness (Glasgow, 2009). Even though theory is important, given the practical nature of this field, the importance of context in conducting research is critical (Glasgow, 2009). This research spans disciplines including the biomedical, social science, and management fields and features an array of different research methods and approaches (Implementation Science, 2014). Implementation research stresses the importance of health service professionals and organizations not only as stakeholders but as sources of variance that require empirical examination (Implementation Science, 2014).

Constructs and Variables of Interest

The main construct of interest for this study is the stages of change construct from the Transtheoretical Model (Prochaska et al., 1992) explained in more detail in Chapter Two. This construct consists of five stages which represent organizational readiness in terms of a decision to adopt an evidence-based program. This is measured by a single item in a survey of community-based health organization staff asking study participants to “Please indicate your level of interest in adopting MI into your program.” The item,
adapted from the work of McGovern, Fox, Xie, and Drake (2004), assesses participants’ interest in adopting motivational interviewing, a counseling approach that attempts to increase the patient or consumer’s awareness of potential problems of the behavior in question (Williams, Blais, et al., 2013). Each response option corresponds to a stage of change in the Transtheoretical Model (i.e., Precontemplation, Contemplation, Preparation, Action, Maintenance) (Prochaska et al., 1992). A more comprehensive discussion of this construct and the underlying theoretical model is presented in Chapter Two.

One key variable of interest is the study group assignment from the original study on dissemination strategies for an evidence-based program. The Treatment group in this study will be those who were exposed to Webinars in addition to an informational packet while those in the Comparison group were provided with an informational packet only. Additional details on the variables of interest are discussed in Chapter Three.

Research Questions

This dissertation seeks to expand on work conducted by the MANILA Consulting Group, Inc. (MANILA) study team on the outcome variable of interest: Decision to Adopt. The decision on how to treat a variable such as this can have a significant impact on the conclusions drawn from the study.

The following research questions are of primary interest in this study:

1) Are there posttest differences in the outcome variable, as measured by posttest Decision to Adopt score, between the Intervention and Comparison groups?
2) Are there posttest differences in the outcome variable, as measured by stage progression (number of stages), between the Intervention and Comparison groups?

3) Are there posttest differences in the outcome variable, as measured by presence of action (dichotomous of yes/no), between the Intervention and Comparison groups?

4) Are there posttest differences in the outcome variable, as measured by presence of stage progression (dichotomous of yes/no), between the Intervention and Comparison groups?

5) Are there posttest differences in the outcome variable, as measured by the stage of change (ordered categorical), between the Intervention and Comparison groups?

6) Do the findings vary across approaches? If so, how do the findings differ?

The first research question aligns with the approach used in the original study (i.e., posttest score). A series of analyses were conducted to address these research questions in turn, taking into account the effects of the experimental group on the outcome variable assessing the decision to adopt an evidence-based program.

Significance of the Research

Federal agencies, such as NIH, stress evaluation efforts of research in “real world” settings given the challenges inherent in implementing evidence-based practices with minimal resources (Brownson, Allen, Duggan, Stamatakis, & Erwin, 2012; Miller, Krusky, Franzen, Cochran, & Zimmerman, 2012). These research efforts should evaluate both health outcomes and process outcomes to optimize the benefits of a specific
intervention and increase the likelihood for the sustainability of the intervention in a specific context (Damschroder et al., 2009). Community-based health organizations are one such setting. In particular, these organizations can benefit from evidence-based practices from improved client outcomes and staff capabilities (Miller et al., 2012). In addition, demonstrating results through evidence-based practice implementation can help secure additional federal and local funding (Miller et al., 2012).

One review of implementation research noted that two-thirds of efforts by organizations to implement change fail (Damschroder et al., 2009). Common reasons for failure include lack of support for the given intervention, poor communication, and no organizational incentives (Damschroder et al., 2009). Given the amount of resources, including time, involved in implementing change, additional knowledge and understanding of organizational decisions around adoption of evidence-based programs is needed (Weiner, Amick, & Lee, 2008).

Community-based organizations are unique health care settings as they serve as a resource center in the community but often have fewer resources and access to expertise than other health care settings (Miller et al., 2012). To ensure evidence-based practices are more accessible and useful for community-based organizations, evaluation efforts for D&I research are critical (Miller et al., 2012; Simpson & Flynn, 2007). Glanz and Bishop (2010) identified a need to increase the use of theoretical frameworks for research at the organizational level to better inform the D&I field.

Some researchers have referenced the importance of developing and improving measures specific to the implementation process (Simpson & Flynn, 2007). Measures, especially those used in conjunction with existing behavioral health models, need to be
useful, testable, and portable for health providers, researchers, and policy makers in real world care settings (Bunton, Baldwin, Flynn, & Whitelaw, 2000; Proctor et al., 2011).

As Proctor et al. (2011) have noted, “a critical yet unresolved issue in the field of implementation science is how to conceptualize and evaluate success” (p. 65). Practical tools for assessing adoption decisions are needed to contribute to D&I research at the organizational level to ultimately improve health outcomes (Weiner et al., 2008). An instrument or tool that demonstrates reliability and validity among community-based health organizations would help advance testing of theories such as the Transtheoretical Model for organizational change (Weiner et al., 2008). Policy makers would also support such efforts that combine experience in the field with methods grounded in research (Simpson & Flynn, 2007). For managers and staff and community-based health organizations, instruments (or items) could also be used as diagnostic tools as part of a needs assessment or planning effort (Weiner et al., 2008). If the organizational representatives are reporting contemplating change, additional capacity building efforts may need to be provided to help advance the organization towards action (Weiner et al., 2008).

In addition, a consistent method of measuring and categorizing change can facilitate collaboration across organizations and increase learning opportunities to advance evidence-based implementation efforts on a larger scale (Simpson & Flynn, 2007). The method used to categorize and subsequently analyze a stage of change variable is critical to explore in the health care field given the potential impact the choice may have on the findings and subsequent recommendations. This is particularly important with the weight individuals often give to conclusions from research studies,
even if the methodological approach may be questionable, which ultimately shapes future thinking and practice (Merbitz et al., 1989). Although a significant amount of work has been done with respect to substance use or other health conditions and the use of stages of change theory, organizational research utilizing stages of change is more limited. By engaging staff and managers in the research planning process, the measures not only advance in terms of utility for those ultimately implementing the intervention but also can contribute more credible evidence to the implementation science field (Simpson & Flynn, 2007).

The need for more research to provide connections between policy, research and evaluation processes is also critical (Head, 2010). As noted above, researcher decisions for measurement, variable treatment, and analysis approach can result in different outcomes and thus different interpretations of results. To further complicate matters, the same results could be interpreted differently depending on the stakeholder. Without understanding perceptions of outcomes and preferences for measuring those outcomes among health services stakeholders, study findings can be constrained or even misused (Proctor et al., 2011).

Documenting whether or not the interpretation of results from the same dataset can differ depending on the method of characterization or manipulation will provide insight into the research and measurement field more broadly and help stimulate use of findings (Patton, 2013). This may include decisions around allocation of dissemination-related funding (including potential targeting of funding to organizations at a specific stage of change deemed more “worthy of dissemination”) or policy guidance around translation strategies (Rahman, Applebaum, Schnelle, & Simmons, 2012).
The decision to adopt an evidence-based program is expected to have long-term effects on implementation process and patient outcomes within an agency. Decisions can even extend to other organizations within larger systems (Panzano & Roth, 2006). Focusing on measurement issues such as this in the short-term provides a practical approach to facilitate faster advancements in the dissemination field toward a larger goal (Rahman et al., 2012).

Given the importance of adopting evidence-based practices, and the amount of time and resources needed to adopt a new program, more research concerning appropriate measurement and treatment of variables is needed among behavioral health and community health organizations. Using data from a Substance Abuse and Mental Health Services Administration (SAMHSA)-funded study to explore the effect of categorizations of stages of change, this dissertation’s comparison of approaches is valuable given the possible statistical and theoretical consequences of variable treatment and subsequent analysis (Kahler, Rogausch, Brunner, & Himmel, 2008).

This documentation is facilitated through a comparison of approaches such as the comparison proposed in the analysis plan for Research Question 6. This appraisal of both methodological and pragmatic factors helps the ultimate intended users understand how credibility and utility may be affected depending on the decision by keeping the discussion practical rather than too technical (Patton, 2013). This approach will be strengthened if stakeholders at various levels actively participate in this process and subsequent discussion which many are incentivized to do, especially with organizational or funding mandates to incorporate evidence-based practices (Donaldson, Rutledge, & Ashley, 2004; Rahman et al., 2012). Sharing and discussing findings with those
ultimately involved in the implementation process is important for making these connections. Particularly with government sponsored research, like in this SAMHSA-funded study, the need for more pragmatic considerations is evident (Head, 2010).

Both researchers and policy- and provider-level stakeholders need to work together to improve utilization of findings. Researchers need to better understand and take into account pragmatic considerations based on contextual factors and groups like policy managers should increase their awareness and understanding of research and resultant outcomes, including strengths and limitations (Head, 2010). This collaboration should lead to improved research in the field on adoption decisions and, ultimately, improved health-related outcomes based on a more effective implementation process.

Limitations of the Study

This dissertation has several limitations to take into account when considering and interpreting study findings, particularly as it relates to generalizing lessons learned more broadly. First, the study takes advantage of data from an existing study so limitations within that context and existing dataset needed to be accounted for. These included factors such as measurement decisions on how information on the stage of change would be collected from respondents (i.e., a single item to measure stage of change) and the need for data augmentation procedures in the original dataset. More broadly, because the study focused on decision to adopt an evidence-based program, these findings may not be as applicable to decisions around other constructs for stage of change variable treatment considerations. These limitations are discussed in more detail in Chapters Three, Four and Seven. However, since this dissertation focuses more on variable treatment and use
of findings in “real-world” situations, these limitations due to using an existing dataset are considered minor.

Broader study limitations that affect the generalizability of findings are more fully discussed in Chapter Seven. These include study design decisions such as excluding contextual variables from the analyses and using hypothetical stakeholder illustrations to discuss variable treatment decisions and potential use of findings rather than the use of qualitative interviews.

Conclusion

This dissertation provides valuable insight into the potential impact of outcome variable treatment decisions on study findings and how such decisions may affect research utilization with examples provided based in the D&I field. Chapter Two presents literature to situate this dissertation in terms of D&I research and measurement issues. It also elaborates on the theoretical basis for the outcome of interest: the decision to adopt an evidence-based program. Chapter Three presents an overview of methods while Chapter Four includes results for the six research questions. Chapters Five through Seven discuss the findings in more detail and present implications, illustrations of implications, limitations, and suggestions for future research.
CHAPTER TWO: REVIEW OF THE LITERATURE

This chapter provides background on the use of evidence-based practices and gaps in D&I research. It also presents needed context for the theoretical model underlying the dissertation’s main construct of interest, stages of change. The levels of measurement and controversies around ordinal versus interval data are presented with stages of change examples from the literature. Finally, this chapter discusses factors that affect methodological decision making and potential implications of these decisions.

Research on Dissemination of Evidence-based Practices

The use of evidence-based practices in clinical settings, both for primary care and behavioral health services, has been strongly encouraged by local and national stakeholders to improve client outcomes (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). This focus began for medicine in the early 1990s and nursing followed in the early 2000s (Burns & Grove, 2009). Since then, implementing evidence-based practices has led to improved outcomes at the patient and provider-level (Burns & Grove, 2009). For this reason, most health care agencies support the use of evidence-based practices given the focus on quality yet cost-effectiveness of care. In fact, implementing and adhering to evidence-based practices are identified in the Joint Commission Hospital National Patient Safety Goals as part of their accreditation program (The Joint Commission, 2013). The Joint Commission provides standards to help health care organizations measure and improve performance needed to deliver high quality care; it is an important group driving improvements to accountability around quality measures (Chassin, Loeb, Schmaltz, &
Wachter, 2010; The Joint Commission, 2013). In addition, in order for hospitals to obtain a “Magnet” designation, a standard of excellence in nursing, staff must implement evidence-based practices as a demonstration of quality care (Wise, 2009).

However, a criticism of evidence-based practice is that there is a lack of evidence on implementation of evidence-based practices (Burns & Grove, 2009). There are of course barriers to implementing evidence-based practices. Successful implementation is reliant on support and action from all relevant stakeholders including administrators and health professionals (Burns & Grove, 2009). Dissemination strategies, such as the use of Webinars, also require their own evidence-base in terms of effectiveness and feasibility to ensure they have a long-term impact to support adoption and implementation of innovative programs (Cookston, Sandler, Braver, & Genalo, 2007; Rahman et al., 2012).

Measuring and observing changes around evidence-based practice dissemination and adoption are critical to organizations and staff as well as the D&I field (Donaldson et al., 2004). There has been a good amount of research on the importance of organizational readiness for change and factors associated with increased readiness yet measure development efforts at the organizational level are more limited (Weiner et al., 2008). Weiner et al. (2008) suggest this may be due to health service researchers investigating different types of organizational change with various theoretical frames of reference which can limit collaboration across efforts. Measuring evidence-based practice dissemination and adoption outcomes is even less common (Donaldson et al., 2004). Donaldson et al. (2004) found that less than 15 percent of health care research organizations are evaluating outcomes as part of their D&I research (Donaldson et al., 2004). Out of those measuring outcomes, measuring organizational-level outcomes was
even less frequent focusing instead on provider- and patient-level outcomes (Dykes, 2003).

D&I research, has a range of stakeholders beyond administrators and clinical staff. Researchers, funders, professional associations, national and state coalitions, and improvement advocates have been engaged in research around the most appropriate, effective, evidence-based approaches to translate research into practice – and this often varies depending on the stakeholder (Rahman et al., 2012). For example, funders are often looking for the most fiscally practical investment while clinical staff may be more focused on the most improved patient outcomes regardless of cost (Rahman et al., 2012). Time is another practical consideration as innovation adoption is a staged process taking months to years to achieve (Rahman et al., 2012).

### Transtheoretical Model and Stages of Change

Research in the health care field is primarily driven by theory testing (i.e., deductive reasoning) or theory development (i.e., inductive reasoning). Theories related to health behavior have been developed and tested extensively in numerous research studies over the past several decades. As evident in the name, the Transtheoretical Model, developed by James Prochaska, Carlo DiClemente, and John Norcross in the early 1990s, borrows from, and cuts across, several theories of health behavior (Brannon & Feist, 2004; Prochaska et al., 1992). Drawing from different theories like the Health Belief Model and the Theory of Reasoned Action, the Transtheoretical Model accommodates and transcends these other models (Whitelaw, Baldwin, Bunton, & Flynn, 2000). The model stresses that behavior change is closely linked with personal values and the
individual needs to perceive that any benefits of change are relevant to their circumstances (Arnold, Hess & Lipner, 2013).

The central, organizing component of the Transtheoretical Model is the use of stages to segment individuals or groups on their readiness to change (Santiago-Rivas, Velicer, Redding, Prochaska, & Paiva, 2013; West, Cafferty, & Ledford, 2013). Stage models have been used to better understand readiness to undertake a range of health behavior changes, particularly in intervention research (Burns & Grove, 2009). In general, stage theories describe behavior change in terms of a set of qualitatively different stages as opposed to a continuous process (Bridle et al., 2005). Bunton et al. (2000) have argued that stage theories are at best descriptive in nature as categorizing individuals into a stage provides no explanation around actual behavior. However, stage theories do provide pragmatic value in the public health field (Bunton et al., 2000).

The number of stages varies across the different models but generally all stages contain categories for those individuals who have not yet decided to change their behavior; those who have; and those who have engaged in change (Bridle et al., 2005). There is some disagreement in the literature about whether or not the stages are, in fact, qualitatively different, and some view distinctions between stages as arbitrary (Bridle et al., 2005). Littell and Girvin (2002) noted evidence that at one point in time, an individual can be involved in more than one (even non-adjacent) stage although this proposition contradicts the idea of a stage-based model of change and questions the explanatory value of the Transtheoretical Model (Bridle et al., 2005; Bunton et al., 2000).

The most popular stage model, and one with extensive research support, is the Transtheoretical Model (Horwath, Schembre, Motl, Dishman, & Nigg, 2013). This model
is used most often to describe behavioral change related to problems such as substance abuse, nutritional decisions, and exercise (Brannon & Feist, 2004; Prochaska, Redding, & Evers, 2002; West et al., 2013). This model has been influencing service planning and implementation at local and federal levels since 1985 in the United States and Europe (Bunton et al., 2000). The Transtheoretical Model has five stages of change through which an individual works in changing a behavior (see Figure 1). The model has undergone a series of changes since it was first introduced; for example, the original model for stages of change had a sixth stage, Termination, following Maintenance, but the current model of five stages of change is the most widely used (Kim et al., 2012; Sutton, 2001).
Figure 1. Stages of Change with Evidence-Based Practice Example. Adapted from Health Psychology, Fifth Edition (p. 76) by L. Brannon & J. Feist, 2004, Belmont: Wadsworth/Thomson Learning.

- **Precontemplation**
  - I am not familiar with MI; *or*
  - I am not interested and do not think this practice would be effective in my program

- **Contemplation**
  - I have considered the practice but see many pros and cons

- **Preparation**
  - I am leaning in the direction of adopting the practice in my program

- **Action**
  - I have just begun to implement the practice in my work

- **Maintenance**
  - I have been using the evidence-based practice and efforts are in place to maintain it
Stages of Change

The first stage is called Precontemplation. Looking at this stage in terms of decision to adopt an evidence-based practice, someone who is in this stage is not interested in or is unaware of the practice and does not have any intention of adopting the practice (Prochaska et al., 2002). The person at this stage may not be taking action because the individual is unaware of the significance of their actions or they are unmotivated to change.

The next stage of the model is Contemplation. In this stage, the individual intends to take action within the next 6 months (Prochaska et al., 2002). This stage is often characterized by weighing the pros and cons of the behavior and the intended consequences of that behavior. For example, someone who is thinking about adopting the evidence-based practice may weigh pros such as more effectively treating clients with cons such as the amount of time needed to devote to training or the costs of implementation. An individual can stay in this stage for a relatively long period of time because this weighing often produces ambivalence towards the behavior and change is delayed (Prochaska et al., 2002).

The third stage of the model, Preparation, is the point where the individual intends to take action and has taken some sort of behavioral step in the direction of taking action. For example, if the individual has weighed the pros and cons and is leaning towards adopting the evidence-based practice, he or she may take the action of going out and talking to other providers who have also implemented the program. While he or she has not yet engaged in the action, he or she has taken some sort of behavioral step to act
(Prochaska et al., 2002). This stage is often equated with “readiness to change” in the organizational health services literature (Weiner et al., 2008).

The fourth stage of the model, Action (also referred to as Implementation), is where the individual finally changes his or her overt behavior (e.g., begins implementing the evidence-based practice), usually for less than 6 months (Prochaska et al., 2002). This may be on a smaller scale such as training a handful of providers on the practice rather than the entire staff. The fifth stage, Maintenance, is where the individual maintains his or her behavior change for more than 6 months (Prochaska et al., 2002). For decisions to adopt, this stage would involve full implementation and include efforts to maintain the practice. Many people have difficulty reaching this stage due to various factors such as cost, time constraints, competing priorities, and a lack of support from leadership (McGovern et al., 2004).

The Precontemplation and Contemplation stages reflect differences in intention while the Preparation through Maintenance stages reflect differences in the duration (and action) of the behavior (Brug et al., 2005). Stages 1 through 3 are motivational in nature while the other two focus on action (Bridle et al., 2005). Certain stages may be more resistant to change due to their stable nature (i.e., Precontemplation and Maintenance) (Russell, Maher, Prochaska, & Johnson, 2012). Preparation and Action stages are considered the most open to change (Russell et al., 2012).

The stages of change construct captures the temporal nature of the change process (Prochaska et al., 2002). However, some argue that the thresholds between stages are arbitrary and the time component distinguishing some stages creates pseudostages rather than true stages and (Brug et al., 2005; Sutton, 2001). For example, if an individual
abstains from alcohol use beyond 6 months, he or she would be categorized as being in the Maintenance stage while another individual abstaining from alcohol use for 5 months and 3 weeks would be classified as in the Action stage. The same may be said for earlier stages which some view as a continuum of “planned time to action” (Sutton, 2001). Sutton (2001) argues that this creates arbitrary segments and that matching interventions to stage is not warranted.

The progression between stages is cyclical rather than linear with individuals often progressing and regressing through the model (Brannon & Feist, 2004; Noordman, De Vet, van Weijden, & van Dulmen, 2013). Individuals may regress across multiple stages and it is not uncommon for people to cycle through the model several times before completing the behavioral change (Brannon & Feist, 2004). Other theories have also addressed this change process for cutting across different circumstances, such as the Precaution Adoption Process Model, which specifies cognitive stages of readiness and action decisions (Glanz & Bishop, 2010).

