

Social Work 3402 Coursebook

Social Work 3402 Coursebook

AUDREY BEGUN



Social Work 3402 Coursebook by Audrey Begun is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/), except where otherwise noted.

Contents

| | |
|---|---|
| Preface & Orientation | 1 |
|---|---|

[Module 1](#)

| | |
|---|----|
| Module 1 Introduction | 5 |
| Module 1 Chapter 1: Review and Overview | 7 |
| Module 1 Chapter 2: What is Evidence-Based Practice? | 18 |
| Module 1 Chapter 3: Practice Evaluation as Evidence | 35 |
| Module 1 Chapter 4: Ethics Related to Intervention Research | 46 |
| Module 1 Conclusions | 61 |
| Module 1 Key Terms and Definitions | 62 |
| Module 1 References | 65 |

[Module 2](#)

| | |
|---|-----|
| Module 2 Introduction | 71 |
| Module 2 Chapter 1: Course Review and Overview | 73 |
| Module 2 Chapter 2: Formulating Practice Evidence Questions | 76 |
| Module 2 Chapter 3: Logic Models and Theory Informed Intervention | 91 |
| Module 2 Chapter 4: Locating Intervention Evidence | 98 |
| Module 2 Chapter 5: Integrating, Implementing, and Evaluating | 121 |
| Module 2 Conclusions | 140 |

| | |
|---|-----|
| <u>Module 2 Key Terms and Definitions</u> | 141 |
| <u>Module 2 References</u> | 143 |

[Module 3](#)

| | |
|---|-----|
| <u>Module 3 Introduction</u> | 149 |
| <u>Module 3 Chapter 1: Overview of Intervention/ Evaluation Research Approaches</u> | 151 |
| <u>Module 3 Chapter 2: Quantitative Design Strategies</u> | 176 |
| <u>Module 3 Chapter 3: Measurement</u> | 211 |
| <u>Module 3 Chapter 4: Participant Recruitment, Retention, and Sampling</u> | 256 |
| <u>Module 3 Summary</u> | 278 |
| <u>Module 3 Key Terms and Definitions</u> | 279 |
| <u>Module 3 References</u> | 285 |

[Module 4](#)

| | |
|--|-----|
| <u>Module 4 Introduction</u> | 293 |
| <u>Module 4 Chapter 1: Review and Extension of Descriptive and Group Comparison Analyses</u> | 295 |
| <u>Module 4 Chapter 2: Paired t-Test</u> | 314 |
| <u>Module 4 Chapter 3: Analysis of Single System Design Data</u> | 323 |
| <u>Module 4 Summary</u> | 343 |
| <u>Module 4 Key Terms and Definitions</u> | 344 |
| <u>Module 4 References</u> | 346 |

[Module 5](#)

| | |
|--|-----|
| <u>Module 5 Introduction</u> | 349 |
|--|-----|

| | |
|--|-----|
| <u>Module 5 Chapter 1: Presenting Evidence for Understanding Social Work Interventions</u> | 351 |
| <u>Module 5 Course Conclusions</u> | 360 |
| <u>Module 5 References</u> | 371 |

Welcome to the online coursebook for SWK 3402 *Research & Statistics for Understanding Social Work Interventions*. There are several tools built into these chapters, designed to enrich and guide your learning.

- Material may be reviewed in three different formats: online (EPUB), downloaded (PDF), or as Mobilepocket (MOBI) format, which is better suited to small-screen devices.
- Embedded interactive “Stop and Think” exercises require internet connectivity, but each can also be downloaded for offline work. These exercises are presented to help you apply what you are reading, challenge yourself, prepare for quizzes, and have a little fun along the way.
- A list of key terms is presented at the end of the module coursebook to explain text appearing as **highlighted in bold italics** throughout the coursebook—in the interactive mode you can click on a highlighted word to jump to its explanation in the key terms section. Use the back arrow to return to where you were reading.
- “**Back to Basics**” textboxes appear in some chapters of some modules. These are designed as refreshers of some concepts that you may have learned in the past and that have applicability to the present material.
- **Green highlighted** boxes appearing in some chapters/modules refer you to learning activities in your online interactive Excel Workbook. These learning activities relate to content introduced in your main coursebook. The intent is for you to complete each Excel Workbook activity as you encounter it, then return to the coursebook.

Should you find it necessary to cite portions of the coursebook, here is the recommended citation: Begun, A.L. (2018). *SWK 3402 online coursebook: Research & statistics for understanding social work*

interventions. The Ohio State University, Open Educational Resources.