Because factors aiding or impeding change are assumed to be different at each stage, interventions adapted to the specific stage of change are expected to lead to successful progression (Bridle et al., 2005; Noordman et al., 2013; West et al., 2013). For example, if someone is in the Precontemplation stage, an intervention may need to focus more on identifying a problem whereas in the Preparation stage, an intervention may be focused on specific suggestions for how to change (Brannon & Feist, 2004). In addition to targeting interventions to specific stages, some other contextual variables may be associated with particular stages – such as females being more likely to be in a given stage for a specified behavior – which is touted as an additional strength of stage-based
models (Weinstein, Lyon, Sandman, & Cutie, 1998). Research has also suggested that the further along an individual is in the stage sequence at baseline, the more likely they are to advance to subsequent stages (Russell et al., 2012).

The stage of change construct has been tested in a number of studies examining behavior change in areas such as alcohol abuse, smoking cessation, psychological distress, and safe sex practices (Prochaska et al., 2005). The “stages” originated in the early 1980s and have been the focus of books and hundreds of articles and empirical studies (Littell & Girvin, 2002). Although useful, research testing the Transtheoretical Model indicates that the model may not be appropriate for understanding all health behaviors. A meta-analysis of studies utilizing the Transtheoretical Model showed that the model works better with some behaviors, like smoking cessation, than with other behaviors, and that it should not be used as a blueprint for all interventions (Rossi in Brannon & Feist, 2004).

Given the Transtheoretical Model’s reliance on readiness for change depending on stage, it has historically been used to create tailored messages (Lustria, Cortese, Noar, & Glueckauf, 2009). Data have shown that individuals often rely on different processes of change depending on the stage they are in. For example, those who are in the earlier, non-action stages often rely on more cognitive or evaluative processes of change while those in the latter, action stages rely more on social support and behavior management techniques (Prochaska et al., 2005).

A decade ago, minimal research examined stages of change for organizations (Berry, Plotnikoff, Raine, Anderson, & Naylor, 2003). More recently, the stage of change construct has been used by organizations and communities to assess readiness for change.
for programs, practices, or policies (Glanz & Bishop, 2010). More specifically, the Transtheoretical Model has been used to better understand evidence-based practice adoption decisions in organizational settings (McGovern et al., 2004; Fleischer & Christie, 2009). Recommendations are emerging around the stage of change and appropriate interventions based on the identified stage (e.g., providing content to increase confidence around innovation adoption for an organization in the Precontemplation stage, disseminating information on adopting best practices amid difficult situations for those in the Action stage) (Russell et al., 2012; Fleischer & Christie, 2009). This approach to matching interventions to staff readiness is expected to reduce staff resistance to change as well as time needed to implement new programs (Berry et al., 2003).

Research has also been conducted on factors influencing progression through stages for decisions to adopt (Simpson & Flynn, 2007). In their working model on adoption decisions, Simpson and Flynn (2007) found that staff knowledge of the evidence-based program and organizational functioning perceptions both influenced stage progression. In a study of health workers by Levesque, Prochaska, Cummins, Terrell, and Miranda (2001), differences in organization role were found. Specifically, administrators were more likely to be in the Maintenance stage than direct care providers while providers were more likely to be in the Precontemplation stage (Levesque et al., 2001).

Other Transtheoretical Model Elements

While the stages are often emphasized the most in this model, there are other elements that are critical to the model. One important aspect of this model includes its flexibility in allowing individuals to move between stages. There are certain activities that people use to progress through each of the stages which are known as processes of
change. There are approximately ten processes that have been supported through research as critical in this model. Methods like self-reevaluation, self-liberation, helping relationships and stimulus control are all such processes (Prochaska et al., 2002). Some of these processes are behavioral in nature while others are cognitive.

Decisional balance, the weighing of pros and cons of a particular behavior, which takes place during the Contemplation and Preparation stage, is also a key component to the Transtheoretical Model. Research by Armitage, Povey and Arden (2003) examines decisional balance and self-efficacy and the role of attitudinal ambivalence. Attitudinal ambivalence is when the individuals hold both positive and negative feelings about a certain behavior (Armitage et al., 2003). This research is useful because many people often get stuck in certain stages and attitudinal ambivalence may play a significant role. In a study on attitudes towards health-related food choices, the greatest levels of ambivalence were shown in the Contemplation, Preparation, and Maintenance stages (Armitage et al., 2003). This concept of attitudinal ambivalence in these stages is useful because of its implications in the progression of individuals from stage to stage.

Another important aspect of the model is the notion of self-efficacy. Self-efficacy is essentially one’s confidence in his or her ability to perform a specific behavior. Action is most likely when levels of self-efficacy are high (Bandura, 2004). It follows then that it is common that Precontemplators report the least self-efficacy while Maintainers report the most self-efficacy (Armitage et al., 2003). However, many researchers have expressed concern around the lack of consensus regarding which variables, such as self-efficacy or decisional balance, predict forward movement between stages (Horwath et al., 2013).
These findings are useful when designing interventions because they allow the developer to tailor the intervention for the target population depending on their needs and characteristics. Many researchers have used the Transtheoretical Model in designing interventions for behavioral health changes (Burns & Grove, 2009). This model has many strengths, particularly the ability of the stages to be both open to change and stable, because this ability mirrors characteristics of behavioral risk factors (Prochaska et al., 2002). Another positive aspect of the model is its focus on those individuals who are not yet ready to act as many interventions are action-based (Prochaska et al., 2002). The Transtheoretical Model does not assume to be able to account for all the different aspects of behavioral change but there is yet to be a single theoretical model which does (Prochaska et al., 2002).

Decision to Adopt

Incorporating evidence-based practices into practice involves four major areas including 1) the identification of relevant and appropriate evidence-based practices, 2) acceptance and decision to adopt the evidence-based practice, 3) implementation of evidence-based practices, and 4) evaluating the usefulness of the evidence-based practice (Proctor, 2004). Once an evidence-based practice change is fully incorporated into an organization, it then becomes a standard of care (Titler, 2008). However, changing individual and organizational level practice is no small endeavor and each area includes intermediate outcomes that need to be examined (Titler, 2008). This is particularly true with decisions to adopt which can often be overlooked in the process (Taxman, 2012).
Adoption, also referred to as uptake, is often viewed as a process beginning with knowledge of the program or practice, followed by forming an attitude toward the program, and finally making a decision on whether to adopt (Proctor, 2004; Proctor et al., 2011). Adoption has also been defined as a continuum beginning with making a decision to change a practice and extending to completely incorporating the change so it is part of one’s routine (Schaffer et al., 2012). Some specify this change by using phrases such as “rate of adoption” or “extent of adoption” (Schaffer et al., 2012). For the purposes of this dissertation, the focus is on decision to adopt consistent with much of the organizational change literature (Simpson & Flynn, 2007).

Adoption and implementation of evidence-based programs and practices can be slow and research around how decisions to adopt are made by organizations is limited, particularly in community health care and behavioral health settings (Williams, Dusablon, Williams, Blais, & Hennessy, 2012; Cookston et al., 2007). This limited understanding of adoption decisions could be due to an assumption that many studies have made believing the adoption decision is unplanned (Taxman, 2012). Researchers in the health field have conducted additional research on evidence-based practice adoption decisions in recent years amid criticism that methods for assessing amenability to adopt new programs have not been well-validated and were lacking altogether for fields such as adolescent substance abuse treatment programs (Saldana, Chapman, Henggeler, & Rowland, 2007). The importance of stakeholders has also been stressed with organizational readiness literature focused on identifying, understanding, and engaging stakeholders involved in evidence-based practice adoption (Cookston et al., 2007).
The decision to adopt an evidence-based practice is complicated with multiple factors converging at various levels, including system, organizational, staff, and program (Aarons, Hurlburt, & Horwitz, 2011). Although the decision is often conceptualized as a fixed event, many organizations often experiment with innovations prior to deciding to implement (Aarons et al., 2011). For most organizations, this decision to adopt is not only strategic but also entails a level of risk given the balance of costs and benefits that need to be weighed by stakeholders (Panzano & Roth, 2006). Some believe that adopting evidence-based programs is a relatively low-risk decision because the benefits of such programs have been established in the research literature (Panzano & Roth, 2006). Yet, the decision to adopt is complex given political and financial considerations. Because of this, obtaining a better understanding of this decision and factors affecting this decision is of utmost importance to improve gaps between research and practice (Panzano & Roth, 2006).

Decisions made at the organizational level are quite complicated given the range of factors and variables affecting those decisions. This is evident in the research regarding decisions on adopting and implementing evidence-based practices. Factors such as leadership, communication, collaboration, goals and values, professional training, organizational culture and climate, and the size of an organization can all influence these decisions (Aarons et al., 2011; Fixsen et al., 2005; Meijers et al., 2006; Proctor, 2004; Satterfield et al., 2009; Taxman, 2012). For example, a counselor may have limited authority in his or her organization to change a care plan for a patient based on current research (Burns & Grove, 2009).
As noted above, some conceptualize the decision-making process regarding evidence-based practice adoption in organizations as a risk assessment, with the decision to adopt the program or practice expected to have an inverse relationship with the perceived risk in adopting the program or practice. This decision can be positively influenced by the organization’s capacity to manage risks and the degree of risk taking historically undertaken by the organization (Panzano et al., 2004). Another important factor in making decisions at the organizational level is the relative advantage of adopting the program or practice, including factors such as ease of implementation and having staff on hand able to handle the implementation requirements. This has been shown to differentiate adopters and non-adopters (Panzano et al., 2004). Organizations that have a climate of innovation (i.e., encouraging new ideas and technology) may be more likely to adopt evidence-based practices than those that do not (Aarons, Horowitz, Dlugosz, & Ehrhart, 2012). Of course, most organizations face barriers around funding which can limit evidence-based practice use and many are not rewarded by outside funders or other stakeholders for providing evidence-based care (Burns & Grove, 2009).

The role of the providers in the organization is critical to the adoption decision-making process given that these individuals usually have responsibility for implementing the new program or innovation which may require a change in practice (Bridges, Bierema, & Valentine, 2007). Even if the organization provides a supportive environment, individual differences in the propensity to adopt among staff could be a barrier to adoption and implementation (Taxman, 2012; Gallo & Barlow, 2012). One such difference could be in staff’s level of familiarity with the evidence-based program or practice, combined with their theoretical orientation. This has been shown to predict
adoption decisions as staff are more likely to adopt an evidence-based practice if it is similar to other practices they currently utilize (Gallo & Barlow, 2012). In cases such as this, staff need support from expert researchers and fellow health professionals to facilitate knowledge transfer (Burns & Grove, 2009).

Decisions to not adopt evidence-based practices by providers in otherwise supportive organizations include an inability to remain current on new developments including no access to research journals or other sources of research findings and evidence-based guidelines, perceived barriers accessing training, selective adoption of evidence-based practices, and misconceptions or preconceptions of evidence-based practices (Gallo & Barlow, 2012). Other studies have pointed to a lack of time as a major barrier to adopting an evidence-based practice given the amount of time needed to learn and implement the practice (Burns & Grove, 2009; Majid et al., 2011; Taxman, 2012).

Levels of Measurement

Brownson et al. (2012) cite a common saying of “what gets measured, gets done” citing the importance of measurement in the field of public health. Measurement theory has played a critical role in social science research for decades (Granberg-Rademacker, 2010). In 1946, Stevens first presented the levels of measurement that are still commonly presented in measurement and statistics textbooks today (e.g., Burns & Grove, 2009; Grimm & Yarnold, 2010; Hair, Black, Babin, & Anderson, 2010). These levels are presented in hierarchical order from lowest to highest: nominal (non-metric), ordinal (non-metric), interval (metric), and ratio (metric) (Burns & Grove, 2009; Hair et al., 2010). Each measurement level represents a distinct theoretical approach to assign numbers to qualitative observations (Granberg-Rademacker, 2010). In *Theory of Data*,

Coombs (1960) postulated that researchers use their best judgment in assigning numbers to observations and proposed a foundational system based on quadrants to classify behavioral observations and interrelate models of measurement.

The general rule in measurement is that the highest level of measurement possible should be used to more precisely measure the magnitude of the attribute and to permit a broad range of mathematical operations to be conducted (Burns & Grove, 2009; Hair et al., 2010). Stevens’ measurement levels are so ingrained in the field of statistics that most statistical packaging programs require users to identify the level of measurement based on Stevens’ taxonomy for each variable prior to running an analysis and may not perform the analysis without first changing the measurement designation of the variable (Velleman & Wilkinson, 1993). Nonetheless, there are many in the research community who oppose Stevens’ views of the “rules” regarding measurement scales and associated statistical techniques, arguing they are not required and are often unrealistic (Harwell & Gatti, 2001).

Nominal data are the lowest level and numbers associated with variables in this category of measurement only represent different categories of the variable with no meaningful ordering or ranking possible (Kachigan, 1991). Data from an ordinal scale can be assigned to various categories of an attribute and have the added benefit of a numerical notion of ordering or rank (Burns & Grove, 2009; Jupiter, 2013). Similar to nominal data, the categories have to be exclusive as well as exhaustive (Burns & Grove, 2009). For ordinal data, one can use any monotonic transformation that preserves order (Knapp, 1990). Ordinal scales are viewed as a quick and inexpensive method of characterizing relatively complex constructs (Merbitz et al., 1989).
Stages of cancer are a useful example in the health field for this type of data as there are clear differences between stages and they possess a numerical notion of ordering. However, as explained below, the stages do not reach the definition of interval-level of measurement because it is not clear whether the difference between two given stages (e.g., Stage 2 and Stage 3) is quantitatively different than the difference between two other stages (e.g., Stage 4 and Stage 5) (Jupiter, 2013). Ordinal data, therefore, are seen as having unequal intervals, and can be referred to as ordered metric scales (Burns & Grove, 2009). Other examples of constructs where ordinal data are usually assumed in the health field include pain intensity, degree of coping, or self-care ability (Burns & Grove, 2009).

Ordinal and Interval Data Controversy

Cliff (1991) argues that ordinal data should be used more often because the ultimate presentation of data in the findings or conclusion section of a report or article is often ordinal and that using interval approaches can lead to interpretation issues. In addition, he notes that with change data, ordinal data are preferable because they are distribution free and invariant under monotonic transformation (Cliff, 1991). Cliff (1991) suggests techniques such as the Wilcoxon Sum Rank test and the Mann-Whitney U statistic as useful approaches for reflecting change across individual observations and through group comparisons. However, Granberg-Rademacker (2010) postulated that researchers are generally uncomfortable or unsatisfied with ordinal data and that they often rely on transformations to deal with the issue even though there have been methodological improvements to statistical techniques available to handle ordinal data.
Likert scale data are also often cited as ordinal level rather than interval because of this emphasis on numerically equal distances – it may not always be accurate to say, for example, that the difference between Strongly Agree and Agree is the same as the distance between Neutral and Agree (Jupiter, 2013). The term “qualitative” is often used for ordinal data since equal distances between values may not necessarily have an equal quantitative distance (Kachigan, 1991). Because of this ambiguity, calculating central tendency based on the mean, which relies on the magnitude of values, could be misleading, and some fundamentalists – those adhering strictly to Stevens’ classification system and guidance – say this aspect precludes using parametric analyses (Burns & Grove, 2009; Townsend, 1990). Rather, the median is recommended as the preferred measure of central tendency for ordinal data because it only takes into account order (Jupiter, 2013). Others believe means can be calculated for ordinal data but that individuals should be careful about the statements or interpretations of such means (Knapp, 1990).

Data from interval scales have numerically equal distances between each point (Burns & Grove, 2009). Interval scales also encompass rules required of the lower level measurement scales (e.g., mutually exclusive and exhaustive categories, rank ordering) (Burns & Grove, 2009). Conservatives believe Likert scales should fall into the ordinal level of measurement because the intervals cannot be assumed to be equal (Jamieson, 2004). However, others, like Grimm and Yarnold (2010), suggest the numerical distance between numbers does not define the difference between ordinal and interval scales but rather the underlying dimension or construct the scale is measuring. Even if the distances between two points are not exactly equal, the errors may be small enough relative to the
measurements and parametric techniques could be used to maximize available data (Velleman & Wilkinson, 1993). Because values on an interval scale are assumed to represent a continuum, the expectation is that the magnitude of the attribute can be estimated. The interval scale does lack a zero point, unlike the highest measurement type: ratio-level scales (Burns & Grove, 2009).

The distinction between the levels of measurement can be especially challenging in the social and behavioral sciences (Grimm & Yarnold, 2010). Even in instances where data are clearly one type or another, the treatment of the data as ordinal versus interval in analyses has been extremely controversial (Burns & Grove, 2009). Those with a more conservative approach argue that ordinal data should be limited to statistical approaches that are designed for ordinal data, like nonparametric procedures (Burns & Grove, 2009). However, many argue that this approach is too constrictive and that very few (if any) measures in the social sciences would meet the interval-level criteria (Burns and Grove, 2009). It appears most researchers are willing to assume interval level data for cases that may seem vague (Grimm & Yarnold, 2010).

This view, from a practical perspective, seems to be mirrored in the literature. In health research, ordinal data are often not analyzed based on approaches recommended in statistical textbooks (Jakobsson, 2004). In a review of three nursing journals from 2003, ordinal data were used in over 30 percent of the 166 articles but only half were deemed to have appropriate data presentation and less than 60 percent were judged as conducting appropriate analyses based on the data according to the conservative rules for analysis (Jakobsson, 2004). In one review of conference proceedings on human factors in computing in the education field, Robertson (2012) found that only 8 percent of the
papers using Likert-type data reported using non-parametric statistics for the analysis. In a similar exploration, Harwell and Gatti (2001) reviewed published papers in the education field for a given year and found that out of over 150 dependent variables, 88 percent used ordinal, Likert scale data yet employed parametric analyses.

For those variables that fall in a gray area – where the distinction between ordinal and interval-level data is not clear – the increased sensitivity and power afforded by using parametric statistics is often cited as a reason for taking a less conservative approach (Grimm & Yarnold, 2010). Some argue that it is almost impossible to construct verbal scales that ensure equal intervals and that these “gray area” variables can be distinguished as “quasi-interval scales” (Kachigan, 1991). This controversy is particularly interesting to explore with the stages of change construct as equal intervals between stages are not always assumed, which is a requirement of interval-level data (Herzog & Blagg, 2007).

Because statistical analyses are considered relative to measurement, this issue of appropriate measurement is deemed by some as of utmost importance (Grimm & Yarnold, 2010). The conservative approach is often championed by various statistical textbook authors which, in turn, influences students and researchers (Binder, 1984). Guidance around analytic techniques is often organized around the measurement scale and associated analysis, with considerations made for both independent and dependent variables (Hair et al., 2010). For example, in Grimm and Yarnold’s textbook, Reading and Understanding Multivariate Statistics (2010), authors note that calculating the mean and standard deviation for ordinal data is relatively futile and that parametric statistics are inappropriate for analyzing ordinal data. The issue of analysis related to ordinal and interval-level data is elaborated further later in this chapter.
Others reject this notion of using the nature of measurement scales to define the statistical analysis and believe decisions should be driven by other factors instead of distinctions around whether a scale is defined as ordinal or interval in nature (Michell, 1986). Clearly, if the scale of measurement is questionable for a given study, the results and subsequent conclusions are likely to be called into question which can affect scientific advancements more generally (Grimm & Yarnold, 2010).

Measuring Stages of Change

Both theoretically and practically, accurately classifying the stage of change for an individual is critical in ensuring that stage-based interventions are effective (Brug et al., 2005; Bridle et al., 2005). This determination requires explicit definitions for both researchers and respondents to facilitate useful analysis and subsequent findings (Bridle et al., 2005). Furthermore, methods to categorize stage, in addition to the population of interest for a study (e.g., substance users, community health service providers), can account for variations in outcomes (Sbrocco, Osborn, Clark, Hsaio, & Carter, 2012). Littell and Girvin (2002) conducted an extensive review of stages of change across health behaviors, reviewing close to 90 studies on the topic. Although there is a lack of agreement around classifying individuals to a specific stage, two main approaches have been used to measure stage of change: algorithms and scales (Brug et al., 2005; Littell & Girvin, 2002).

Algorithm Approach

The algorithms category includes groupings of items with yes/no answers to several questions around current behavior and intention to change as well as using single
items with individual statements representing each stage (Littell & Girvin, 2002). The review revealed inconsistencies in studies adopting an algorithm approach such as using different time frames as reference points. Others have noted that because of the lack of a “gold standard” for stage allocation, the validity and reliability of stage of change measures are called into question as a result. Furthermore, others question the validity of stages in the Transtheoretical Model altogether (Brug et al., 2005). Sutton (2001) noted the effect of different algorithm approaches on study outcomes in his reference to studies on a group of smokers. The studies produced very different stage distributions and the associated difference in study conclusions had an effect on programming decisions for stage-based interventions (Sutton, 2001).

The most common approach to representing stage categories is through mutually exclusive, dichotomous variables; however, a subset of researchers used the algorithms to create an ordinal variable based on the stages of change (Littell & Girvin, 2002). Algorithms can be seen as having an advantage in measuring stage of change as they can place participants in mutually exclusive stage categories a priori (Littell & Girvin, 2002).

In a study of stage of change related to weight loss intentions and activities, Sbrocco et al. (2012) used an algorithm based on four yes/no questions in which a combination of responses resulted in stage of change placement. For example, individuals providing a “No” response to two to three of the questions were designated as Contemplators while those that indicated they maintained active changes for more than 6 months were designated as Maintainers (Sbrocco et al., 2012). The authors noted this method of categorization had established reliability among varying problem behaviors (Sbrocco et al., 2012).
Another staging algorithm example comes from a study that examined the pre-action stages (i.e., Precontemplation, Contemplation, Preparation) in relation to risk of sun exposure (Santiago-Rivas et al., 2013). Subjects were deemed “at-risk” for sun exposure if they indicated on a seven-item scale that they do not consistently use sun protection and have not used it for the past 12 months. Among these respondents, researchers assigned a stage of change based on other items (e.g., Contemplation stage if there was an intent to use sun protection in the next 12 months) (Santiago-Rivas et al., 2013). In another example, Liau et al. (2011) used an algorithm based on two items: 1) “Have you seriously thought about changing your behavior starting sometime in the next 6 months?” and 2) “Are you planning on changing your behavior sometime in the next 30 days?” using responses from the first item to assign the Precontemplation stage (“No” responses) and the Action/Maintenance stages (“Presently trying” responses). They used responses from second item to assign the Contemplation stage (“No” responses) and the Preparation stage (“Yes” responses) (Liau et al., 2011).

The use of single items is considered an “algorithm approach” in which individual statements represent each stage of change (Littell & Girvin, 2002). An individual Likert-type item collects information from a respondent on membership based on a list of ordered alternatives with each item providing “a discrete approximation of the continuous latent variable” (Clason & Dormody, 1994). Proponents of using single-item measures point to practical benefits (e.g., simplicity, efficiency of the approach, reduction of respondent burden and monotony, reduced costs, and increased response rates given shorter response times) as well as psychometric benefits (e.g., increases in face and construct validity, reductions in common method variance, reduced respondent bias, and
accurate assessments of global measures) (Fuchs & Diamantopoulos, 2009; Jordan & Turner, 2008; Van Keulen, Mesters, Van Mechelen, & De Vries, 2010).

For attributes that are more concrete and where there is little confusion over where someone may be placed, a single item is generally seen as sufficient (Rossiter, 2002). Studies in the marketing field have pointed to examples such as price perceptions or buying intentions (Fuchs & Diamantopoulos, 2009). One could argue the decision to adopt an evidence-based program at an organization is concrete in nature and thus a single item would be sufficient. In fact, single-item measures are often used in studies concerning organizational behavior and industrial psychology (Jordan & Turner, 2008).

This one item approach can be seen as controversial, however. Critics suggest single-item scales are more susceptible to measurement error and have fewer desirable psychometric properties than multiple item scales (Van Keulen et al., 2010). Data from individual items are likely to be skewed, often displaying floor or ceiling effects in their distributions, and mean differences may merely be a function of sample size rather than differences in the latent variable (Clason & Dormody, 1994).

Given the potential risks associated with use of single items, the use of multiple-item scales is usually recommended in most empirical settings (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). For example, Carifo and Perla (2007) suggest single items should only be analyzed in cases focused on item analyses or the initial exploratory phase in a study. However, the majority of the scale items for stage of change are not well developed (Littell & Girvin, 2002). In addition, respondents may overestimate their behavior when using multiple item scales (Van Keulen et al., 2010).
A study by Bergkvist and Rossiter (2007) showed that findings based on concrete, single item measures are equivalent to multiple-item measures. Specifically, if a construct is narrow in scope, unidimensional, and unambiguous (i.e., concrete) to the respondent, single items may be more appropriate measures than multiple items (Jordan & Turner, 2008). This guideline is especially true for single items that measure observable attributes – like “I have just begun to implement Motivational Interviewing in my work” – as opposed to cognitive or affective attributes for a construct (Jordan & Turner, 2008). Weinstein and Sandman (1992) suggested that assessing stage with a single item measure would be valuable in studying stage theories in health research given its ease of use.

Analyses of single-item data can also be contentious, particularly when considering the assumption in parametric analyses of normally distributed data (Clason & Dormody, 1994). In addition, there are concerns with Type I errors or inflated alphas due to repeated statistical testing when analyzing individual items separately (Carifo & Perla, 2007). The discrete nature of the information, such as distinct stages of change, should be taken into account to appropriately analyze single items from Likert scales (e.g., use of an ordered categorical variable versus a mean score) (Clason & Dormody, 1994).

The use of a single-item to measure stage of change is not uncommon in the literature (Berry et al., 2003; Kim et al., 2012; Weinstein & Sandman, 1992). In a study on hygiene practices, researchers asked participants to select the statement that best described their current practices from the following list: 1) I do not consider hand hygiene practices to be important for health; 2) I agree that hand hygiene behavior is very important for health, and although I do not currently practice it well now, I will in the near future; 3) I realize that hand hygiene behavior is very important for health, and I
have been practicing it for at least six months (Kim et al., 2012). Researchers assigned those selecting the first option to the Precontemplation stage, the second option to a grouping of the Contemplation and Preparation stages, and the third to a grouping of the Action and Maintenance stages (Kim et al., 2012). Berry et al. (2003) used a single item from a larger survey instrument to reflect each stage of change at the organizational level and noted that this served as a construct validity test in their study and an effective approach.

There have also been instances of algorithms that did not rely on yes/no questions or single-item measures. For example, West et al. (2013) studied exercise behaviors and used the reported number of exercise minutes logged weekly to group the sample into stages of change. The study team grouped the stages into three categories of Contemplation, Preparation, and Action and used the ADA goal of 150 minutes per week as the threshold for the Action and Maintenance stages (West et al., 2013).

*Scale Approach*

The other common method of assessing stage of change discussed by Littell and Girvin (2002) is the use of scales representing each stage (Lopes, Prieto, Delgado, Gamito, & Trigo, 2010). Modifications are often made to the wording based on the problem (e.g., smoking cessation versus exercise levels) although this is not always the case (Littell & Girvin, 2002). Using scales to measure stage of change often includes utilizing Likert scaling, an approach often used to measure opinions, beliefs, or attitudes around an underlying or latent variable (Clason & Dormody, 1994; DeVellis, 2012).

The Likert response format takes into account real time responses as well as the scoring or coding of the response leading to a more efficient form of measurement
(Carifo & Perla, 2007). The Likert scale has several advantages including ease of use, distinct consideration criteria, and a limited number of questions that can provide higher reliability estimates than other types of rating scales (Shaw & Wright, 1967). The equal intervals component is critical as the difference in the agreement or endorsement of a given option to the option next to it should be the same as for any other pair of options (DeVellis, 2012). When referring to Likert scales, some researchers assume the scale is a summation of item scores, similar to Likert’s original work, while others assume the term “scale” refers to the number of response alternatives for individual items (Clason & Dormody, 1994; Likert, 1932). This distinction should be made clear when conducting research given implications for analysis and interpretation (Carifo & Perla, 2007).

Using scales, researchers employ different methods to classify the stage of change including using the highest raw score, highest standardized score, or manipulating tied scores to either advance or regress into an adjacent stage (Littell & Girvin, 2002). Littell and Girvin (2002) noted not only that approaches to classification appear fairly arbitrary but, more importantly, that the different approaches to classification produce different results.

The most commonly used scale to measure readiness to change is the University of Rhode Island Change Assessment (URICA) although there are several variants including the Readiness to Change Questionnaire (RCQ), Treatment Motivation Questionnaire (TMQ), and the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) (Lopes et al., 2010). Items on the URICA correspond with a given stage. For example, two items associated with Precontemplation include “As far as I’m concerned, I don’t have any problems that need changing” and “All this talk about
psychology is boring. Why can’t people just forget about their problems?” (Sutton, 2001, p. 177).

Some of the commonly used scales only focus on specific stages. Researchers developed the RCQ to assess readiness to change among those abusing alcohol in a clinical setting (Kwon et al., 2012). This 12-item questionnaire only focuses on three stages within the stage of change construct (Precontemplation, Contemplation, and Action) but yielded Cronbach’s alpha coefficients of 0.73 and higher for each stage’s subscale (Kwon et al., 2012). The 32-item Stages of Change Questionnaire (STOCQ) contains 8 items assigned to four stages of the model: Precontemplation, Contemplation, Action, and Maintenance (D’Sylva, Graffam, Hardcastle, & Shinkfield, 2012). This scale has been tested in numerous settings and analyses produced high correlations between individual stage items and good separation between the stages (D’Sylva et al., 2012).

An adaptation of the STOCQ is the Pain Stages of Change Questionnaire (PSOCQ) which measures readiness to adopt a specific approach to chronic pain management. This instrument also consists of items using a five-point Likert scale rating grouped into the same four stages as the STOCQ (Dysvik et al., 2010). Lopes et al. (2010) evaluated the 19-item SOCRATES using classical test theory and found good internal consistency. However, after subsequently applying a Rasch model, the team found the functioning of the five categories corresponding to each of the stages of change was not optimal and after re-analyzing the data, found a three-category system to be more appropriate (Lopes et al., 2010).

However, others have called into question the utility of these popular scales, notably the URICA, RCQ, and the SOCRATES, citing evidence that they are not
measuring discrete stages of change (Sutton, 2001). In fact, Sutton (2001) suggested that scales may never be a useful means of assessing stage of change because they provide an individual with a score on a continuous dimension which is theoretically at odds with the stages of change construct. This suggestion questions whether or not the measurement is faulty or the model itself.

Comparison of Approaches

Although both methods of categorization have been deemed appropriate for measuring stage of change, the differences in stage criteria as well as classification have been inconsistent both within, and between the stage of change algorithms and scales (Littell & Girvin, 2002). In a meta-analysis by Lustria et al. (2009), it was not clear how stages of change were assessed or whether assessments were similar based on the given topical area, such as smoking cessation. These inconsistencies and issues around reliability and validity for the stage of change construct based on the Transtheoretical Model are particularly concerning when making decisions based on stage membership (Bridle et al., 2005).

Some argue that self-reporting is necessary since stage of change requires an assessment of intention (Ingledew, Markland, & Medley, 1998). However, other researchers have relied on observations to assign a stage of change (Noordman et al., 2013; Proctor et al., 2011). For example, in one study examining unhealthy behaviors by Noordman et al. (2013), researchers relied on observations of communications between the patient and a nurse and assigned a stage of change for the problem behavior identified. The authors prioritized earlier stages in situations where there may have been ambiguity in assigning a stage (e.g., ambiguity between Precontemplation and
Contemplation would be designated as Precontemplation (Noordman et al., 2013). Others have even combined both approaches by classifying stage of change by self-reported responses on a survey and then cross-validating responses with key informant interviews (Russell et al., 2012).

There is also no consensus on the best approach to allocate people to stages of change and few comparison studies of methods to measure stages of change have been conducted (Brug et al., 2005; Sutton, 2001). Several authors suggest this lack of consensus may be due to issues with the validity of the Transtheoretical Model stages themselves (Brug et al., 2005; Sutton, 2001).

**Distinct Stages**

By design, the stages of change are considered to be ordered and discrete in nature (Littell & Girvin, 2002). The order is based on the sequential transitions in the change process with skipping stages an unexpected occurrence (Littell & Girvin, 2002). The discrete nature is due to the qualitative different states identified in the readiness process (Littell & Girvin, 2002). In fact, the relationship between a particular stage of change and treatment of a specific behavior has been identified as evidence supporting the concept of stage of change as discrete data (D’Sylva et al., 2012). At any given point, an individual is expected to be in only one stage – thus each stage is considered mutually exclusive – and there should be no overlap between stages (Littell & Girvin, 2002; Martin, Velicer, & Fava, 1996).

Some researchers have found that the stages in the Transtheoretical Model are not mutually exclusive and that an individual can be categorized into more than one stage at the same time (e.g., Action and Maintenance) (D’Sylva et al., 2012). For example,
Herzog and Blagg (2007) found a divergence of concepts in stages of change related to smoking cessation when conducting a qualitative examination. D’Sylva and colleagues (2012) have suggested that the stage of change construct forces artificial categories and does not take into account the complexity of behavior change. In conducting a principal components analysis of a stages of change scale, McConnaughy, Prochaska, and Velicer (1983) found evidence supporting all but one stage in the model, Preparation. Items designed to measure this stage loaded highly on the adjacent stages of Contemplation and Action suggesting adding complexity to the issue (McConnaughy et al., 1983).

In an analysis examining the idea of qualitatively distinct stages of change, Kraft, Sutton, and Reynolds (1999) found no evidence that the Contemplation and Preparation stages warranted distinction from each other. They did find that Precontemplation emerged as a qualitatively distinct stage (Kraft et al., 1999). Lam et al. (2006) found support for the notion of stage of change as a continuum rather than discrete stages when examining general health promotion action. Other studies have focused more on potential differences within each stage of change pointing to the opportunity for additional tailoring of messaging based on these subgroup differences (Santiago-Rivas et al., 2013).

### Stage of Change Variable Treatment

In examining stages of change, defining what constitutes progress can be surprisingly complex. Some view positive outcomes in the Transtheoretical Model as any forward movement although others emphasize the need for action, or actual behavior change, as the main outcome of interest (Bridle et al., 2005). Others argue that movement from one stage (e.g., Precontemplation) to another (e.g., Contemplation) offers minimal
value and is seen as a “soft” outcome (West, 2005). In addition, a researcher may be interested in those originating in only one stage (for example a subject in the Precontemplation stage and their patterns of movement). Finally, it is not uncommon for choices around variable treatment to be driven by the sample being studied. For example, in one study on hand hygiene behavior in adolescents, the nature of the research and limitations due to the participants’ age influenced the decision to divide groups into three stages (Precontemplation, Contemplation, and Action) (Kim et al., 2012).

Nigg (2002) reports that the most common approach to analyzing stages of change is counting any movement forward in the stages as progress (Nigg, 2002). Although whether or not progression leads to behavior change is a contentious issue, many believe any forward movement through stages is evidence of effectiveness (Bridle et al., 2005). This belief is particularly evident the closer individuals are to the Maintenance stage (West, 2005). Treating forward movement as evidence of effectiveness is often done by operationalizing stages of change as an outcome variable and using the proportion of individuals that have moved to an Action or Maintenance stage (Nigg, 2002).

Horwath et al. (2013) followed a similar approach, of counting forward movement as progress, in their study of processes of change for fruit and vegetable consumption (measured by a 36-item Processes of Change instrument). If the stage transition score, calculated by subtracting Time 2 from Time 1, was greater than zero (0), it indicated stage progression and was labeled “successful.” If the score was zero (0) or lower, this was labeled “unsuccessful” (Horwath et al., 2013). These researchers also
identified a transition as “implausible” if the stage transition score was greater than 2 because of the limited time frame of the study (within 6 months) (Horwath et al., 2013).

The majority of studies regarding stage of change have focused more on transitions from non-action to action (Weinstein, Rothstein, et al., 1998). However, the definition of “action” could be set at a different criterion depending on the study. For example, if the goal of the study was to move individuals out of the Precontemplation stage, reaching the Contemplation stage could be used as the threshold. Two strengths of this approach are its straightforward nature and easy explanatory value (Nigg, 2002). The other main advantage is the ease of analysis, especially when used in randomized control trials examining the efficiency of an intervention. If all of the individuals participating in the study are relatively equal in terms of stage distributions, the proportion that reach the pre-specified stage of change (e.g., Action) would represent an estimate of treatment success (Nigg, 2002).

This approach does have distinct disadvantages as well with the primary concern of the loss of information through dichotomization, particularly in the case of stages of change which is a theoretically defined construct with five explicit stages (Nigg, 2002). This is of particular concern with stage-based intervention research since the distinct stages are often of interest to better tailor the intervention (Prochaska et al., 2002). Statistically, dichotomizing can lead to a decrease in statistical power and with this variable considered at a minimum, ordinal, using a nominal-level analytic technique would result in additional sensitivity losses (Nigg, 2002). Another disadvantage is that it is generally difficult to include covariates in the analysis (Nigg, 2002). If this approach (conceptualizing stages of change as the proportion of individuals moving to an Action or
Maintenance stage) is used, nonparametric techniques, like the McNemar test, that are sensitive to change in ordinal-level data over time would be more appropriate than parametric analyses (Nigg, 2002).

Using a trichotomous variable is also a possibility especially since most stage theories distinguish individuals based on three main categories: (1) those who have not decided to change the defined behavior, (2) those who have decided to change, and (3) those who are already engaged in change (Bridle et al., 2005). This distinction is quite clear in the Transtheoretical Model specifically between Contemplation (category 1), Preparation (category 2), and overt action (category 3) (Glanz & Bishop, 2010). Noordman and colleagues (2013) grouped the stages of change together similarly clustering Precontemplation and Contemplation into one category, Action and Maintenance into another category, and Preparation as its own category, in their study of clients’ behavior (e.g., smoking, dietary habits). While the authors noted small sample sizes in the Precontemplation and Maintenance stages, they also cited similar studies that adopted a comparable grouping approach in their analyses (Noordman et al., 2013).

Another approach to constructing a trichotomous variable would be grouping by “progress,” “regression,” or “no change.” This option can be preferable to dichotomization because it better reflects patterns of movement and includes the option of “no change” which can be very common with stages of change, particularly if data measurement periods are close together (Nigg, 2002). Some researchers have been unable to use the trichotomous approach incorporating regression because too few participants reported regressing stages (thus limiting treatment to dichotomous options of progression or no change) (Berry et al., 2003).
However, many of the limitations that were associated with dichotomization remain when using trichotomization. As noted above, information is lost when collapsing categories, particularly the distinction between stages. This can be a key limitation because someone moving from the Precontemplation stage to an Action stage would be characterized the same as someone moving from Precontemplation to Contemplation.

The mean score has also been used to categorize stage of change, particularly since this is a common approach in outcome evaluations (e.g., examining mean change in an intervention group compared to a control or comparison group) (Lipsey & Cordray, 2000). This mean score approach has also been taken by those who believe the stages are arbitrary and that stage descriptions have minimal value for those interpreting results (Sutton, 2001; West, 2005). An example of a smoker who plans to quit within the next 30 days as compared to stopping within the next 31 days illustrates this “arbitrary” distinction (Sutton, 2001). West (2005) cites this example as evidence that treating stages as categorical data can be problematic. Others support this idea citing empirical support for treating the variable as a continuous model of change (Bridle et al., 2005). Opponents of using the mean score argue that critical stage information is lost with this approach and this loss of information affects the ability to understand how the intervention affected participants and individuals most affected (Lipsey & Cordray, 2000).

In outcome research in the nursing field, many researchers have moved away from using the mean score to test for significant differences focusing instead on analyzing change scores and using exploratory approaches for outlier identification (Burns & Grove, 2009). This trend includes trying to better understand patterns of change, particularly when examining stage-based health theories (Burns & Grove, 2009).
One of the issues more specific to the stages of change construct is the cyclical nature of the stages of change where individuals are expected to progress as well as regress as they work through the stages (Brannon & Feist, 2004). Others suggest the most important aspect in studying change is consistency in the direction of change (Cliff, 1991).

Connected with this issue of patterns of change is capturing enough time points for more appropriate measurements of change. The timing and frequency of data collection efforts are critical. If follow-up assessments are insufficient, it is difficult to judge the effectiveness of an intervention, particularly whether or not action stages are maintained (Nigg, 2002). It is often the case that the ideal baseline position for an intervention related to changing a health behavior is a pre-action stage (i.e., Precontemplation, Contemplation) which necessitates a data collection period at least 6 months in length but often longer to account for transitions from earlier stages to action stages (Burns & Grove, 2009; Nigg, 2002). If the study period is not long enough, assessing movement between specific stages, such as between Precontemplation to Contemplation, is often used to assess progress (or regression) as an outcome variable (Nigg, 2002). Even though this approach does not take into account all stages of the Transtheoretical Model, this approach does complement the idea of stage-based, tailored interventions where the intent of the intervention may be more focused on movement to just one stage (Nigg, 2002).

Analytical Considerations

Non-parametric procedures, such as contingency table analysis or specialized structural equation models, are specifically designed to analyze ordinal data and are
based on ranks (Harwell & Gatti, 2001; Jakobsson, 2004). Fundamentalists believe only statistical methods designed for ordinal data, like nonparametric procedures, should be used for ordinal data – adhering strictly to Stevens’ original interpretation (Burns & Grove, 2009).

With cross-sectional data, stage of change as a construct is often treated as an ordered categorical variable (Nigg, 2002). Therefore, using an Analysis of Variance (ANOVA) to demonstrate differences in stages or a classification analysis approach like discriminant function analysis would be appropriate (Nigg, 2002). With more complicated datasets like time series in which subjects are measured for changes over a series of time points, more complex analytic approaches are more appropriate such as latent transition analysis or generalized estimating equations (Nigg, 2002). These analyses can take into account the various stages an individual may be in throughout the study period and capture progression, regression, and maintenance.