MODULE I

Module 1 Introduction

Our first module in this second of our two-course sequence allows us to become oriented to the focal topic of the course: *Research and Statistics for Understanding of Social Work Intervention*. This first module extends concepts learned in the first course (SWK 3401: *Research and Statistics for Understanding Social Problems and Diverse Populations*). Major research and empirical evidence concepts are placed in the context of social work practice and social work intervention at all levels of practice. The topic of research ethics is revisited, placed in the context of intervention research. This attention to intervention research is critically important to social work practice because good intentions are not enough: sometimes interventions are helpful, sometimes they fail to be helpful, some interventions are more efficient than others, and sometimes they actually are harmful. Therefore, social workers have an ethical responsibility for accountability and to engage with evidence to ensure that the best possible outcomes are realized when we intervene. Our entire SWK 3402 course is about learning to ask and answer these types of questions about social work intervention.

READING OBJECTIVES

After engaging with this module's reading and learning resources, you should be able to:

- Relate major concepts from SWK 3401 to social work intervention research;
- Understand the principles of evidence-based and evidence-informed social work practice;
- Identify the role of evaluation in social work practice;

- Explain ethical issues that might arise specifically with intervention research studies;
- Define key terms related to social work research and research ethics.

Module 1 Chapter 1: Review and Overview

A major difference between social work and many other disciplines is our profession's emphasis on intervening to help create or influence change. Social workers go beyond trying to understand social issues, social problems, social phenomena, and diverse populations: we also apply this knowledge to tackle social problems and promote social justice. This chapter provides a context for understanding whether the things we do in social work practice are helpful. We begin with a brief review of major concepts learned in the first course that have direct application to the learning objectives of this second course, SWK 3402.

In this chapter you will:

- Revisit key concepts from SWK 3401 related to aims of SWK 3402;
- Explore basic concepts about social work intervention;
- Discover how empirical evidence fits into social work practice.

Review of Key Concepts

It is impossible to rehash all the important material from our previous course. However, it is helpful to revisit a few concepts with great relevance to our current course: Research and Statistics for Understanding Social Work Intervention. Here are nine concepts from SWK 3401 and how they relate to our new course, SWK 3402.

- The first important set of concepts that carry over from our previous course is the role played by empirical evidence and critical thinking in the social work profession (Module 1). This

theme carries through our current course, SWK 3402. We will see, once again, the importance of this form of knowledge for informing professional practice decisions, as well as for evaluating intervention impact.

- Among the topics explored in SWK 3401 was the set of relevant standards in our professional Code of Ethics (Module 1). The points concerning the social workers' Code of Ethics continue to provide an important context for understanding and appreciating the role played by research and evidence in understanding social work intervention.
- You learned about the ethical and responsible conduct of research, including what is discussed in the NASW Code of Ethics (Module 1). The points from Standard 5.01 (d) through (q) all continue to have great relevance as we explore intervention research in our current course.
- The role of theory in developing social work knowledge represented another key topic from SWK 3401 (Module 2). As you will see in SWK 3402, theory continues to play a significant role as we think about social work intervention.
- Another topic that retains its relevance between our previous and current courses involves developing a research question (Module 2). While this content is relevant to designing and assessing research for and about social work intervention, skills related to “causal” questions are relevant for practice, program, and policy evaluation, as well.
- A great deal of emphasis in our current course is dedicated to locating and assessing evidence available in the research literature. What you learned in our earlier course in relation to understanding social work problems and diverse populations (Module 2) is directly applicable to locating and assessing evidence about and to inform social work interventions.
- In our present course, we continue to extend what you learned about research approaches and study designs (Module 3) to the design and implementation of research studies. This includes topics related to study participants and measuring

variables of interest (research methodology). As before, we explore how each study design decision is related to the nature of the research questions being addressed.

- The statistical analysis lessons learned in our SWK 3401 course (Module 4) continue to apply in analyzing data related to social work intervention. While some of the same analytic approaches are relevant, we learn new statistical approaches that are particularly appropriate for answering research questions about social work interventions.
- Finally, we extend in our current course what you learned about presenting evidence about social work problems and diverse populations (Module 5) to presenting evidence about social work intervention.

In summary, a great deal of content learned about research and statistics for understanding social work problems, diverse populations, and social phenomena (SWK 3401) forms the base for understanding social work intervention (SWK 3402). We begin our new voyage by examining what is meant by “social work intervention.”

Understanding Social Work Intervention

When we act to facilitate the process of change, we are engaging in **intervention**.