Parametric tests – such as the t-test, multiple regression, and ANOVA – are designed for continuous data, rely on normal distributions with equal variance, and are therefore said to require interval level data (Jakobsson, 2004; Townsend, 1990). Parametric tests are seen as more efficient methods for estimating data but only for interval or ratio scales and are purported to provide the possibility of clearer interpretations (Jakobsson, 2004; Vigderhous, 1977).

Some researchers point to the robustness of parametric techniques that help minimize the chance or likelihood of erroneous conclusions when certain statistical assumptions are violated by treating ordinal data as interval level data (Norman, Velicer, Fava, & Prochaska, 2000). Others, like Parker, McDaniel, and Crumpton-Young (2002)
suggest that violating assumptions of normality associated with parametric tests can be easy with a five-level ordinal scale. They note that approaches like regression, correlation and even creating scatterplots capitalize on the distance between adjacent categories being constant or equal and that using ordinal data results in a loss of face validity if these approaches are used (Parker et al., 2002). Of course, an alternative approach is to rescale ordinal data to an interval scale and then employ parametric analyses to analyze the interval level data although this can be complicated (Harwell & Gatti, 2001).

Using parametric statistics for ordinal data is often seen as appropriate if there is a large enough sample size (greater than 300 participants) and the presence of seven or more occupied categories in the outcome of interest (Kahler et al., 2008). However, Norman et al. (2000) noted that parametric statistics can be used with ordinal, Likert data, small sample sizes, and non-normal distributions without concern for wrong conclusions. Others agree that the extent to which responses have characteristics such as equal intervals is not critical (Weiss, 1986). Many researchers use methods such as t-tests and ANOVA to analyze ordinal data, thus treating them as interval level (Burns & Grove, 2009; Harwell & Gatti, 2001). In fact, some have suggested that measurement issues in general are irrelevant to the statistical treatment of the data (Anderson (1961) in Kurshid & Sahai, 1993). This is particularly important in light of practical considerations that the majority of data in the social sciences would not meet the more strict definition of interval level data (Burns & Grove, 2009; Weiss, 1986).

Some studies have specifically explored the analytical considerations by examining the differences between using parametric and non-parametric statistics on a single dataset. Vigderhous (1977) found regression estimates stemming from ordinal and
interval statistical analyses were not comparable. This finding provides evidence that level of data measurement and corresponding analyses can affect findings and subsequent conclusions (Vigderhous, 1977). Knapp (1990) pointed to research where findings have shown odd results when using parametric techniques with ordinal data such as means for one group being higher than the other on an original scale but lower than the other on a scale that has undergone permissible transformations.

Kahler et al. (2008) demonstrated that applying parametric procedures to ordered, categorical quality of life scores yielded different findings than those in which nonparametric analyses were used. They reasoned the inappropriate nature of applying parametric techniques to ordinal data resulted in uninterpretable parametric findings since equal differences between two points in their quality of life data cannot be assumed (Kahler et al., 2008). They also pointed to studies by researchers in the fields of psychiatry and dentistry that resulted in findings suggesting only nonparametric analyses should be used for ordered categorical data.

In addition, some analyses may be more sensitive to differences between ordinal and interval data than others. For example, Woodward, Hunter, and Kadlec (2002) found that a least squares regression using raw scores was more affected when using ordinal data than ordinal multiple regression. However, when data were transformed to ranks, the least squares regression had similar results to the ordinal multiple regression in terms of performance.

Others argue the controversy and attention surrounding ordinal and interval level measurement and analysis is misplaced and that research should be focused more on the research questions and purpose of the study (Binder, 1984; Clason & Dormody, 1994;
Velleman & Wilkinson, 1993). Scale types are not necessarily fundamental attributes of the data but rather dependent on what is being measured and intended types of conclusions – this *what* and *how* can change with data transformations or new supplemental information that aids interpretations of findings (Velleman & Wilkinson, 1993). Clason and Dormody (1994) suggest decisions on analytic techniques for Likert-scale items should be driven by the likelihood of providing meaningful findings and maintaining data richness. Proponents of this approach cite the number of research advances and valuable findings stemming from analyses that would be deemed “inappropriate” by those taking a conservative view (Binder, 1984; Clason & Dormody, 1994; Harwell & Gatti, 2001).

Instead of looking for the “true” system, why not experiment with more than one assignment of numbers, with validation depending on the numbering system’s ability to predict an outside variable, or its ability to clarify the particular problem? Certainly, in a young science some measurement experimentation is not only justifiable, but necessary. Instead of looking for the “true” system, why not focus on the most useful system that is guided either by empirical evidence in terms of prediction, or by any theoretical reasoning on the problem? (Labovitz, 1967 in Binder, 1974, p. 475).

Velleman and Wilkinson (1993) agreed with this point and noted that limiting studies and subsequent analyses based on an a priori assignment of scale type could be viewed as irresponsible research practices because they do not take into account other empirical evidence or contextual considerations.
Factors Affecting Methodological Decision-Making

Methodological decision-making, or decisions around planning and implementing the research strategy for a given study (Mock, 1972; Schumm, 2012), is a complicated process. Decisions on research strategies include areas such as sampling, definition of terms, identification of indicators, and analyses (MacDonald et al., 2006; Schumm, 2012). These are weighed against data issues, resource limitations, and political constraints to ensure study outcomes are both clear and meaningful to stakeholders (Bamberger, Rugh, & Mabry, 2012; Schumm, 2012).

Methodological decisions are usually made in conjunction with best practices in the given research field although there can be deviations which can affect outcomes and interpretations (Schumm, 2012). Over four decades ago, Mock (1972) noted the potential for influence by outside groups and the probable impact on decisions. Specifically, he stressed a likely tension between those in the “scientific community” and those evaluating research (e.g., clients) (Mock, 1972).

Choosing an appropriate evaluation design involves incorporating considerations of study scope and methodology to determine if the approach is feasible (Bamberger et al., 2012). Study designs should match research questions important to stakeholders as this positively influences the acceptance and implementation of recommendations from the study findings (Patient-Centered Outcomes Research Institute (PCORI), 2012). Problems occur when study design or research questions are influenced by the availability of data or other external constraints (PCORI, 2012). The analytic approach should be a means to answer the research question and research questions should not be limited by methodological constraint concerns (PCORI, 2012).
Similarly, decisions around analytic approaches should be carefully weighed, considered, and documented. As noted in PCORI’s Methodology Report (2012), choices for study design are nuanced and require tradeoffs among the limitations to each design. Appropriate design is not simply a binary decision (PCORI, 2012). However, the reality is that many times the limitations can and do inform choice of study designs and research questions, particularly in resource-constrained environments. The final choice, particularly in “real-world” evaluations is a combination of technical considerations, client preferences, and other contextual factors (Bamberger et al., 2012).

This process becomes more complicated when numerous stakeholders are invested in the outcomes. Characteristics of study designs are likely to be prioritized differently depending on the stakeholder group (e.g., policymakers, researchers, providers) which can sometimes lead to incompatible definitions of success (PCORI, 2012). However, to help maximize external validity for outcomes, involving as many stakeholder groups as possible and conducting research in a range of settings is important (Proctor et al., 2011). Defining the measure, its content, goals and thresholds, and the ultimate purpose all require judgment decisions by one or more stakeholders (Behn, 2003; Khorsan, Coulter, Hawk, & Choate, 2008).

Criteria for measure selection include, but are not limited to, study purpose or intended use, content relevance, reliability and validity, respondent and administrator burden, cost, appropriateness with a chosen theoretical model, availability, and cultural sensitivity (Khorsan et al., 2008). Factors such as sensitivity to change, or clinical responsiveness, and score precision are also important (Lipsey & Cordray, 2000; Khorsan
et al., 2008). For a stage of change outcome, documenting “any progress” may be more important than dichotomizing by action versus non-action.

In program evaluation, the changes an intervention is expected to produce are not always well-specified. Therefore, a critical piece of the evaluation process is defining goals to facilitate measuring outcomes (Lipsey & Cordray, 2000). Within this, the decisions involved around how to treat an outcome variable, such as the stage of change construct for decision to adopt, are key to understand given the potential to influence how findings are interpreted and used. If stakeholders, particularly those that will ultimately use these findings, are involved in this process and understand the strengths and weaknesses of proposed methods, this helps support their continued commitment to using findings from the study and advancing the dissemination field (Patton, 2013).

This dissertation borrows heavily from the principles of utilization-focused evaluation in which an evaluation is designed to be responsive to the needs of “primary intended users” so decisions around the use of methods and measurement are guided by ultimate intended use (Patton, 2013; Bamberger et al., 2012). While all evaluations are expected to focus on utility (e.g., constructive purpose, meeting needs of stakeholders, relevant information), utilization-focused evaluations are a shift in focus from evaluations focusing specifically on the program or intervention of interest to stakeholder intended uses and judgments on value and worth, and the evaluator serving in a facilitator capacity (Patton, 1994). Utilization-focused evaluations are considered both “highly personalistic and situational” which leaves them vulnerable to attacks on external credibility (MacDonald et al., 2001; Patton, 1994, p. 317).
This approach focusing on intended users is especially relevant when considering variable treatment decisions in addition to overall study design. Choices are driven by attention to utility. Tradeoffs between independence in evaluations and use of evaluation results need to be carefully considered (Patton, 1994). Specifically, in Patton’s (2013) “Checklist for Utilization-focused Evaluations,” Step 10 specifies “Negotiate appropriate methods to generate credible findings that support intended use by intended users” (p. 11).

The negotiation process for identifying appropriate methods is complex because the evaluator needs to work with stakeholders on a range of issues in determining methodology and data treatment decisions including appropriateness, practicality, cost-effectiveness, and ethical considerations (Patton, 2013). Theory is, of course, important but the role of context is especially critical (Glasgow, 2009). Stakeholders such as policy managers and service providers, often groups using the information in real-world settings, likely have different viewpoints on the validity and reliability of evidence and the acceptable data collection and analysis burden compared to researchers and analysts (Head, 2010; Proctor et al., 2011). For example, those conducting D&I research are often focused on addressing pragmatic decisions for various stakeholders (e.g., policy makers, service providers) such as program effectiveness in a given setting or implementation costs (Glasgow, 2009).

It is common in the policymaking environment for conflicts to emerge between contextual factors and methodological rigor (Head, 2010). Many policy decisions are the product of the interplay of facts, norms, and preferred courses of action (Head, 2010). For example, policymakers may be more sensitive to cost, while service providers may be
more interested in feasibility or burden issues (Proctor et al., 2011). Researchers and analysts, on the other hand, often view methodological rigor as a key factor in providing credible evidence (Fleischer & Christie, 2009). This can create a tension across groups that must be managed.

Even when presented with the same evidence, stakeholder opinions and interpretations of findings can vary greatly depending on local context (Damschroder et al., 2009; Head, 2010). Thus, engaging stakeholders and understanding their priorities is critical to facilitating effective implementation at the local level (Damschroder et al., 2009). Decisions about measure selection and variable treatment depend on a range of factors including stakeholder priorities, study timelines, and assessment costs (Glasgow, 2009). This conversation around priorities and challenges can be facilitated by the researcher. In fact, for an effort to implement evidence-based programs in diverse community settings, Miller et al. (2012) found that having conversations between stakeholders facilitated partners’ understanding of significant issues and limitations such as needing flexible timelines for Institutional Review Board considerations.

The issues above underscore the importance of appraisal and evaluation of the needs of stakeholders and contextual considerations by the researcher along with the need for methodological rigor (Simpson & Flynn, 2007). In the field of program evaluation, evaluators are held responsible for what is seen as “misevaluation” which includes a flawed evaluation study design, inappropriate methodology, poor data collection and/or analysis, and low quality reporting (Fleischer & Christie, 2009).
Conclusion

The stages of change construct is often used in health research although measurement and analytic considerations, as well as theoretical limitations, may have influenced study findings. Chapter Three presents an overview of the original study and specifies the analysis approach for this dissertation incorporating measurement and analytical considerations presented in Chapter Two.
CHAPTER THREE: METHODS

This chapter discusses how the research for the study was conducted. It provides a brief overview of the original study including the sample and data collection instruments. The six study research questions are presented along with the respective analysis approach. The chapter ends with a discussion on the potential use of findings for two hypothetical stakeholders.

Sample

A study team at MANILA Consulting Group, Inc., under contract to SAMHSA developed a clustered, randomized controlled trial to examine the influence of dissemination strategies on decisions to adopt a community behavioral health practice. Through this 2-year study researchers sought to better understand effective approaches to dissemination of evidence-based practices (Williams, Williams, et al., 2013).

Recruitment for the original study began by drawing from two major national membership organizations: the National Association of Community Health Centers and the National Council for Community Behavioral Healthcare. Organizations where motivational interviewing was routinely practiced and supported were excluded from the study. This process resulted in 92 organizations with 49 from the community health sector and 43 from the community behavioral health sector (Williams, Williams, et al., 2013). The organizations were based in a range of geographic settings (i.e., urban, suburban, rural).
The individual-level participants were then selected based on organizational recommendations for staff involved in the decision-making process for new clinical practice adoption (Williams, Williams, et al., 2013). Participants’ positions in the organization consisted of those in director or administrative positions and those in staff or practitioner positions. In return for their participation in the study, participants received a gift certificate at the conclusion of the study, and organizations received a participation plaque, a motivational interviewing reference book, and a thank you letter from the funding organization, SAMHSA (Williams, Dusablon, et al., 2013). The complete description of recruitment and assignment can be found in earlier publications (Williams, Williams, et al., 2013; Williams, Dusablon, et al., 2013).

Variables Measured and Instruments Used

The baseline and follow-up surveys for the study collected information regarding organizational barriers to implementing evidence-based practices, management strategies to support the implementation of evidence-based practices, organizational readiness for change, individual attitudes toward evidence-based practices, individual readiness for change, stage of adoption decision, and consumer involvement in the organizational decision-making process. The instrument also included organizational information such as the number of clients served, client demographics, geographic location (e.g., urban, suburban), and financial capacity (Williams, Dusablon, et al., 2013).

The variable of interest for this dissertation is the main outcome measure in the original study (focused on the decision to adopt motivational interviewing). This measure, adapted from the work of McGovern et al. (2004), assessed participants’ interest in adopting motivational interviewing, a counseling approach that attempts to
increase the patient or consumer’s awareness of potential problems of the behavior in question (Williams, Blais, et al., 2013). Variations of this measure have also been used in other studies (McGovern et al., 2004). The single-item measure asked study participants to “Please indicate your level of interest in adopting MI into your program.” The item responses were based on the stages of change construct in the Transtheoretical Model and provided construct validity support for the measure. The response options and the corresponding stages of change are presented in Table 1 (McGovern et al., 2004; Prochaska et al., 2002; Williams, Blais, et al., 2013). The response option “I am not familiar with MI” was recoded as Precontemplation. This approach not only increased the sample size but aligns with the definitions in the Transtheoretical Model: someone who is not interested in, or is unaware of, the practice and does not have any intention of adopting the practice (Prochaska et al., 2002).

Table 1. Response Options and Corresponding Stages of Change

<table>
<thead>
<tr>
<th>Response Option</th>
<th>Stage of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am not familiar with MI (1)</td>
<td>Precontemplation</td>
</tr>
<tr>
<td>I am not interested and do not think this practice would be effective in my program (1)</td>
<td>Precontemplation</td>
</tr>
<tr>
<td>I have considered MI but see many pros and cons (2)</td>
<td>Contemplation</td>
</tr>
<tr>
<td>I am leaning in the direction of adopting MI in my program (3)</td>
<td>Preparation</td>
</tr>
<tr>
<td>I have just begun to implement MI in my work (4)</td>
<td>Action</td>
</tr>
<tr>
<td>I have been using MI, and efforts are in place to maintain it (5)</td>
<td>Maintenance</td>
</tr>
</tbody>
</table>

In the baseline survey, this item measured the baseline stage in the adoption-decision process prior to implementing the dissemination strategies while the follow-up survey provided a post-intervention measure of change. In addition, the group for the two dissemination strategies – Webinars plus informational packet (Intervention group) and
the informational packet only (Comparison group) was used as the key variable of interest.

Data Collection Procedures

Community health centers and community behavioral health organizations in the study were matched as pairs based on the type, size, and setting of the organization. The study team randomly assigned one organization in each pair to a study condition (exposure to a dissemination strategy) and assigned the other organization to the alternate study condition. The study team tested two dissemination strategies: a passive strategy consisting of an information packet and a more active strategy including an informational packet and two interactive Webinars (Williams, Williams, et al., 2013). This dissertation refers to the active strategy as the “Intervention” and the passive strategy as the “Comparison” group. The original study team obtained approval from the Institutional Review Board of MANILA Consulting Group as well as the Office of Management and Budget (Williams, Dusablon, et al., 2013).

The team administered three surveys: a baseline survey and two follow-up surveys. The team administered the baseline survey immediately following completion of consent forms to participate in the study and the follow-up survey after study participants received all of the intervention components (Williams, Williams, et al., 2013). The first follow-up survey was administered 1 month following the intervention while the second follow-up survey was administered 3 months after the intervention. The study team developed three versions of the survey (one for the “main director”, one for “other directors”, and one for staff) and only altered certain scales based on the role of the
respondent. The team collected survey data through a Web-based platform (Qualtrics). Participants had a 2-week window in which to complete each 25 to 30 minute survey.

Data Treatment

Because this is an existing dataset, instances of missing data were pre-treated prior to obtaining the de-identified dataset. The study team assigned a code of “-88” for instances where the participant did not receive the question (due to skip patterns or whether questions were not on their version of the survey) and a value of “-66” for “don’t know” responses. If the respondent did not answer enough items to reliably calculate scale scores (for example, did not take the last half of the survey), analysts included a value of “-55” in the dataset. Finally, the study team utilized mean value imputation for “real” missing data where the participant did not answer and should have. There were very few instances of this type of missing data (J. Williams, personal communication, February 15, 2012).

Due to administrative constraints in the original study, the two follow-up assessments were administered more closely together than planned. As such, there were no significant differences between the two follow-up administrations (Williams, Williams et al., 2013). For this dissertation, data from the last follow-up (follow-up 2) were used given the higher number of cases and additional opportunity for movement along the stages of change continuum. In addition to the data treatment by the original study team, cases where the baseline (pretest) or follow-up 2 (posttest) responses were unavailable were coded as missing (-55) for the Decision to Adopt outcome variable given the focus on difference scores for several of the research questions.
As noted in the Variables Measured and Instruments Used section, respondents who selected “I am not familiar with MI” were recoded from “0” to “1” so they are assigned to the Precontemplation stage. Although the original study excluded individuals that reported being in the Action or Maintenance stage at baseline given the nature of the intervention, (Williams, Williams, et al., 2013), these 64 cases were retained in the current study since the focus is on how analytic and data treatment decisions affect findings rather than the impact of the original intervention.

Scatterplots were generated to identify potential influential data points. In examining outliers, 11 cases emerged as potential influential points based on residual diagnostics such as Cook’s distance and the studentized residuals. Residual diagnostics are useful for researchers in determining whether there is something occurring in the dataset that has not been accounted for by the model. These 11 cases were examined further to understand why these were extreme and the potential influence on the statistical analysis. After examination of the residuals, data plots, and raw data, only two of these cases were problematic in that responses were not consistent with expectations (e.g., indicating the intervention was being implemented at the pretest but in Precontemplation or Contemplation stages in the posttest). These two cases were thus excluded from the analysis.

Analyses

Given the distribution of respondents across the different stages of change categories, as well as the number of stages for Research Question 2, sample sizes were adequate to run each intended analysis. The ability of the original dataset to detect differences was tested by comparing the group means on the posttest between the
Intervention and Comparison groups. Results showed that the original data were not able to detect differences. This could be due to the Comparison group experiencing improvements given a “usual care” condition as opposed to no treatment in the original study. Therefore, the likelihood for differences in success for the Intervention group would likely be higher if there was a true control group and thus illustrating differences between groups may have been more pronounced. To ensure the data had adequate power to detect differences among the five research questions for this dissertation, data were augmented by adding a constant of one (1) for the Intervention group. Instances where the addition of the constant yielded a score of “6,” were classified as “5” to maintain alignment to the Stages of Change construct.

Data simulation would be an alternative approach to data augmentation. Data simulation can be used to demonstrate potential differences between measurement and analysis approaches relatively quickly and inexpensively. However, using the data augmentation approach was preferable over data simulation because this study is focused on real-world evaluation situations. Thus the data, while augmented, reflect more realistic scenarios, particularly as they relate to the stage of change construct and underlying theory of change. Adding a value of one to the Intervention group still preserves the general distribution of respondents along the stage of change continuum and maintains the needed context for data interpretations that would be meaningful for stakeholders.

The benefit of using data augmentation also extends to other evaluation decisions including data preparation considerations for analysis (e.g., missing data, outliers, plausible change scenarios).
Research Question 1

Are there posttest differences in the outcome variable, as measured by posttest Decision to Adopt score, between the Intervention and Comparison groups?

Research Question 1 focused on the differences in groups in terms of Decision to Adopt as measured by the stage of change item – this is a similar approach to the treatment of the variable in the original study. Experimental condition (Intervention versus Comparison Group) was the major independent variable with posttest values on Decision to Adopt score as the outcome variable. The following model illustrates how this approach assesses treatment effects:

\[ \text{DecAdoptScore} = \alpha_0 + \beta_1 \times \text{Intervention} + \varepsilon \]

where \( \text{DecAdoptScore} \) represents an outcome variable; \( \text{Intervention} \) is a dichotomous variable indicating the respondent’s experimental condition; and \( \varepsilon \) is the error term.

Research Question 2

Are there posttest differences in the outcome variable, as measured by stage progression (number of stages), between the Intervention and Comparison groups?

Research Question 2 also addresses differences in groups in terms of Decision to Adopt but instead focuses on number of stages as a measure of progression (or regression) rather than using the posttest score. With this approach, each respondent was assigned a value based on the difference in the number of stages between the pretest to the posttest. For example, those respondents reporting progressing a single stage were assigned a value of 1, those progressing two stages a score of 2, regressing a single stage a score of -1 (negative one), etc. This approach is recommended as a useful alternative to other approaches since it takes into account the direction and magnitude of change and
also avoids issues with losing information from dichotomizing data (Nigg, 2002). This research question utilized a similar analytic approach to that used to answer Research Question 1:

\[ \text{NumStagesDecAdopt} = \alpha_0 + \beta_1 \times \text{Intervention} + \varepsilon \]

where \( \text{NumStagesDecAdopt} \) represents the outcome variable and \( \text{Intervention} \) represents the respondent’s experimental condition.

**Research Question 3**

*Are there posttest differences in the outcome variable, as measured by action versus non-action, between the Intervention and Comparison groups?*

Research Question 3 focused on the difference in terms of action versus non-action, an approach many studies have taken by grouping pre-action stages and action stages (Weinstein, Rothstein, et al., 1998). This question was addressed using logistic regression techniques. Logistic regression is used when one is interested in the probability of an outcome. Specifically, the presence of action at the posttest (recoded as Action=1, Non-action=0) was the dichotomous outcome variable of interest. Since the outcome variable was dichotomous, the relationship between the outcome variable and the independent variable (in this case the treatment group) is not linear which is a violation of the assumption of linearity in ordinary least squares regression. In addition, the errors are not normally distributed (the error term for a discrete variable follows a binomial distribution). Logistic regression requires the use of the maximum likelihood procedure (rather than least squares estimates) because of the nonlinear nature of the logistic transformation (Hair et al., 2010). Therefore, applying a linear regression equation is inappropriate.
Instead, odds, or more specifically a logistic transformation of the odds called a logit, was used as the outcome variable. This is illustrated in the logistic regression model for this research question examining the presence of an action stage:

$$\text{logit}(P) = \alpha_0 + \beta_1 \times \text{Intervention} + \varepsilon$$

This equation predicts whether a case will be classified into Action for Decision to Adopt at the posttest as opposed to non-action. Again Intervention was the dichotomous variable indicating the respondent’s experimental condition.

**Research Question 4**

*Are there posttest differences in the outcome variable, as measured by stage progression (yes/no), between the Intervention and Comparison groups?*

Research Question 4 used a similar analytic approach as Research Question 2 (a logistic regression model) to focus on differences in terms of the presence or absence of stage progression (recoded as Presence=1, Absence=0). Maintenance and regression to a previous stage were treated as an absence in this equation:

$$\text{logit}(P) = \alpha_0 + \beta_1 \times \text{Intervention} + \varepsilon$$

This equation predicts whether a case will be classified into progression for Decision to Adopt at the posttest as opposed to non-progression. Again Intervention is the dichotomous variable indicating the respondent’s experimental condition.

**Research Question 5**

*Are there posttest differences in the outcome variable, as measured by the stage of change (categorical), between the Intervention and Comparison groups?*

For Research Question 5, the outcome variable was treated as ordered, categorical data where the categories correspond to the five stages of change. This research question
was investigated through the use of logistic regression but using a cumulative logit (or ordered logistic regression) model to accommodate the ordinal nature of the outcome variable (Menard, 2001):

\[
\text{logit}(P) = \alpha_0 + \beta_1 \cdot \text{Intervention} + \varepsilon
\]

where \( k \) represents the five stages of change (i.e., Precontemplation = 1, Contemplation = 2, Preparation = 3, Action = 4; Maintenance = 5). This equation attempts to predict that a case will be classified into a specific stage compared to other stages. The ordered logistic regression model is preferable to the multinomial logistic regression model which does not assume ordering of the categories (e.g., the categories are nominal in nature) (Menard, 2001). Again \( \text{Intervention} \) is the dichotomous variable indicating the respondent’s experimental condition.

**Research Question 6**

*Are there differences in the significance and magnitude of findings across approaches? If so, how do the findings differ?*

To address Research Question 6, which is the primary question of interest, findings from Research Questions 1 through 5 were evaluated on a range of methodological and pragmatic factors detailed below. Figure 2 provides a visual representation of the five different treatment approaches used in each research question. The number and proportions of participants in each stages of change grouping will be presented across all five scenarios in Chapter Four (Moher et al., 2010).
Figure 2. Illustration of Research Question Variable Treatments

*Research Question 1: Posttest Score for Outcome Variable*

*Research Question 2: Number of Stages Progressed (or Regressed) for Outcome Variable*

*Research Question 3: Dichotomous Action and Non-Action for Outcome Variable*

*Research Question 4: Dichotomous Progress and No Progress for Outcome Variable*

*Research Question 5: Ordered Categorical Stage for Outcome Variable*
In order to help assure that results will be used as intended, there must be a proper balance between technical considerations and practical ones (Patton, 2013). Practical considerations are particularly important for policy makers who often take an informal approach to assessing evidence such as relying on anecdotes or opinion surveys (Choi et al., 2005). Even though the research questions do vary, the approaches are compared to provide a means to better understand the strengths, weaknesses, and general nuances for each approach to help aid future decision-making (Connell, Lynch, & Waring, 2001). Presenting the information in this manner facilitates data interpretation and increases understanding of the potential impact of different approaches (Patton, 2013).

The factors for comparison draw from two major sources: the Consolidated Standards of Reporting Trials (CONSORT) statement (Schulz, Altman, & Moher, 2010) and the Standards for Reporting on Empirical Social Science Research (American Educational Research Association (AERA), 2006). The CONSORT statement is a checklist of items that are considered essential when reporting findings from a study or trial (Moher et al., 2010). It is primarily intended for randomized control trials but is also suitable for other designs such as factorial or cluster designs (Moher et al., 2010). Developed in the early 1990s by an expert panel of journal editors and methodologists, this checklist emphasizes clarity, comprehensiveness, and transparency to help readers, editors, and other stakeholders critically evaluate and interpret findings from a study (Moher et al., 2010). The CONSORT statement is revised periodically to ensure recommendations reflect the most current evidence in the field for reporting requirements (Moher et al., 2010).
The Standards for Reporting on Empirical Social Science Research provide guidelines on a broad range of social science research studies to ensure authors provide adequate evidence and transparency (AERA, 2006). These guidelines facilitate use of findings by other stakeholders interested in the study. The standards are broken out into eight different areas and may vary depending on the type of research methodology used (AERA, 2006). These types of comparison factors provide insight into methodological quality (such as potential issues around power to detect an association) and provide credible evidence for different stakeholders (Patton, 2013; West & Dupras, 2013).

Specific elements include:

- An index of the quantitative relation between variables (e.g., regression coefficient, odds ratio) and associated significance level;
- Indication of uncertainty of index (e.g., standard error, confidence interval); and
- Analytic considerations that may affect inferences (e.g., ceiling or floor effects, violations of assumptions) (AERA, 2006; Donaldson et al., 2004).

Pragmatic factors include the practical significance based on interpretation of findings as well as compromises and limitations with the selected approach. Practical significance is critical to include given arguments that statistical significance alone is a poor standard (Lipsey & Cordray, 2000). Limitations are useful to document in comparisons, particularly when working with different stakeholders, to ensure the users understand the implications of decision-making around research strategies (Connell et al., 2001). Within limitations, the approximate time investment for each approach is documented. This is particularly important for stakeholders that influence programming or policy as timing of information can often take precedence over the quality of the data (Choi et al., 2005;
Connell et al., 2001; Patton, 2013). Limitations are often omitted from study reports yet they contain critical information for decision-making (Moher et al., 2010).

Table 3 illustrates the comparison format to facilitate this analysis. Comparison factors are borrowed from the CONSORT Statement (Schulz et al., 2010), the Standards for Reporting on Empirical Social Science Research (AERA, 2006), and Patton’s (2013) utilization-focused checklist. Following each analysis for Research Questions 1 through 5, relevant information was pulled and entered into the table. Each general area (i.e., Methods, Results, and Discussion) was then examined, in turn, and summative conclusions were drawn based on the results of each discussion.
<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
<th>Research Question 1</th>
<th>Research Question 2</th>
<th>Research Question 3</th>
<th>Research Question 4</th>
<th>Research Question 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Variable</td>
<td>Treatment of single-item stage of change measure</td>
<td>Posttest Score</td>
<td>Number of Stages</td>
<td>Dichotomous Variable (Action / Non-Action)</td>
<td>Dichotomous Variable (Progress / No Progress)</td>
<td>Ordered Categorical Variable</td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Methods</td>
<td>Statistical methods used to compare primary outcomes</td>
<td>Ordinary Least Squares Regression</td>
<td>Ordinary Least Squares Regression</td>
<td>Binary Logistic Regression</td>
<td>Binary Logistic Regression</td>
<td>Ordered Logistic Regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>Number of participants analyzed for the primary outcome</td>
<td>n = (Intervention)</td>
<td>n = (Comparison)</td>
<td>n = (Intervention)</td>
<td>n = (Comparison)</td>
<td>n = (Intervention)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Results for each group (intervention and comparison)</td>
<td>R² = b = t = d = p-value =</td>
<td>R² = b = t = d = p-value =</td>
<td>Pseudo R² = Odds ratio = Wald = p-value =</td>
<td>Pseudo R² = Odds ratio = Wald = p-value =</td>
<td>Pseudo R² = Odds ratio = Wald = p-value =</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Estimation</strong></td>
<td>Precision (e.g., 95% confidence interval)</td>
<td>95% CI = S.E. =</td>
<td>95% CI = S.E. =</td>
<td>95% CI = S.E. =</td>
<td>95% CI = S.E. =</td>
<td>95% CI = S.E. =</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discussion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>Limitations such as sources of potential bias or imprecision</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
</tr>
<tr>
<td></td>
<td>Interpretation consistent with results and considering other relevant evidence</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
</tr>
</tbody>
</table>

Table 3. Research Question 6 Discussion Comparison Format
This comparison of both methodological and practical factors incorporates an extensive discussion in Chapters Four and Five of if, and how, variable treatment decisions affect interpretations of findings. It is often this careful integration of information that is most useful in guiding programmatic and policy changes and a deeper understanding of the intervention of interest (Lipsey & Cordray, 2000). Drawing from the principles of utilization-focused evaluation, presented in Chapter Two, methodological decisions and the presentation of findings should be responsive to the needs of primary intended users (Bamberger et al., 2012; Patton, 2013).

Although there are likely numerous audiences for this information, this dissertation focuses on two hypothetical stakeholders in the evaluation process for illustrative purposes. The first stakeholder is a federal program officer interested in findings around dissemination approaches and how they influence decision to adopt evidence-based programs and practices. This stakeholder is able to influence policy and programming decisions. The program officer is representative of a primary intended user of evaluation findings, specifically a stakeholder in a position to decide something about the intervention being evaluated (MacDonald et al., 2006). The second hypothetical stakeholder is a consultant working on research and evaluation studies with various clients and responsible for reporting on the findings of the intervention in the original study. While the examples for each stakeholder are based in the literature and real-world practice around evidence-based program adoption and D&I research, the stakeholder illustrations are used to stress the implications with respect to utilization-focused evaluation or measurement considerations as opposed to recommendations around evidence-based practice adoption.
Conclusion

Five analyses were conducted using different approaches to variable treatment for this construct (e.g., posttest score, dichotomous outcome). Research Question 6 focuses on the comparison of the approaches used in Research Questions 1 through 5 and is intended to provide insight into whether and how study findings may differ and the implications of these differences for policy makers and other stakeholders. Chapter Four presents the findings from these analyses in detail.
Chapter Four: Results

This chapter presents the findings of Research Questions 1 through 5 and includes an emphasis on Research Question 6 where former research questions are compared.

Sample Description

The final dataset contained 208 cases with 110 respondents in the Intervention group and 98 in the Comparison group (see Table 4.1). Data were analyzed using IBM SPSS Statistics for Windows, Version 21.0 (IBM Corp., Armonk, NY). Consistent with the original study, there were no significant differences between the Intervention and Comparison groups at the pretest time point on the Decision to Adopt outcome (Williams, Williams et al., 2013). The mean score for the Decision to Adopt outcome for the Intervention group at baseline was 2.23 (SD=1.153) while the mean for the Comparison group was 2.27 (SD=1.040).

Table 4.1
Sample Sizes for Group Condition

<table>
<thead>
<tr>
<th>Group Condition</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>110 (52.9%)</td>
</tr>
<tr>
<td>Comparison</td>
<td>98 (47.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>208 (100%)</td>
</tr>
</tbody>
</table>

Research Question 1

*Are there differences in the posttest in adoption decision, as measured by the Decision to Adopt score?*

Research Question 1 focuses on the differences in groups in terms of Decision to Adopt as measured by the stage of change item. Table 4.2 presents the distribution of
cases based on the posttest score by study group. The Intervention group (M=4.25) had a higher mean than the Comparison group (M=3.45) at the posttest.

Table 4.2
Distribution Based on Posttest Score\(^1\) (Research Question 1)

<table>
<thead>
<tr>
<th>Group</th>
<th>1 N (%)</th>
<th>2 N (%)</th>
<th>3 N (%)</th>
<th>4 N (%)</th>
<th>5 N (%)</th>
<th>Mean Posttest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>8 (8.2%)</td>
<td>10 (10.2%)</td>
<td>37 (37.8%)</td>
<td>16 (16.3%)</td>
<td>27 (27.6%)</td>
<td>3.45</td>
</tr>
<tr>
<td>Intervention</td>
<td>0 (0.0%)</td>
<td>8 (7.3%)</td>
<td>16 (14.5%)</td>
<td>27 (24.5%)</td>
<td>59 (53.6%)</td>
<td>4.25</td>
</tr>
</tbody>
</table>

\(^1\) Note distributions for all research questions based on augmented data (i.e., adding a factor of 1 to the Intervention group)

This difference in groups was assessed using an ordinary least squares regression (see output in Figure 4.1). In the simple regression, Group Condition accounted for a significant 11.8% of the variance in the Decision to Adopt posttest score \([R^2 = 0.118, b = .796, t = 5.240, p < 0.001]\). In general, the Intervention group had significantly higher scores on Decision to Adopt in the posttest than the Comparison group. The positive value of the coefficient, 0.796, indicates an approximately 0.8 point advantage for those in the Intervention group compared to those in the Comparison group.

Figure 4.1
SPSS Output for Research Question 1

Model Summary\(^a\)

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R^2</th>
<th>Adjusted R^2</th>
<th>Std. Error of Estimate</th>
<th>R^2 Change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.343(^a)</td>
<td>.118</td>
<td>.113</td>
<td>1.094</td>
<td>.118</td>
<td>27.463</td>
<td>1</td>
<td>206</td>
<td>.000</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Constant), Intervention

\(^b\) Dependent Variable: RQ 1: Posttest value
### ANOVA

<table>
<thead>
<tr>
<th>Mode 1</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>32.878</td>
<td>1</td>
<td>32.878</td>
<td>27.463</td>
<td>0.000b</td>
</tr>
<tr>
<td>Residual</td>
<td>246.618</td>
<td>206</td>
<td>1.197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>279.495</td>
<td>207</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Dependent Variable: RQ 1: Posttest value  
b. Predictors: (Constant), Intervention

### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>95.0% Confidence Interval for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>3.449</td>
<td>.111</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>.796</td>
<td>.152</td>
</tr>
</tbody>
</table>

5. Dependent Variable: RQ 1: Posttest value

### Research Question 2

*Are there differences in the posttest in adoption decision, as measured by stage progression (number of stages)?*

Research Question 2 focuses on the differences in groups in terms of adoption decision but instead uses number of stages as a measure of progression (or regression) rather than the posttest score. Table 4.3 presents the distribution of cases based on the number of stages by study group. The mean number of stages (progressed or regressed) for the Comparison group (M=1.16) was lower than the Intervention group (M=1.85).
Table 4.3
Distribution Based on Number of Stages (Research Question 2)

<table>
<thead>
<tr>
<th>Group</th>
<th>-1 N (%)</th>
<th>0 N (%)</th>
<th>1 N (%)</th>
<th>2 N (%)</th>
<th>3 N (%)</th>
<th>4 N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>3 (3.1%)</td>
<td>13 (13.3%)</td>
<td>56 (57.1%)</td>
<td>20 (20.4%)</td>
<td>3 (3.1%)</td>
<td>3 (3.1%)</td>
</tr>
<tr>
<td>Intervention</td>
<td>1 (0.9%)</td>
<td>4 (3.6%)</td>
<td>43 (39.1%)</td>
<td>31 (28.2%)</td>
<td>25 (22.7%)</td>
<td>6 (5.5%)</td>
</tr>
</tbody>
</table>

Mean Number of Stages:
Comparison: 1.16
Intervention: 1.85

This question was also assessed using ordinary least squares regression with output presented in Figure 4.2. In the simple regression, Group Condition accounted for a significant 11.0% of the variance \([R^2 = 0.110, b = .682, t = 5.038, p < 0.001]\) in number of stages for Decision to Adopt.

Figure 4.2
SPSS Output from Research Question 2

**Model Summary**

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>(R^2)</th>
<th>Adjusted (R^2)</th>
<th>Std. Error of Estimate</th>
<th>(R^2) Change</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.331&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.110</td>
<td>.105</td>
<td>.975</td>
<td>.110</td>
<td>25.381</td>
<td>1</td>
<td>206</td>
<td>.000&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Predictors: (Constant), Intervention
<sup>b</sup> Dependent Variable: RQ 2: Number of stages

**ANOVA**

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>24.119</td>
<td>1</td>
<td>24.119</td>
<td>25.381</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>105.760</td>
<td>206</td>
<td>.950</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>219.880</td>
<td>207</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Dependent Variable: RQ 2: Number of stages
<sup>b</sup> Predictors: (Constant), Intervention
Research Question 3

Are there differences in the posttest in adoption decision, as measured by action versus non-action?

Research Question 3 focuses on the difference in terms of action versus non-action, an approach many studies have taken by grouping pre-action stages and action stages (Weinstein, Rothstein, et al., 1998). This question was examined using binary logistic regression where action was dichotomized. Table 4.4 presents the distribution of the Decision to Adopt variable broken out by Action (i.e., Action or Maintenance stage at posttest) and Non-Action (i.e., Precontemplation, Contemplation, and Preparation).

Table 4.4
Distribution Based on Dichotomized Action (Research Question 3)

<table>
<thead>
<tr>
<th>Group</th>
<th>Non-Action N (%)</th>
<th>Action N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>55 (56.1%)</td>
<td>43 (43.9%)</td>
</tr>
<tr>
<td>Intervention</td>
<td>24 (21.8%)</td>
<td>86 (78.2%)</td>
</tr>
</tbody>
</table>

The results of the logistic regression, presented in Figure 4.3, revealed a statistically significant relationship between the Action variable and Group Condition (Wald=24.467, SE=0.308, p<0.001). The exponentiated coefficient (Exp(B)=4.583) represents the relative level of the outcome variable for the represented group compared to the omitted group (Hair et al., 2010). This coefficient is considered the best means of interpreting the impact of a dummy
variable in a logistic regression (Hair et al., 2010). In terms of directionality, because the exponentiated coefficient is greater than 1, there is a positive relationship (e.g., for odds increase with a positive change in the independent variable). In examining magnitude, those in the Intervention group (Intervention=1) are 4.6 times more likely to be in the Action or Maintenance stage than the Comparison group. The pseudo $R^2$ measure, based on Nagelkerke’s $R^2$, is 0.162, indicating the Group Condition variable accounts for about 16% of the variation in the model.

Figure 4.3
*SPSS Output from Research Question 3*

Model Summary

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log likelihood</th>
<th>Cox &amp; Snell R Square</th>
<th>Nagelkerke R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>249.795&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.119</td>
<td>0.162</td>
</tr>
</tbody>
</table>

5. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>95% C.I. for EXP(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention (1)</td>
<td>1.522</td>
<td>.308</td>
<td>24.467</td>
<td>1</td>
<td>.000</td>
<td>4.583</td>
</tr>
<tr>
<td>Constant</td>
<td>-.246</td>
<td>.204</td>
<td>1.462</td>
<td>1</td>
<td>.227</td>
<td>.782</td>
</tr>
</tbody>
</table>
Research Question 4

Are there differences in the posttest in adoption decision, as measured by stage progression (yes/no)?