“In social work, interventions are intentionally implemented change strategies which aim to impede or eradicate harm, or introduce betterment beyond harm eradication, thus social work intervention encompasses a range of psychotherapies, treatments, and programs. Interventions may be single or complex” (Sundell & Olsson, 2017, p. 1).

The important aspects of this quote are that social work interventions:

- are intentional
- are implemented to create change
- aim to prevent or eliminate harm and/or to promote positive outcomes
- include a range of strategies at the individual, dyad, group, community, policy, and global level
- include strategies with different degrees of simplicity and complexity.

A major defining characteristic of social work intervention is our profession's practices of intervening at all levels of functioning and social systems—other disciplines or professions generally focus on one or two levels. Here are examples of the multiple ways that social workers intervene at different levels.

Human Biology. Social workers use evidence related to human biology to further their understanding of human behavior, to inform their intervention strategies, and evaluate their intervention efforts. For example, social workers in the maternal and child health arena have a long history of working to improve child development outcomes by intervening to ensure that mothers, infants and young children have access to the basic nutritional and environmental resources necessary for children to grow up healthy and to reach their greatest developmental potential.

Individual. Social workers utilize evidence concerning the “whole” person—biology, psychology, and social world—to assess, intervene, and evaluate interventions with individuals at all ages and stages of life (this combines the biopsychosocial and lifespan perspectives). One branch of “micro” social work practice, called psychiatric social work, addresses the intrapsychic needs of individuals who experience or are at risk of emotional, behavioral, or mental disorders. This is often accomplished through individual

counseling and other psychiatric services but may include a host of preventive and environmental interventions, as well.

Interpersonal. Social work professionals are acutely aware of the significance that the social environment plays in human experience, and that individuals function within a system of close interpersonal relationships. Evidence concerning interpersonal relationships is used by social workers both to understand individuals' experiences and how individuals influence others, as well. This evidence informs their intervention strategies, as well as their evaluation of the interventions they implement. A great deal of social work intervention is aimed at improving functioning in human relationships, including (but not limited to) couples, parents and children, siblings, extended family systems, and other important social groups. For example, one of the Grand Challenges for Social Work (see aaswsw.org) is dedicated to eradicating social isolation and another to stopping family violence—two significant aspects of interpersonal relations.

Communities. Facilitating change within neighborhoods and communities is a long-standing tradition in social work. Community development, community organization, community empowerment, and capacity building are “macro” approaches in which social workers are engaged to improve community members' well-being. Communities are often self-defined, rather than defined by geography, policy, or other externally imposed authority. Social change efforts that impact larger segments of society often begin with change at the community level. Social workers apply evidence concerning social indicators and other community behavior evidence in planning, implementing, and evaluating intervention at this level.

Organizations and Institutions. Social workers often intervene with organizations, agencies, programs, service delivery systems, and social institutions to effect change both on behalf of the health of the organization and on behalf of the clients these organizations and institutions serve. This type of intervention might be in the form of change agent with the organization or advocacy on behalf of

clients affected by the organization. Organizations have both formal and informal systems in operation, and both types have powerful influences on the behavior of individuals in the system. Thus, social workers engage with evidence about individuals, interpersonal relations, and organizational behavior to inform and evaluate their organization and institution level interventions.

Policy. Policy is a form of intervention, and social workers are often engaged in activities that help shape policy decisions or how policy is implemented. Social work interventions include policy established by organizations and social institutions, as well as public policy at the local, regional, state, national, and global/international level. Evidence at all levels is used to engage in policy practice, in terms of shaping, implementing, and evaluating policy that influences the way social workers practice and clients' lives.

While the breadth of practice domains contributes to the power of social work as a profession, it also contributes to certain challenges. Consider the wide array of practice contexts where social workers intervene:

- child welfare
- corrections/criminal justice systems
- schools and education
- gerontology
- health care systems
- hospice
- intimate partner/domestic/relationship violence
- mental health
- military/veterans affairs
- developmental disabilities
- public health
- substance misuse/addiction
- housing
- food security
- income maintenance/assistance

- workplace/employee assistance
- disaster relief
- community development
- government/policy
- and more...

Now consider the challenge that this diversity creates in terms of the profession's developing an intervention knowledge base. Some knowledge is relevant across this wide spectrum of levels and settings for social work practice, but a great deal of the knowledge on which social workers depend is specific to a particular social problem, social system, or population. For example, knowledge needed to protect young children from caregiver maltreatment may both overlap and differ from knowledge needed to protect vulnerable older adults and adults with intellectual disabilities from neglect or abuse by their care providers.