Research Question 4 uses a similar analytic approach as Research Question 3, a binary logistic regression model, to focus on differences in terms of the presence or absence of stage progression. Table 4.5 presents the distribution between the Comparison and Intervention group based on progress (advancing 1 or more stages) and no progress.

Table 4.5
Distribution Based on Dichotomized Progress (Research Question 4)

<table>
<thead>
<tr>
<th>Group</th>
<th>No Progress N (%)</th>
<th>Progress N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>16 (16.3%)</td>
<td>82 (82%)</td>
</tr>
<tr>
<td>Intervention</td>
<td>5 (4.5%)</td>
<td>105 (95.5%)</td>
</tr>
</tbody>
</table>

The results of this logistic regression model, presented in Figure 4.4, revealed a statistically significant relationship between the Progress variable and Group Condition (Wald=6.999, SE=0.533, p<0.009). The exponentiated coefficient (Exp(B)=4.098) reflects a positive relationship and suggests those in the Intervention group (Intervention=1) are 4.1 times more likely to progress one or more stages than the Comparison group. Nagelkerke’s $R^2$ is 0.081 indicating the Group Condition variable accounts for about 8 percent of the variation in the model.
Figure 4.4

SPSS Output from Research Question 4

Model Summary

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log likelihood</th>
<th>Cox &amp; Snell R Square</th>
<th>Nagelkerke R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>127.908a</td>
<td>0.039</td>
<td>0.081</td>
</tr>
</tbody>
</table>

5. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1a</td>
<td>Intervention (1)</td>
<td>1.410</td>
<td>.533</td>
<td>6.999</td>
<td>1.008</td>
<td>4.098</td>
<td>1.441</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>1.634</td>
<td>.273</td>
<td>35.750</td>
<td>.000</td>
<td>5.125</td>
<td></td>
</tr>
</tbody>
</table>

5. Variable(s) entered on step 1: Intervention

Research Question 5

Are there differences in the posttest in adoption decision, as measured by the stage of change (categorical)?

For Research Question 5, the outcome variable was treated as ordered, categorical data where the categories will correspond to the five stages of change. Table 4.6 summarizes the observed distribution of the ordered categorical variable (Decision to Adopt) by Group Condition.

Table 4.6

Distribution Based on Ordered Categorical Variable (Research Question 5)

<table>
<thead>
<tr>
<th>Group</th>
<th>Precontemplation N (%)</th>
<th>Contemplation N (%)</th>
<th>Preparation N (%)</th>
<th>Action N (%)</th>
<th>Maintenance N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>8 (8.2%)</td>
<td>10 (10.2%)</td>
<td>37 (37.8%)</td>
<td>16 (16.3%)</td>
<td>27 (27.6%)</td>
</tr>
<tr>
<td>Intervention</td>
<td>0 (0.0%)</td>
<td>8 (7.3%)</td>
<td>16 (14.5%)</td>
<td>27 (24.5%)</td>
<td>59 (53.6%)</td>
</tr>
</tbody>
</table>
This research question was also investigated through the use of logistic regression but using a cumulative logit (or ordered logistic regression) model to accommodate the ordinal nature of the outcome variable (Menard, 2001). The ordinal logistic regression model’s “pseudo $R^2$” value (Nagelkerke=0.117) indicates the Group Condition accounts for approximately 12% of the variation between respondents in their attainment of a given stage of change. Figure 4.5 presents the output from the ordinal logistic regression and Table 4.7 presents the parameters from the ordinal regression along with the cumulative odds, proportion, and category probability. The constant odds ratio for this proportional odds model is also presented. The odds ratio indicates the odds for the Intervention group achieving a higher level in the stages of change is 3.6 times the odds of the Comparison group.

Figure 4.5
*SPSS Output from Research Question 5*

<table>
<thead>
<tr>
<th>Goodness-of-Fit</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>8.304</td>
<td>3</td>
<td>.040</td>
</tr>
<tr>
<td>Deviance</td>
<td>10.097</td>
<td>3</td>
<td>.018</td>
</tr>
</tbody>
</table>

Link function: Logit.

<table>
<thead>
<tr>
<th>Pseudo R-Square</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox and Snell</td>
<td>.109</td>
</tr>
<tr>
<td>Nagelkerke</td>
<td>.117</td>
</tr>
<tr>
<td>McFadden</td>
<td>.042</td>
</tr>
</tbody>
</table>

Link function: Logit.
Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>95% Confidence Interval Bound</th>
<th>Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention=0</td>
<td>-1.281</td>
<td>.266</td>
<td>23.253</td>
<td>1</td>
<td>.000</td>
<td>-1.801</td>
<td>-.760</td>
</tr>
<tr>
<td>Intervention=1</td>
<td>0</td>
<td>a</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. This parameter is set to zero because it is redundant.

Table 4.7

Parameters from the Ordinal Regression of Group Condition on the Outcome Variable

<table>
<thead>
<tr>
<th>Comparison</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cumulative logit</strong></td>
<td>--</td>
<td>-2.712</td>
<td>-1.411</td>
<td>0.154</td>
<td>1.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cumulative odds</strong></td>
<td>--</td>
<td>0.07</td>
<td>0.24</td>
<td>1.17</td>
<td>2.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cumulative proportion</strong></td>
<td>1</td>
<td>0.94</td>
<td>0.80</td>
<td>0.46</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Category probability</strong></td>
<td>0.06</td>
<td>0.13</td>
<td>0.34</td>
<td>0.21</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Intervention**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cumulative logit</strong></td>
<td>--</td>
<td>-3.993*W1</td>
<td>-2.692*W2</td>
<td>-1.127*W3</td>
<td>-0.195 W4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cumulative odds</strong></td>
<td>--</td>
<td>0.02</td>
<td>0.07</td>
<td>0.32</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cumulative proportion</strong></td>
<td>1</td>
<td>0.98</td>
<td>0.94</td>
<td>0.76</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Category probability</strong></td>
<td>0.02</td>
<td>0.05</td>
<td>0.18</td>
<td>0.21</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Odds Ratio</strong> (Comparison/Intervention)</td>
<td>--</td>
<td>3.600</td>
<td>3.600</td>
<td>3.600</td>
<td>3.600</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<0.001

W1 Wald=98.738; W2 Wald=98.184; W3 Wald=30.569; W4 Wald=1.092

There was a significant difference in probabilities for all thresholds (p<0.001) except Maintenance where the probability was 55% for the Intervention and 25% for the Comparison. The probability of being at Precontemplation was 2% for the Intervention group compared to 6% for the Comparison. The probability of being at Contemplation for the Intervention group was 5% compared to the Comparison group at 13%. In moving to
Preparation, the Intervention group had an 18% probability compared to the Comparison group’s probability of 34%. For Action, the probability of being at the Action stage was 21% for both the Intervention and Comparison groups.

Research Question 6

*Are there differences in the significance and magnitude of findings across approaches? If so, how do the findings differ?*

As discussed in Chapter Three, the primary question of interest, Research Question 6, involves the comparison of the findings from Research Questions 1 through 5 on a range of methodological and pragmatic factors (high-level overview presented in Table 4.8).
Table 4.8
Research Question 6 Comparison

<table>
<thead>
<tr>
<th>Area</th>
<th>Research Question 1</th>
<th>Research Question 2</th>
<th>Research Question 3</th>
<th>Research Question 4</th>
<th>Research Question 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td>Posttest Score</td>
<td>Number of Stages</td>
<td>Dichotomous Variable (Action / Non-Action)</td>
<td>Dichotomous Variable (Progress / No Progress)</td>
<td>Ordered Categorical Variable</td>
</tr>
<tr>
<td></td>
<td>Ordinary Least Squares Regression</td>
<td>Ordinary Least Squares Regression</td>
<td>Binary Logistic Regression</td>
<td>Binary Logistic Regression</td>
<td>Ordered Logistic Regression</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>n = 111 (Intervention)</td>
<td>n = 111 (Intervention)</td>
<td>n = 111 (Intervention)</td>
<td>n = 111 (Intervention)</td>
<td>n = 111 (Intervention)</td>
</tr>
<tr>
<td></td>
<td>n = 99 (Control)</td>
<td>n = 99 (Control)</td>
<td>n = 99 (Control)</td>
<td>n = 99 (Control)</td>
<td>n = 99 (Control)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>R² = 0.118</td>
<td>R² = 0.110</td>
<td>Pseudo R² = 0.162</td>
<td>Pseudo R² = 0.081</td>
<td>Pseudo R² = 0.117</td>
</tr>
<tr>
<td></td>
<td>b = 0.796</td>
<td>b = 0.682</td>
<td>Odds ratio = 4,583</td>
<td>Odds ratio = 4.098</td>
<td>Odds ratio = 3.600</td>
</tr>
<tr>
<td></td>
<td>t = 5.240</td>
<td>t = 5.083</td>
<td>Wald = 24.467</td>
<td>Wald = 6.999</td>
<td>Wald (see table 4.7)</td>
</tr>
<tr>
<td></td>
<td>d = 0.731</td>
<td>d = 0.708</td>
<td>p-value = &lt;0.001</td>
<td>p-value = &lt;0.001</td>
<td>p-value (see table 4.7)</td>
</tr>
<tr>
<td></td>
<td>p-value = &lt;0.001</td>
<td>p-value = &lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation</td>
<td>95% CI = 0.497–1.096</td>
<td>95% CI = 0.415–0.949</td>
<td>95% CI = 2.507–8.379</td>
<td>95% CI = 1.441–11.650</td>
<td>95% CI (see table 4.7)</td>
</tr>
<tr>
<td></td>
<td>S.E. = 0.152</td>
<td>S.E. = 0.135</td>
<td>S.E. = 0.308</td>
<td>S.E. = 0.533</td>
<td>S.E. (see table 4.7)</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major Finding</td>
<td>The Intervention group advanced significantly more than the Comparison group</td>
<td>The Intervention group advanced through more stages of change than the Comparison group</td>
<td>Those in the Intervention group are 4.6 times more likely to be in the Action or Maintenance stage than the Comparison group</td>
<td>Those in the Intervention group are 4.1 times more likely to progress than the Comparison group</td>
<td>The odds for the Intervention group achieving a higher level in the stages of change is 3.6 times the odds of the Comparison group</td>
</tr>
<tr>
<td>Major Limitation(s)</td>
<td>• Does not take into account theorized discrete nature of stages</td>
<td>• Implausible values of change (greater than 3 stages) may limit sample size</td>
<td>• Dichotomizing variable loses stage information</td>
<td>• Dichotomizing variable loses stage information</td>
<td>• More difficult to interpret • Use may be limited in resource-constrained settings</td>
</tr>
</tbody>
</table>
Methodological Factors

Sample size and distribution. Figure 4.6 provides a visual representation of the five different treatment approaches used in each research question, as presented in Chapter Three, but now includes the number of cases in each grouping (Moher et al., 2010). Each analysis included 98 cases from the Comparison group and 110 cases from the Intervention group. As noted in Chapter Three, sample sizes were adequate to run each intended analysis with the original distribution of respondents across the different stages of change categories, as well as the number of stages for Research Question 2. This was a concern given how certain approaches, like logistic regression, require larger samples than ordinary least squares regression (Hair et al., 2010).
Figure 4.6 Visual Comparison of Cases Within Groupings

**Research Question 1: Posttest Score**

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Comparison</th>
<th>M=3.45</th>
<th>N=98</th>
<th>Intervention</th>
<th>M=4.25</th>
<th>N=110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation</td>
<td></td>
<td></td>
<td></td>
<td>Contemplation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td></td>
<td></td>
<td></td>
<td>Action</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Research Question 2: Number of Stages**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Comparison</th>
<th>N</th>
<th>Intervention</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td></td>
<td>3</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>13</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>56</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>20</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

**Research Question 3: Action and Non-Action**

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Non-Action</th>
<th>N=55</th>
<th>Action</th>
<th>N=43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contemplation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Non-Action</th>
<th>N=24</th>
<th>Action</th>
<th>N=86</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contemplation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Research Question 4: Progress and No Progress**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Comparison</th>
<th>N=16</th>
<th>Intervention</th>
<th>N=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Research Question 5: Categorical Stage**

![Stage Diagram]

**Significance, magnitude, and estimation.** Table 4.9 presents the comparison of methodological factors for the five research questions. Comparing the p-values across questions, all analyses yielded statistically significant differences in the posttest between the Intervention and Comparison groups. This finding is unsurprising as data were augmented for the Intervention group to ensure this.

Table 4.9
**Comparison of Methodological Factors**

<table>
<thead>
<tr>
<th>Research Question</th>
<th>p-value</th>
<th>R² (or Pseudo R²)</th>
<th>Effect Size*</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Continuous posttest score)</td>
<td>&lt;0.001</td>
<td>0.118</td>
<td>0.731</td>
<td>0.497–1.096</td>
</tr>
<tr>
<td>2 (Number of stages)</td>
<td>&lt;0.001</td>
<td>0.110</td>
<td>0.708</td>
<td>0.415–0.949</td>
</tr>
<tr>
<td>3 (Action/Non-action)</td>
<td>&lt;0.001</td>
<td>0.162</td>
<td>0.839</td>
<td>2.507–8.379</td>
</tr>
<tr>
<td>4 (Progress/No progress)</td>
<td>&lt;0.009</td>
<td>0.081</td>
<td>0.778</td>
<td>1.441–11.650</td>
</tr>
<tr>
<td>5 (Ordered categorical)</td>
<td>n/a</td>
<td>0.117</td>
<td>0.737</td>
<td>0.709-0.764</td>
</tr>
</tbody>
</table>

The R² (or pseudo R²) values accounted for approximately 11% of the variance in outcomes for three of the research questions (Research Question 1 (Posttest Score), Research Question 2 (Number of Stages), and Research Question 5 (Categorical)). Of note, the percent of variance for the Action/Non-Action dichotomous variable (16.2%) was double the percent for the Progress/No Progress dichotomous variable (8.1%).

For the purposes of further comparison, the same effect size, d, was used across research questions and calculated using the Campbell Collaboration’s web-based calculator (http://www.campbellcollaboration.org/resources/effect_size_input.php). This
standardized mean-difference is typically used when there is a continuous outcome variable although methods have been developed to accommodate other types of outcome variables (e.g., dichotomous) (Lipsey & Wilson, 2001). The effect size is complementary to statistical significance, providing an indication of the strength of the relationship (Kotrlik, Williams, & Jabor, 2011).

Using guidance developed by the What Works Clearinghouse, effects that are greater than 0.25 standard deviations units are considered substantively important suggesting a meaningful difference between the Intervention and Comparison groups (Gersten & Hitchcock, 2009). Further, based on Cohen’s “rules of thumb” for effect size interpretation (Lipsey & Wilson, 2001), there was a large effect for all research question approaches which is not surprising given data augmentation procedures. However, the effect sizes did range from 0.708 (Research Question 2: Number of Stages) to 0.839 (Research Question 3: Action) based on different variable treatment approaches.

In terms of estimation, the confidence interval provides a measure of uncertainty around the intervention effects, providing a range of values for where one can be reasonably sure the true effect lies (Higgins & Green, 2011). Specifically, confidence intervals provide a point estimate of the population parameter and an associated interval to reflect likely error (e.g., the estimate’s precision) (Cumming & Finch, 2001). The imprecision of an estimate can be due to factors such as sample size and measurement error (Cummings & Finch, 2001).

Confidence intervals provide information that is accessible which supports understanding and interpreting a given analysis (Cumming & Finch, 2001). For example, the upper and lower bounds allow for interpreting findings and the potential implications
at either end of the given range. This may be particularly important if the results at one end of a range are clinically meaningful but not at the other (Medina & Zurakowski, 2003). In addition, confidence intervals are presented in the same units of the variable of interest, unlike a p-value, that also facilitates interpretation of findings (Medina & Zurakowski, 2003).

For Research Question 1, the confidence interval for the effect size ranges from 0.45 to 1.10 for a point estimate of 0.73. This interval is wider than ideal as it ranges from a medium effect interpretation to a large effect interpretation. However, even at the lowest end of the range, the Intervention group still shows an improvement over the Comparison group and thus could be considered practically significant. For Research Question 2, the confidence interval ranges from 0.42 to 0.95, with an effect size point estimate of 0.71, which presents a similar interpretation to Research Question 1 (that the precision ranges from a medium to large effect but that at a minimum, there is a 95% confidence that the true estimate is at least a medium effect).

Research Question 3 yielded a confidence interval of 2.51 to 8.34, for the odds ratio point estimate of 4.58, which is a wide range. Therefore, while the interpretation of the point estimate would be that the Intervention group is 4.6 times more likely to achieve the action stage, the 95% confidence interval lower limit shows that this could be as low as 2.5 times as likely or as high as 8.3 times as likely which indicates low precision. However, there still is a notable effect with the lowest range of the estimate (e.g., 2.5 times more likely).

The confidence interval for Research Question 4 was 1.44 to 11.65 which is the widest interval range indicating the least amount of estimation precision. While this still
includes an odds ratio of 1.4 at the lowest range (indicating the Intervention group is 1.5 times as likely to progress compared to the Comparison group), one should not be confident in the precision of the point estimate (OR=4.10). Finally, the confidence interval for Research Question 5 is 0.71 to 0.76 with an effect size estimate of 0.74 which is the most precise estimate for all five of the research questions.

Variations in estimates are due to factors including observed factors, unmeasured factors, and sampling error (Greenberg, 2007). For example, in terms of observed factors, the Intervention group shows a larger effect size for Research Question 3 (Action vs Non-Action) compared to Research Question 4 (Progress or No Progress). This is reasonable given that the Comparison group was not a true control and received a condition that led to progress (i.e., influenced Decision to Adopt), albeit not as much progress as the Intervention group. Sampling error, or the expected variation in an estimated parameter due to the sample (Hair et al., 2010), is also a likely source of the variance. Thus, a larger sample size would reduce sampling error across the five research questions and result in more precise estimates.

As noted in Chapter Three, data augmentation procedures were necessary (adding a constant of 1 to the Intervention group) as the original dataset was not able to detect differences between the Intervention and Comparison group. The data augmentation procedures also contributed to the significant p-values for Research Questions 1 through 5, which limited the ability to compare statistical significance in Research Question 6. Although this consideration is important to keep in mind, the original study’s dataset still allows for a useful demonstration of the implications for variable treatment decisions and serves as a more effective platform than using simulated data. In addition, it is important
to note that this comparison of approaches should not be considered data dredging (or “hypothesis fishing”) as the hypotheses for Research Questions 1 through 5 are not intended to be used or answered in isolation but rather presented for demonstrating potential differences (Dahl, Grotle, Benth, & Natvig, 2008).

**Pragmatic Factors**

*Practical significance.* The confidence intervals combined with the point estimates can provide useful information on the utility of intervention (Higgins & Green, 2011). However, much of this depends on the expectations for what is considered useful. For example, if the expectation for the study is that the intervention will show any effect and the size of the effect is inconsequential, the wide confidence interval bands for certain research questions may be seen as less problematic in terms of practical significance. Alternatively, if the expectation is that the Intervention group needs to produce a large effect compared to the Comparison group to be considered useful, the confidence intervals play a larger role with less precision around the certainty of a large effect for Research Questions 1 (posttest score), 2 (number of stages), and 4 (progress).

*Limitations.* Each research question has limitations when compared to either the execution of the analysis or the utility of subsequent findings. As discussed in Chapter Two, and elaborated in Chapter Five, these limitations can play a significant role in decision-making for a given study. For example, Research Questions 1 (posttest score) and 2 (number of stages), ordinary least squares regression techniques were employed. Depending on an individual’s field of research or discipline area, these may be viewed as simpler and more common analytic approaches than those used in Research Questions 3 through 5 and thus preferable. Many stakeholders familiar with research are conversant
with simple linear regression approaches, particularly since linear regression is often presented and discussed in most introductory statistics textbooks. Non-linear models, like binary logistic regression models, are usually touched on but not always explored in as much depth. Again, this varies depending on the field of research (e.g., the public health field often utilizes binary logistic regression to convey presence or absence of a disease).

Ordinal logistic regression (used for Research Question 5) is often not discussed in detail, and is not readily available in statistical packages (particularly earlier versions) and thus can be considered the most challenging to conduct. Even in this dissertation, the ordinal regression approach took the longest to conduct, interpret, and present findings appropriately. In low-resource settings, particularly where individuals with advanced statistics knowledge are not readily available, more advanced techniques may not be practical or feasible. In addition, the likelihood for errors in interpretation is higher with more complex procedures (Bamberger et al., 2012).

Although simplicity of findings is largely a product of the analytic approach and underlying dataset, the onus is typically on the researcher to present the findings as simply as possible relative to the audience (Choi et al., 2005). This means understanding the primary stakeholders for the study findings is critical to informing decisions around which approach will be comprehensible to that audience. The first two research questions may be considered the easiest to interpret given the widespread use of ordinary least squares regression. However, those in certain fields (e.g., public health) may prefer approaches like used in Research Questions 3 and 4, which involve the use of binary logistic regression. More broadly, for those audiences that understand probabilities, this can be easy to interpret but certainly challenging to those not used to seeing findings
presented in this format. Again, there are alternatives to presentation in logistic regression (so the burden would be on the researcher to present the findings in a manner that facilitates the easiest understanding for the client) but it is likely that these are more difficult than Research Questions 1 and 2. Research Question 5 is likely the most challenging to interpret given the nature of the outcome variable and the five different stages.

Another major limitation to certain research questions, occurring at the question formulation stage, concerns the Transtheoretical Model and stages of change construct that is the guiding model for the original study. For example, Research Question 1 does not take into account the theorized discrete nature of the stages by treating the outcome variable as continuous (Prochaska et al., 2002). However, as noted in Chapter Two, some researchers argue that the stages are not discrete citing instances where individuals can be in more than one stage at a time or that certain stages are not qualitatively distinct (D’Sylva et al., 2012; Kraft et al., 1999; Lam et al., 2006).

Research Questions 3 and 4 can also be considered problematic in this context of alignment with the Transtheoretical Model because of the loss of stage of change information when dichotomizing the outcome variable. The treatment of the outcome variable in these cases effectively reduces the amount of information available to understand the impact of the intervention according to the stages of change (Nigg, 2002).

Finally, although not problematic in this particular sample given the distribution and relaxed assumptions on what would constitute “plausible” progress, the choice of categorization approaches can affect the minimum sample size needed for analysis. For example, for Research Question 2, if there are enough implausible values of change in the
time period (e.g., greater than 3 stages), treating these cases as missing may limit sample size and the power of the analysis to detect differences. Only two cases emerged as implausible given the relaxed assumptions on expectations for progression and regression because of the intervention and nature of the outcome variable (e.g., the decision to adopt an evidence-based practice may be more sensitive to change in a shorter time period than other outcome behaviors often associated with stages of change like the decision to quit smoking).

Conclusion

Five analyses were conducted using different approaches to variable treatment for this construct (e.g., posttest score, dichotomous outcome) and results documented to answer Research Questions 1 through 5. Each Research Question yielded a significant difference between the Intervention and Comparison groups for the Decision to Adopt outcome. Research Question 6 focused on the comparison of the approaches used in Research Questions 1 through 5, presenting the relative strengths and limitations compared to other questions for both methodological and pragmatic factors. Chapter Five presents the discussion including hypothetical examples of how findings could be used in practice.
Chapter Five: Discussion

Across Research Questions 1 through 5, the Intervention group tended to have better outcomes for stage of change in terms of Decision to Adopt than the Comparison group. Given the data augmentation procedures, this finding is unsurprising. The more interesting component lies in how “better outcome” was defined in Research Questions 1 through 5 and the implications of those definitions on data treatment and subsequent analysis. Specifically, Research Question 6 demonstrates that methodological choices around data treatment and subsequent analysis result in similar yet distinct findings and presented different practical constraints and limitations. This chapter presents a more detailed discussion of these choices and their implications. It also presents considerations for methodological decision-making to guide stakeholders for future studies.

Implications for Stage of Change Variable Treatment and Analysis

There were five approaches to treating the stage of change outcome variable. These aligned with different characterizations of a “positive outcome” or success and generally relied on the underlying theoretical properties of the variable. The treatment approach for Research Question 1 used a simple posttest score for the outcome variable. Because the variable data stemmed from a single item, the response options aligned with the stages of change for Decision to Adopt. Using this raw posttest score is a simple approach that assumes equal distances between each stage (interval level data) which aligns with certain interpretations of the Transtheoretical Model. Given the continuous interval level data, ordinary least squares regression was an appropriate analytical choice to test differences between groups. However, while certainly defensible given the
circumstances, some may argue that this approach is not appropriate for the stage of change construct because of the discrete nature of the stages. In addition, as described in Chapter Two, there is a measurement issue around treating ordinal level data as interval level. Specifically, several authors found that research findings differed and were uninterpretable since equal distances between two points could not be assumed (Kahler et al., 2008; Knapp, 1990; Vigderhous, 1997).

The treatment approach for Research Question 2 relied on the number of stages progressed (or regressed). This method of variable treatment assumes more than one stage of progress is plausible in the given time frame, that skipping stages may be feasible, and also assumes equal distances between each stage. From a pragmatic perspective, this approach offers a distinct advantage by taking into account both progression and regression which aligns with the dynamic nature of the Transtheoretical Model. However, opponents may argue that this rationale is flawed. For example, Littell and Girvin (2002) note that skipping stages is not expected within the Transtheoretical Model. In addition, this approach does not account for the qualitative aspect of each of the stages so critical stage information is not captured. From an analytical perspective, similar to Research Question 1, Research Question 2 can take advantage of ordinary least squares regression which is often considered straightforward and easy to interpret.

Research Question 3 used the Action stage as a threshold for Action and Non-Action as the outcome variable. This approach is in line with a good majority of studies that focus on this transition (Weinstein, Rothstein, et al., 1998). This method is often seen as straightforward and easy to explain which can be critical for certain stakeholders when it comes to evaluation use. In addition, analysis is also straightforward, particularly in
comparing groups from randomized control trials (Nigg, 2002). Specifically, the binary logistic regression provides understandable findings in terms of odds but other approaches could be used such as nonparametric procedures like the McNemar test (Nigg, 2002). Interestingly, the main strength of this approach, its simplicity, can also be viewed as a weakness in that information is lost with dichotomization. This, of course, from a Transtheoretical Model perspective, could be seen as an insurmountable barrier for those interested in understanding the intervention in terms of each stage of change.

Research Question 4 focused on progress. This is a common approach with the stages of change construct and the subsequent interpretation that any forward movement through the stages as evidence for effectiveness (Bridle et al., 2005). However, this approach does not take into account different baseline stages and how individuals may differ. In other words, those progressing from Precontemplation to Maintenance (moving through all five stages) are treated the same as individuals moving from Precontemplation to Contemplation (one stage). From an analysis perspective, Research Question 4 shares the same benefits and challenges as Research Question 3 in terms of decision-making (i.e., considerations for binary logistic regression).

The treatment approach for the outcome variable in Research Question 5 relied on the theoretical perspective that the stages of change are considered ordered and discrete by design in the Transtheoretical Model (Littell & Girvin, 2002). Because theoretically these are not viewed as interval-level data in this treatment approach, ordinal logistic regression was used. Specifically, this analysis accounts for the theory that there are distinct stages within the stage of change construct and the distance between these stages is not equal. As described in Chapter Two, this approach may be one of the most
contentious issues because some analysts may argue that even though data are ordinal in nature, they can be treated as interval-level data given the robustness of parametric techniques (Norman et al., 2000). Others argue that this results in a reduction in face validity and affects interpretations of findings.

Summary

Both variable treatment decisions and subsequent analytic decisions were driven by the research question (and underlying theory), data limitations, and resource limitations. Although the importance of variable treatment and analytic decisions are evident in this dissertation, the primary driver of these decisions, particularly in this study and dealing with a complex construct like stages of change, is inextricably tied to the evaluation question for the intervention. Yet, it is evident making variable treatment decisions forces stakeholders to think through theoretical implications of their research questions more fully.

For example, Research Question 3 focused on achieving “action” in reference to the Action stage in the Transtheoretical Model yet another interpretation and subsequent treatment of the data could have been to define action as reaching the “Preparation” stage (defined as taking a behavioral step to act) (Prochaska et al., 2002). It follows that analytical decisions are primarily driven by variable treatment approaches (which are assumed to be informed by the research questions). However, there are still decisions that can be made at this stage in planning that can influence findings and interpretation of those findings. The chosen analytic approach should be practical, credible, and suitable for the question of interest (Green, Lipsey, Schwandt, Smith, & Tharp, 2007).
This is well-demonstrated through the comparison of Research Questions 1 and 5 because the same “number” is used as the outcome variable in the dataset but the first analysis approach uses parametric statistics to account for the interval level data and the latter uses non-parametric statistics to account for the ordinal level data. As evident in the research questions, the distinction between the research questions here gets at a core difference in theory for stage of change: the former that there are equal distances between stages and the latter that there are not. This is not something that is easily determined as evidenced in Chapter Two. However, the more important takeaway is whether, and how, the chosen approach best fits with the intervention of interest and whether it is appropriate to assume equal distances between stages for the given study.

The underlying theory also affects decisions surrounding data preparation for analysis. This may be most evident in decisions on missing data treatment or data transformations, which can ultimately impact the necessary sample size and ability to conduct certain analyses. For example, if a given study has a high number of responses that are deemed implausible because of the underlying theory (e.g., advancing four stages in a one month period), this can greatly affect the statistical analysis, particularly in terms of power. Of course, simple study considerations can also affect statistical method choice. For example, certain approaches, like logistic regression, require larger samples than ordinary least squares regression (Hair et al., 2010).

Methodological Decision-Making Considerations

The core issue in the use of evidence, regardless of the stakeholder, is the difference in decision-making imperatives (Choi et al., 2005). Scientific evidence can conflict with values and beliefs of a given stakeholder and so some may selectively look
for, and use, evidence to support claims (Choi et al., 2005). While not ideal (and potentially unethical in some cases), this reflects real-world conditions particularly in program evaluation settings where decision-making is typically stakeholder controlled (MacDonald et al., 2001).

The influences of judgments described above showcase the potential impact of prior experience, influence of the research question, and considerations around intended use of findings by stakeholders (Green et al., 2007). Given the array of approaches to interpreting and using findings, and the range of potential stakeholders of interest, there are a number of lessons learned from this dissertation that can be applied to the methodological decision-making process to yield credible findings within a given context.

These findings are primarily relevant to evaluators who generally hold primary responsibility for evaluation and research design decisions. The representative of the organization funding the study is also responsible for awareness of the implications of these methodological decisions to ensure that evaluation efforts are meeting the needs of intended users and thus stewards of those funds need to be attentive to these issues. These individuals need to be involved and understand the implications of variable treatment at a minimum. However, all stakeholders involved in using the evaluation findings should be at the very least knowledgeable about the implications of defining success or failure and subsequent approaches to variable treatment and analysis.

Clearly a pragmatic framework is needed to help guide decision-making in different situations and present factors that influence method choice (Julnes & Rog, 2007a). The following section details key considerations with respect to defining the
program logic and study design, measurement, and analytic decisions. These considerations are drawn from the literature based on key areas of interest in this dissertation. Of course, as noted in Chapter Two, final decisions, particularly in “real-world” evaluations are a combination of technical considerations, client and evaluator preferences, and other contextual factors (Bamberger et al., 2012).

**Defining program logic.** Defining what is considered “success” is critical to using information from a study and this should be done in the early stages of planning (as part of stakeholder engagement and describing the intervention and its intended outcomes) (MacDonald et al., 2001). This decision and the selection of an appropriate treatment for the outcome measure should be based on the most important expected effects for the intervention (Tunis, Stryer, & Clancy, 2003). As noted in Chapter Two, defining success can be particularly challenging with the stages of change construct (Bridle et al., 2005; West, 2005).

Evaluation questions are critical to the process because they inform evaluation plan development by defining components of the program to be addressed and thus pragmatic decisions around study design and data collection and analysis (MacDonald et al., 2006). There are established frameworks and standards for program evaluation that are intended to improve the conceptualization and conduct of evaluations among professionals (MacDonald et al., 2006).

The assumptions and purpose of conducting the evaluation or research study should be clear and documented. Assumptions may be based on program theory or evaluation theory or more broadly on scientific paradigms. For example, if an evaluator is following the principles of empowerment evaluation, this could affect methodological
decisions differently than someone being guided by the principles of utilization-focused evaluation (Donaldson, 2009).

The purpose can also vary (accountability focused versus promotion focused) or intend to serve multiple purposes. Variables and measures should be clearly tied to the evaluation or research questions (Donaldson, 2009). If the goal of the evaluation is to promote the use of the intervention, stakeholders may prefer to use “any progress” as a measure of success. Further, the approach to treatment or classification of the variable should be carefully weighed to ensure appropriate alignment with the evaluation questions and underlying assumptions. For example, if the evaluation question focuses on making a decision to adopt the intervention, variable treatment would focus on the stage where that decision takes place (Action).

Thus, formulating and prioritizing evaluation (or research) questions is a critical component that can drive methodological decision-making. Chelimsky (2007, p. 22) noted “it is impossible to assign a priority of importance to a method because methods depend on questions, which cannot be ranked, and all methods present advantages and disadvantages relative to a question posed.” This underscores the importance of working with stakeholders during the study design to clearly define the logic of the program, through tools like logic models, so success is clearly defined as well as the elements that are expected to influence the outcome (Julnes & Rog, 2007b). Stakeholders in the federal space generally support this practice of defining program logic. Many government agencies encourage the use of logic models to define and illustrate relationships among program elements and specify how success is defined (Datta, 2007).
The nature of the program or intervention would also influence study design and evaluation question considerations (Datta, 2007). Considerations include how often the intervention needs to be implemented (one-time or a series of intervention points), how quickly it is hypothesized to have an impact on the outcome of interest, whether the intervention can ethically be withheld from a control or comparison group, and the intended population for the intervention. For example, with the stage of change variable and treatment decisions, the anticipated effect of the intervention and the hypothesized time range would influence the evaluation questions and interpretation of findings around whether movement across multiple stages is reasonable.

Study design, measurement and analytical method. One key consideration in study design, measurement, and analytical considerations is the degree of desired confidence in conclusions (Julnes & Rog, 2009). Programs where there are just considerations around incremental program changes may require less confidence in findings than situations where decisions are being made around continuing funding for a given program or deciding to expand the use of an intervention (Julnes & Rog, 2009). This is reasonable given that it is easier to reverse incremental changes if subsequent evidence provides different or contradictory information or conclusions (Julnes & Rog, 2009).

Stakeholders in the federal space are often focused on establishing precise estimates around the magnitude of program effects because these estimates are important to agencies like the Office of Management and Budget (an actuary of policy change) (Julnes & Rog, 2009). The precision of estimates links to considerations of sample size as well as other key variables identified in defining the program logic (Greenberg, 2007).
A range of analytical approaches is available to evaluators and researchers. As noted in Chapter Two, a critical component in making measurement and analytical decisions is aligning the methods with the research or evaluation questions of interest. The methodology chosen should be practical, credible, and suitable for the research questions (Green et al., 2007). Inherent in this choice are judgments rather than simply relying on evidence.

In fact, using a hierarchy of evidence for methodological choices can undermine this alignment (Green et al., 2007). One methodological approach may be considered “the best” or a “gold standard” but may be expensive and or difficult to implement (Green et al., 2007). If this is the case and people continue to attempt to implement them with unsuccessful results, stakeholders may conclude there is no evidence or that evaluations are not worth the resources invested in them and undermine needed support for the method and/or evaluations in general (Green et al., 2007).

The politics around methodology can also influence methods choice (Datta, 2007). An individual’s scientific paradigm, influenced by personal experience and professional standards, can affect decisions on superiority of approaches to yield credible evidence (Chelimsky, 2007; Donaldson, 2009). This is illustrated in Chapter Two with measurement and analytic debates such as ordinal versus interval level data or the use of single-item versus multiple-item scales for stages of change. Just like agency culture may favor one approach over another, so may the evaluator based on his or her training and experience (Datta, 2007).

In addition to individual preferences, it is critical for an evaluator or applied researcher to understand agency preferences in terms of methodological approaches. For
example, an agency may prefer using mean scores in characterizing stages of change because it is consistent with other studies within the agency. This may influence not only the choice of methods but the initial evaluation questions of interest. Neglecting patterns of practice within the funding agency can therefore also affect eventual use of findings (Julnes & Rog, 2007b). If a method does differ from the typical practice of an agency (e.g., the use of ordinal logistic regression), it would benefit the evaluator to clearly explain why a deviation from the typical approach is needed to answer the questions of interest.

There is not an established research model for program evaluation given the challenges, and sometimes inappropriate nature, in non-academic settings (MacDonald et al., 2006). There are a number of political, time, or resource constraints that can impact a study and these should be documented along with the study assumptions (Bamberger et al., 2012). These can be used to inform weighing the pros and cons of alternate approaches to yield the most credible evidence given the contextual factors (Donaldson, 2009). For example, budget and time constraints can greatly influence the ability to obtain a large sample size. As noted in Chapter Four, this affects sampling error which affects the precision of estimates.

Another example is staffing resource constraints. Depending on how the stages of change construct is measured and the anticipated patterns of change defined in the program logic, more complex statistical procedures may require staffing capabilities that are beyond the scope of the evaluator (or his or her evaluation team). In this case, the evaluator might need to adjust the scope and design of the evaluation (e.g., collect
supplemental qualitative data instead) to answer the evaluation questions of interest given the resource constraints.

**Questions to Consider**

Table 5.1 presents questions specific to methodological decision-making for a stage of change variable based on the two main areas described above: defining program logic; and study design, measurement and analytical method. These questions can and should be used at the planning stage of the evaluation by different stakeholders to inform the most appropriate design. In addition, these questions are useful in facilitating the thoughtful interpretation of findings and how variable treatment decisions may have affected these findings after analyses have been completed.
Table 5.1 Sample Considerations for Methodological Decision-making for a Stage of Change Outcome Variable

<table>
<thead>
<tr>
<th>Area</th>
<th>Questions for Consideration</th>
</tr>
</thead>
</table>
| Defining Program Logic        | • How is the program logic defined? Does the program logic support or reflect a particular variable treatment?  
• What is the intent of the evaluation (i.e., what aspect of program performance needs to be examined)?  
• How is the outcome of interest defined (e.g., a specific stage, progression)?  
• How should the outcome of interest be measured?  
• What change is expected in this outcome? Are there anticipated patterns of change?  
• What is the timeframe for the intervention and when is it expected to affect stage of change?  
• Would this vary by stage of change and if so, how?  
• What is the intervention and how it is expected to affect change in the outcome variable? Is the intervention stage-based in nature?  
• Are there other key variables that are necessary for change? How are these defined and measured? |
| Study Design, Measurement, and Analytical Approaches | • How are other key variables defined and measured?  
• How will these be collected and when?  
• What is the needed power to conduct the study?  
• What is the feasibility of obtaining useful conclusions on intervention effects?  
• What level of confidence is desired?  
• What external factors exist that may affect the ability to conduct the study (e.g., budgetary limitations, political constraints)?  
• Are there any ethical considerations?  
• What analytical approaches can best answer the evaluation questions?  
• Are the data interval or ordinal level?  
• What is the distribution?  
• What resources are available to conduct the evaluation (e.g., existing staff, skill sets, statistical software availability)? |
Chapter Six: Practical Application

A key component of D&I research, and evaluation studies in general, involves collaboration between researchers and stakeholders to ensure research outputs are relevant and useful for decision-making (Velentgas, Mesters, Van Mechelen, & De Vries, 2013). This chapter presents examples of methodological considerations using two hypothetical stakeholders: a federal program officer and an evaluation consultant serving in a principal investigator role for a federal consulting firm. These are illustrative of the types of stakeholders and considerations but certainly do not encompass the full range of stakeholders or considerations inherent to individual and organizational preferences and constraints. This section also presents potential interpretations and use of evaluation findings based on the decisions made for Research Questions 1 through 5. Relevancy of study findings to research, evaluation, and measurement more broadly is also discussed.

Defining Program Logic

A key component of a recent methodological report to come out of PCORI prioritized measuring outcomes that people notice and care about and specified that these outcomes should be clearly defined and relevant to decision-makers (PCORI, 2012). For example, a federal program officer stakeholder may have specified that the purpose of the intervention is to facilitate consideration of using more evidence-based programs. Progress (or Research Question 4) might be a better indicator than action (Research Question 3) as even earlier stages in the model indicate consideration of whether the intervention is a good fit. Alternatively, the federal program officer may value the use of the Transtheoretical Model and be more interested in defining success in terms of stage of
change. In this case, Research Question 2 (taking into account the number of stages progressing or regressing) or Research Question 5 (ordinal categorical responses) might be better aligned with addressing the intended use.

This debate can become even more complicated when taking into account other expectations and constraints. Specifically, the agency’s administrative culture and priorities could drive a federal program officer’s variable treatment recommendations (Hanney, Gonzalez-Block, Buxton, & Kogan, 2003). This could result in “any progress” being seen as a successful outcome for the intervention and thus the preference would be for a research question focusing on the presence or absence of progress (Research Question 4).

Another consideration is a situation where the federal program officer (or his or her leadership) may only have an interest in one type of group at baseline (e.g., those in the Precontemplation stage). This interest would help define and inform what is considered success for the given intervention for that group. Specifically, focusing on Research Questions 2 (number of stages), 4 (progress or no progress), or 5 (ordinal categorical) might be the most appropriate questions to consider.