Social work intervention is rooted in the principles, values, and ethics of our profession. This includes adopting at least six central perspectives:

- a biopsychosocial perspective
- a lifespan developmental perspective
- a person-in-environment perspective
- a strengths perspective
- an emphasis on client self-determination
- an emphasis on evidence-informed practice.

Where Empirical Evidence Fits into Social Work Practice

In our earlier course, you learned to recognize empirical evidence as one form of knowledge used by social workers in understanding social work problems, diverse populations, and social phenomena.

Empirical evidence is equally important to informing and understanding social work interventions.

Briar (1974) identified five expectations regarding the fit between empirical evidence and social work practice. These expectations include:

1. "Social work practice is not random but guided by a body of knowledge and skills that are discernible and transmittable and that facilitate attainment of its socially worthwhile and publicly sanctioned goals.
2. Social work is responsible for maintaining the institutions necessary to update its knowledge base, to train its practitioners, and to oversee the proper and ethical discharge of its services.
3. Practicing as part of a publicly sanctioned profession, social workers must be accountable—to clients, to peers, and to a variety of sanctioning bodies. Accountability is most rudimentarily expressed by a commitment to and a responsibility for demonstrating that practice is effective and efficient. Demonstrating the effectiveness of practice requires evidence that service goals have been achieved and that goal attainment was causally linked to the activities (programs, methods) undertaken to reach the goals. Demonstrating efficient practice also requires evidence that service goals are attained with the least possible cost—in time, money, effort, and client suffering. Thus, in order for social work practice to be accountable it must be subject to scrutiny according to acceptable evidentiary standards.
4. A fourth basic premise is that professional knowledge in general, and knowledge that guides intervention in particular, must be tested and supported by empirical evidence obtained and evaluated according to prevailing scientific standards. Such standards must include, at the minimum, an explicit and systematic procedure for defining, gathering, and analyzing relevant evidence.
5. Finally, human behavior, individually and in the aggregate, as

well as the process of behavior change, is complex and multi-determined. Hence, interventions cannot be viewed as or expected to be uniformly applicable or universally effective. Their effectiveness is likely to vary in relation to the outcomes that are pursued, to the problem and other client characteristics, and to factors of the helping and in vivo situation” (pp. 2-3).

What This Means for Social Work. In summary of Briar’s points, the global questions we always ask are:

- what works,
- for whom does it work, and
- under what conditions does it work?

There exist two sides to the evidence “coin” regarding the social work profession. First, social workers seek and utilize evidence to inform our practice decisions. Second, we seek and utilize evidence to evaluate the impact of our intervention efforts and modifying our strategies when indicated by the evidence. Figure 1-1 depicts the relationship between these two facets of social work intervention and evidence.

Figure 1-1. Relationships between social work evidence and intervention



Why this matters so much to our profession was noted in our module introduction: the statement, “good intentions don’t always lead to good results” (Miron, 2015). An example suggested by Briar (1974) was a study of an intensive services intervention provided with the best of intentions to frail elderly clients, but in the end, was associated with a higher mortality rate than the treatment-as-usual condition: 25% compared to 18% (Blenkner, Bloom, & Nielsen, 1971). It was not the services provided by social workers that caused the higher rate of death; the group receiving the intensive services intervention was more likely to be moved to a nursing home setting (34% compared to 20%) than the other group, and this change in placement was likely responsible for the increased mortality. An unintended harmful outcome brought forth by the person trying to help or heal is called an **iatrogenic effect** of intervention. And, at the very least, we need to be sure that in intervening we are doing no harm.



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=25>

Module 1 Chapter 2: What is Evidence-Based Practice?

The left-hand side of the evidence-intervention-evidence figure (Figure 1-1) is the focus of this chapter: informing practice decisions with evidence.



This effort could be referred to as addressing “avoidable ignorance” (Gambrill and Gibbs, 2017, p. 73). Selecting an intervention strategy should be guided by evidence of “its potential to achieve the desired goals for clients” (Briar, 1974, p. 2). In this chapter you will learn:

- Definition and elements of evidence-based practice
- How evidence-based practice compares to other types of practice informed by evidence

Defining Evidence-Informed Practice, Evidence-Based Practices, & Evidence-Based

Practice in Social Work

As you move through your social work education, you are likely to encounter the term “evidence-based practice” and the expectation that social workers engage in practice based on evidence. On the surface, this seems like a simple concept: social workers engage in practice that is informed by or based on evidence and that evidence evaluating practice is used, as well. The concept of evidence-based practice, however, is considerably more complex and nuanced than it seems on the surface. It is helpful to distinguish between evidence-informed practice, evidence-based practices (note this is pluralized with the letter “s”), and evidence-based practice (not pluralized).