In the research and evaluation fields, it is critical that stakeholders and decision-makers have a clear understanding of goals prior to defining outcomes of interest and how these outcomes should be characterized. Focusing on intended use will also avoid overburdening respondents because data collection will be limited to information needed for the specific use (MacDonald et al., 2006). Then, there should be a clear rationale for approach to classifying outcomes and this rationale should be discussed in conjunction with key stakeholders and decision-makers.
Study Design, Measurement and Analytic Approaches

It is likely that the client will look to the consultant as the “expert” on the best design and analytic approach for an evaluation study (Bamberger et al., 2012). Yet, factors influencing decisions on preferred variable treatment and analytical approaches can vary greatly between the consultant and the federal program officer. For example, the consultant may have ideological biases around the “correct” methodology to use which can impact methodological decisions and ultimate use of findings (Bamberger et al., 2012). The federal program officer, operating under greater fiscal, administrative, and political constraints than the consultant, would likely weight expert recommendations with other contextual factors as the desire for scientific rigor is often only one consideration for policymakers (Bamberger et al., 2012; Choi et al., 2005). This may depend on the Federal agency as certain agencies have standards around “rigor” (Julnes & Rog, 2009).

As noted in Chapter Two, measurement considerations are also critical for conducting evaluation studies. Therefore, an evaluator with a more conservative methodological approach may favor Research Question 5, which preserves the discrete nature of the stages of change and does not assume interval-level data.

Practical constraints, like funding, can influence recommendations around analytical approaches. If the federal program officer values simplicity, he or she may recommend the consultant utilize a less complicated analytical approach so it is easier to explain to his or her leadership. As noted in the limitations in Chapter Five, this would favor Research Questions like 1 (posttest score) and 2 (number of stages) that have continuous outcome variables and utilize linear regression models. Although simpler
methodological approaches are not necessarily less scientifically rigorous, not using a more complex approach when the research question calls for it could compromise findings from the study. For example, as demonstrated in this study, if the evaluation question of interest is focused on reaching the Action stage, analyzing the posttest score (because the ordinary least squares regression is considered simpler to conduct) would not answer the main question of interest as directly as other approaches like binary logistic regression.

Interpreting and Using Evaluation Findings

*Interpreting findings*. Chapter Four detailed specific methodological factors to compare findings from the five research questions. Each of the research questions resulted in statistically significant differences between the Intervention and Comparison group although other statistics provide additional detail to aid in interpreting the findings. For example, Research Question 3 (action) yielded the highest $R^2$ value of 16.2 percent, almost double that of Research Question 4 (progress). The consultant would be more likely to give heavier weights to such a figure in interpreting the finding while the federal program officer may or may not be as interested in these statistics.

As noted by Lavis, Robertson, Woodside, McLeod, and Abelson (2003), “decision makers rarely use a regression coefficient to help them solve a particular problem” (p. 223). Depending on the federal program officer’s background, there are often varying attitudes around the use of evidence, and what constitutes “good” evidence, for policymaking and programmatic decisions, which leads some stakeholders in the policy environment to be more receptive to research findings than others (Choi et al.,
That being said, scientific rigor and objectivity are both principles that are often highly valued by those in the research and policymaking communities (Hanney et al., 2003).

It is often the responsibility of the consultant to distill these data in a meaningful way to explain to the program officer in plainer terms the implications of the statistical findings. That is not to say the federal program officer is not interested in these statistics but given time constraints and competing commitments, these stakeholders generally do not have the time to investigate and interpret these data.

Chapter Four listed several limitations of each approach in pragmatic factors. These limitations are an area to which the federal program officer should pay particular attention. For example, one of the major limitations of Research Questions 3 and 4 is that dichotomizing the outcome measure loses critical stage information and thus limits the interpretations that can be made (focused just on action/non-action or progress/no progress). Limitations can often be overlooked, either as a product of too much information or because the limitations are at odds with other political factors (Hanney et al., 2003).

*Using evaluation findings in practice.* There are a number of potential uses of evaluation findings. Broadly, many evaluations focus on demonstrating the effectiveness of a given program, informing program improvement, demonstrating accountability, and serving as justification for funding (MacDonald et al., 2001). It is often the case where there are multiple intended uses. This may include a combination of justifying funding by promoting positive findings on effectiveness from the study and informing programmatic decisions such as encouraging use of the intervention among staff (MacDonald et al.,
2001). As noted in Chapter Two, particularly with utilization-focused evaluations, these intended uses would be specified during the planning stage and accounted for in the evaluation design (MacDonald et al., 2001; Patton, 1994).

For example, a federal program officer may be focused on informing budget allocations for the following year and/or justifying the use of funds on the intervention. In considering this use, Research Question 3 may be the most impactful in that the federal program officer would be able to say the Intervention group resulted in a higher likelihood for “action” than the Comparison group. For funders, this may be more impactful than a focus on movement from Precontemplation to Contemplation (captured as success in Research Questions like 1 and 4).

Another area of use would be focusing use of findings on program promotion or improving the intervention’s image. This would be a common intended use among a federal program officer given the amount of funding often dedicated to these types of studies and intervention development. However, this may be potentially frustrating to a more conservative consultant especially given the potential for attacks on credibility (MacDonald et al., 2001). However, as long as these expectations are made clear in the planning stages and clearly communicated in presenting findings, this is a valid use of evaluation findings, particularly in utilization-focused evaluations (Patton, 1994).

Study findings may also indirectly benefit policy by informing policy discussions or approaches to funding future grants (Elliott & Popay, 2000). For example, the federal program officer may recommend using a specific approach to categorizing stages of change as an initial screening tool for grant applications (e.g., only those categorized as “action” will be given funding). Another area of use may be informing guidance around
data management systems. For example, based on study findings, the federal program officer may influence policy by requiring monitoring systems for the evidence-based program translation process with a special emphasis on decisions to adopt (Durlak & DuPre, 2008).

Although identifying “success” is certainly important, it may also not be the ultimate intended goal of the study. The goal may be to simply improve understanding of the impact of the intervention. In this case, comparing actual outcomes versus intended outcomes can be used to make needed programmatic adjustments in the future such as testing an alternative intervention or changing certain components of the intervention in the study (MacDonald et al., 2001). This would still involve variable treatment decisions but choices may be less political than other scenarios.

Along these lines, this may also inform a potential use of what outcomes are considered “realistic.” This is particularly relevant with a stage of change variable where the time needed to progress through the stages is often of interest. Because the decision to adopt an evidence-based practice is not a commonly used outcome, this may be particularly useful to explore within stages of change to understand timing expectations in terms of movement across stages.

**Competing findings or projects.** One of the biggest disconnects between the consultant and the federal program officer may be the degree to which they weight the impact of research findings. For a consultant, he or she may strongly believe there should be actions stemming from study findings alone and might be willing to make a decision around the intervention based on the effects shown in the analyses (e.g., the Intervention group performed significantly better than the Comparison group in terms of reaching the
Action stage). Thus, decisions around variable treatment would need to take this into account. However, the political context for a federal program officer often requires that he or she consider the impact and feedback of multiple stakeholder groups which can affect decision-making based on study outcomes (Choi et al., 2005).

It is likely that for the federal program officer, the study is just one of many under consideration. For example, he or she may be overseeing four studies, each examining different interventions and their impact on decisions to adopt an evidence-based program. The federal program officer may suggest variable treatment approaches that align with other studies or may choose alternate approaches if he or she is more interested in different approaches to assessing the outcome variable. In addition, because it is likely inputs from different studies are stemming from different groups, the federal program officer may be receiving disconnected, or conflicting, advice and thus an increased burden to make sense of findings (Brownson et al., 2011). This challenge for federal program officers also stresses the importance of presenting findings as simply and concisely as possible to assist in the likelihood of use (Brownson et al., 2011).

Alternatively, there may be other types of feedback being received (e.g., personal anecdotes, cost information) by the federal program officer about which the consultant may not be aware. For example, the federal program officer may be overseeing another project where a different intervention is being used for the same purpose and showing even better outcomes for a reduced cost. Therefore, decisions around variable treatment may be focused on defining “progress” as success as opposed to reaching a specific stage to show more improved outcomes and thus cost savings.
More broadly, beyond the specific uses related to decisions around the intervention of interest, the consultant could use findings from across the five research questions to inform future study designs and help communicate with his or her clients around the importance of measurement decisions on study designs and outcomes (Durlak & DuPre, 2008). For example, in the study design phase, discussions on the methodological approach could be better focused on the client’s goals for the intervention to ensure more appropriate measures (e.g., do you want participants to just show stage progress or move them to action?). This exchange on aligning goals and measures would keep the client engaged throughout the process and increase the likelihood that he or she would be able to use the information (Lewis, 2011).

Finally, and this issue is especially important in utilization-focused evaluations, the research results need to be communicated clearly to inform decision-making. Evaluators need to provide “big picture” takeaways without “dumbing down” results (Choi et al., 2005). This includes making the strengths and limitations of a study design clear so tradeoffs are apparent and stakeholders can better engage in decision-making. For example, the consultant could use these study findings to communicate how dichotomizing the outcome variable into action or non-action can simplify analyses and interpretation but also can result in a loss of stage information for respondents. By improving education and understanding of how different methods affect the conduct and evaluation of research, with studies like this dissertation, both evaluators and the public can improve use of research findings (PCORI, 2012).
Summary

Broadly, this hypothetical illustration showcases some of the distinguishing characteristics between how a federal program officer and a consultant may make decisions and interpret findings in Research Questions 1 through 5 in the context of different political, data, and resource constraints. Depending on an individual’s role, background, and organizational priorities, the decision-making process and interpretation of findings could vary. Therefore, this comparison is simply a device to illustrate two examples of stakeholders and how complicated study decisions and use of findings can vary based on the existing dataset.
Chapter Seven: Conclusion and Future Research

Chapter Five detailed a more complete discussion of Research Question 6 and the implications for stage of change variable treatment. In addition, the chapter provided some considerations for informing methodological decision-making based on study findings and the literature base more broadly. Chapter Six then presented a practical application using two illustrative examples to demonstrate how decisions and use of findings might vary across two different stakeholders. Both of these chapters stress the importance of stakeholder values throughout the evaluation cycle. Although the importance of values is emphasized in existing literature around the use of findings, particularly in utilization-focused evaluation, this dissertation stresses the significance of these personal values in earlier stages like asking the evaluation questions, determining how to categorize constructs, and assigning value to the resulting codes. This final chapter presents overarching conclusions based on information presented in Chapters Four and Five, a discussion of limitations, and directions for future research based on study findings.

Conclusion

Clearly questions remain in the field around “appropriate” study design, measurement, and subsequent analyses and reporting. As seen in earlier chapters, some tend to oversimplify opposing positions in methodological debates, like ordinal versus interval level data, which can lead to unproductive dialogue around what is best in certain circumstances and ignoring inherent gray areas (Julnes & Rog, 2009). Yet it is this
dialogue that can help advance thinking around methodological decision-making in real-
work contexts. Continued research on evaluation use may help provide some insight here.

This dissertation provides a first step in this direction by demonstrating some of
the methodological and pragmatic factors that can differ using the same dataset. While
the results were not drastically different across the five research questions (i.e., the
Intervention group showed a statistically significant improvement in each case), the
findings were substantial enough to warrant concern about always using one approach
without thoroughly thinking through the situation. This was particularly the case with
practical significance and the importance of defining success in the evaluation questions
to ensure findings are useful for stakeholders. In addition, it presented the very real
problem that many evaluations face in that certain approaches can be more limiting in a
practical context (e.g., ordinary least squares regression may be much more common in
certain fields and easier to interpret than the ordinal logistic regression).

Further, given the prominent use of the stages of change construct in the literature
and the importance of the Transtheoretical Model in health research, stages of change
continues to be a critical outcome measure concept used in various applied research and
evaluation studies. This dissertation provided insight into some key methodological
decision-making considerations based on one approach to assessing the stage of change
outcome variable.

Because of the focus on credible evidence in the field of applied research and
evaluation, continued dialogue and research is needed to help guide the choice of
methods, particularly in federal evaluations (Donaldson, 2009; Julnes & Rog, 2007b).
This includes improving the understanding around how different assumptions and
contextual factors can affect stakeholder use of evidence (Donaldson, 2009). With continued work and research on how methodological decisions may affect research findings, both evaluators and stakeholders will benefit.

Practical Recommendations

Based on the findings from this dissertation, there are a few recommendations specific to evaluation planning that may yield more credible evidence for evaluators and their stakeholders. The first is for evaluators to work closely with stakeholders to clearly define the program logic and specify what would be considered successful outcomes. This does require time, and patience, from both parties but will facilitate measurement and analysis discussions during evaluation planning. This also ensures expectations are managed for both the evaluator and the stakeholders. The second recommendation is to review the literature. While this is common practice in many research studies (in terms of defining the need for the program), a review of literature attuned to measurement considerations based on the construct of interest, and in light of evaluation constraints, is a critical step. This includes ensuring whether the underlying theory or framework is still relevant and appropriate to the construct of interest. As noted in Chapter Two, even the Transtheoretical Model has undergone significant changes since it was first introduced and is not applicable to all health behaviors, which has implications for decision-making.

The third recommendation is to ensure the analysis plan aligns with the evaluation questions of interest and therefore the underlying program logic. This seems like common sense but often the analysis plan can become disconnected from evaluation questions once surveys are being developed or interviews are conducted. The analysis
should be grounded in the evaluation questions and data treatment decisions should be based on the defined conceptualization of the outcome variable. If questions remain around data treatment, it would be valuable to test the differences between the variable treatments.

It should be noted that testing for these differences should only be done if needed to inform decision-making and attention should be paid to avoid “hypothesis fishing” where a range of tests are conducted for a given hypothesis (Dahl et al., 2008). This dissertation focused on five approaches to demonstrate a range of potential outcomes and implications of decisions – the intent was not to inform programmatic decision-making about the intervention being studied (which would be a likely hypothesis in a real-world evaluation situation). However, in a practical situation, earlier recommended evaluation planning steps like defining success or the measurement approach would narrow down data treatment and analysis options. If a lingering concern remained, such as whether treating the data as ordinal or interval may affect study findings, it would then be appropriate to test this and make needed adjustments as appropriate (such as applying the Bonferroni correction) (Dahl et al., 2008).

The analysis plan should also be revisited after data are collected to ensure the plan is still appropriate. For example, if all of the respondents fall into one category, this can affect plans and subsequent interpretation. Finally, the importance of communication with stakeholders cannot be understated. This may even involve informal training where appropriate so they become more empowered to question or understand methodological decisions and understand results. The importance of communicating limitations also remains a concern, particularly in evaluation, and the impact on findings. These should
not be lost in the discussion and should be incorporated in reports, presentations and conversations.

**Limitations of the Study**

This dissertation has several factors that limit the interpretations and the generalizability of the findings related to if, and how, findings differ in this study based on variable treatment, and subsequent analysis, decisions. These limitations are related to three main areas including design, methods, and sample detailed below.

The exclusion of any contextual variables in the analysis (such as respondent role or organizational resources that could affect decision-making around an evidence-based program) could be considered a key study design limitation. Contextual variables are a crucial component of program evaluation and a main difference between research and evaluation in that contextual variables as seen as essential information as opposed to something to be controlled in a statistical model (MacDonald et al., 2001). Further, simplistic designs that do not include moderators and mediator variables are not well-suited to evaluate complex programs or interventions (Julnes & Rog, 2007b). Although the exclusion of contextual variables was necessary for the study design because the focus was on measurement and impact on evaluation use, this is a notable consideration as most studies necessitate the inclusion of contextual variables for more real-world applicability and improved predictive models (e.g., influencing the size of the intervention effect). Contextual variables could also affect decisions around variable treatment, particularly if the program logic differs depending on certain types of groups (e.g., an administrator may be more likely to reach the Action stage than a non-administrator).
The main limitation related to methods is the use of stakeholder illustrations in Chapter Six. These illustrations are based on the literature and common reactions, or approaches, to interpreting or treating data rather than primary data collection activities. This limited the ability to generalize findings to a specific stakeholder group or audience. Collecting qualitative data from a range of stakeholders on their perspectives for variable treatment and evaluation use in this study would have yielded more information around decision-making choices. However, given the number of contingencies and contextual factors noted throughout this dissertation, a qualitative sample would simply provide more illustrations of this process rather than definitive guidance on decision-making.

Finally, this study relied on an existing sample as a way to study variable treatment with a stage of change outcome variable. Thus, the study is limited in that variable treatment decisions are considered in the context of decisions to adopt an evidence-based program among community-based health organizations. Using stage of change for this construct is reasonable but is not as well-studied or utilized as other constructs like tobacco use or healthy eating habits. Another limit to generalizability in this context relates to the approach for measuring stage of change in data collection. The original study used a single-item to measure stages of change as the outcome variable but as detailed in Chapter Two, comparisons among approaches like multi-item algorithms or scales may yield different findings. Examining additional datasets would provide more evidence for generalizability of findings and potentially other considerations that are key to methodological decision-making. However, this is not a major limitation as decisions about variable treatment are not solely dependent on a given theory of change or the variable being measured but rather a range of factors as demonstrated in this study.
Directions for Future Research

There are several areas where future research would be beneficial to improve the generalizability of findings. The first is to explore the impact of methodological decision-making using data from studies that measure stages of change through approaches other than a single-item scale (e.g., multiple item scale, algorithm). The use of the single-item scale was a result of the existing dataset used from the original study. Comparisons among approaches for treating stage of change, such as other algorithms or multi-item scales, may yield different findings. As described in Chapter Two, there are numerous approaches to assessing stage of change and categorizing it as an outcome variable, so expanding this analysis to other assessments could be a valuable addition to this discussion.

The second recommended area of future research is to explore methodological decision-making around stage of change with other types of study designs. For example, while this study focused on a clustered, randomized controlled trial design, qualitative and mixed-methods designs are quite common in evaluation and would require similar investigation into methodological decision-making and use. This would require an examination of both methodological and pragmatic criteria adapted from Research Question 6 which may need to be adapted based on the type of study (Julnes & Rog, 2009). Specifically, the incorporation of qualitative data may provide interesting insights into the stage of change construct and influence variable treatment decisions. For example, a review of qualitative data may reveal a lack of discrete differences between
two stages of change for the construct of interest and treating data as categorical could therefore be inappropriate.

The final suggested area for future research involves investigating more fully the impact of evaluating evidence in policy settings. Even if evaluators and researchers could agree on the most appropriate methodological designs in certain contexts, policymaking requires more than considering evidence from just one study (or a collection of studies) (Green et al., 2007). Given the political environment, it would be interesting to engage federal workers in the policy sphere to get their perspectives on credible evidence and methodological approaches.

Summary

This dissertation highlighted differences in approaches for classifying and treating a stage of change variable and demonstrated the impact decisions on approaches can have on study findings. Moreover, the role of different stakeholders’ values and preferences paired with practical challenges in real-world evaluations were emphasized as key influencing factors for methodological decision-making as well as eventual use of study findings. The proposed directions for future research seek to continue advancing this understanding of the impact of methodological decisions in different contexts and help improve the utility of evaluations more broadly.


Horwath, C. C., Schembre, S. M., Motl, R. W., Dishman, R. K., & Nigg, C. R. (2013). Does the Transtheoretical Model of behavior change provide a useful basis for interventions to promote fruit and vegetable consumption? American Journal of Health Promotion, 27(6), 351.


The Joint Commission (2013). National Patient Safety Goals effective January 1, 2014: Hospital Accreditation Program. The Joint Commission. Available at:


*Introduction to program evaluation for comprehensive tobacco control programs.* U.S. 
Department of Health and Human Services, Centers for Disease Control and Prevention, 
National Center for Chronic Disease Prevention and Health Promotion, Office of 
Smoking and Health. Available at: 
http://www.cdc.gov/tobacco/tobacco_control_programs/surveillance_evaluation/evaluati 
on_manual/pdfs/evaluation.pdf.

Adopting evidence-based practice in clinical decision making: nurses’ perceptions, 


Practice*, 20(3), 368-375.

practices and readiness to adopt evidence-based practices: Dissemination research in an 
adiction treatment system. *Journal of Substance Abuse Treatment*, 26(4), 305.


(2006). Assessing the relationships between contextual factors and research utilization in 

on Quantitative Applications in the Social Sciences, series no. 07-106). Thousand Oaks, 
CA: Sage.

Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of 

398-407.

to translate evidence-based programs to community settings: bridging the gap between 
research and practice. *Health Promot Pract*, 13(4), 559-566.

Mock, T.J. (1972). A decision tree approach to the methodological decision process. *The 


Patient-Centered Outcomes Research Institute (PCORI). *Public comment draft report of the Patient-Centered Outcomes Research Institute Methodology Committee* presented on July 23, 2012, and revised thereafter.


John Wiley & Sons, Inc.


Van Keulen, H. M., Mesters, I., Van Mechelen, W., & De Vries, H. (2010). Single-item and multiple-item measures of adherence to public health behavior guidelines were incongruent. *Journal of Clinical Epidemiology, 63*(1), 75-84.