Evidence-Informed Practice. In our earlier course you learned about the ways that social work interventions might be developed based on empirical evidence. Informative evidence included epidemiology, etiology, and prior studies of intervention results: **efficacy studies** and **effectiveness studies**. Together, these sources of information, when applied to practice, comprise what is meant by the term evidence-informed practice. The term evidence-informed practices (EIPs) is sometimes used, as well.

Evidence-Based Practices. The concept of **evidence-based practices (EBPs)** is pretty much what it sounds like: practices based on empirical evidence. When professional practices are based on evidence, practitioners utilize empirical evidence concerning the outcomes of specific intervention approaches as the basis for designing their own interventions. The key to EBPs is that the practices utilized have an evidence base supporting their use, and that they are being implemented with a good deal of fidelity to the interventions as originally studied. The more there is “drift” from the studied intervention, the less relevant the supporting evidence

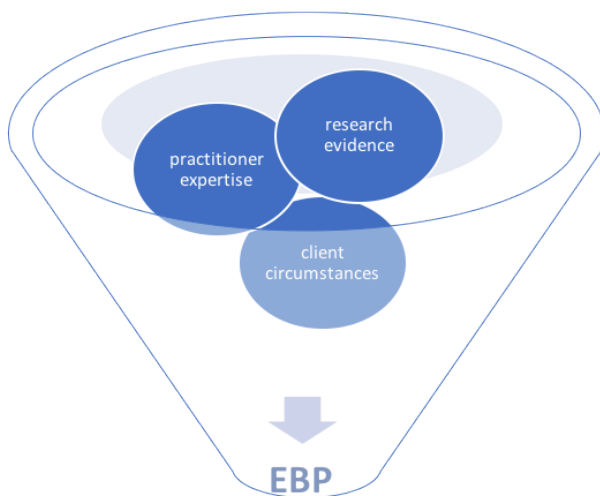
becomes, and the more the practices look to be evidence-informed (EIPs) rather than evidence-based practices (EBPs).

Consider the example of the Duluth Model for a coordinated community response to domestic violence that emerged in the intervention literature during the 1970s and 1980s (see Shepard & Pence, 1999). Duluth Model advocates provided a great deal of detail concerning the components of their coordinated community response model, and evidence of its effectiveness in addressing the problem of intimate partner (domestic) violence. Subsequently, many communities adopted the model as representing EBPs, describing themselves as being “Duluth Model” programs. In many cases they integrated only the batterer treatment program approach; many programs incorporated their own tools and local flair into the batterer treatment programs, as well. Unfortunately, the Duluth Model approach to intervening around intimate partner violence included batterer treatment programs alongside a coordinated community response model. What many programs left out was the intensive community organization and empowerment work that was also part of the model—leading to significant changes in local policy, as well as policing and court practices. The evidence supporting implementation of the Duluth Model was not relevant to the truncated intervention approach (batterer treatment only), nor was it relevant when the treatment program was modified with different tools. Communities disappointed that their own outcomes did not match those reported in Duluth were often communities where there had been considerable loss of fidelity to the full model—their practices might or might not have been evidence-informed (EIPs), but they were not the EBPs known as the Duluth Model. Evidence from the original Duluth Model programs lost its relevance the more the programs changed the shape of their practices—pieces of the puzzle no longer fit together.



Evidence-Based Practice. *Evidence-based practice (EBP)*, without the “s,” is where further complexity is introduced. EBP is both an ideology and a social work methodology (Gibbs, 2003). As an *ideology*, EBP involves a commitment to applying the very best available, relevant evidence to problems encountered in practice. “It requires changes in how we locate and integrate research into practice” (Gibbs & Gambrill, 2002, p. 453). This means engaging with evidence at all points of contact with clients, throughout the entire course of the helping relationship. As a *methodology*, EBP involves applying a very specific process to practice decision-making—assessing what is known and what is unknown about a practice problem (Gambrill & Gibbs, 2017). In the EBP process, empirical evidence is integrated with the practitioner’s professional experience and the client’s (or patient’s) circumstances, values, and preferences (Strauss et al., 2019), as depicted in Figure 2-1. Hence, more sources of knowledge inform practice decisions.

Figure 2-1. The EBP model



Engaging in EBP requires the practitioner to engage in a sequence of six steps which have been modified here from clinical medicine (Strauss et al., 2019) to social work practice at multiple intervention levels (see Gibbs, 2003). Engaging in the EBP process is appropriate when practice situations are uncertain or ambiguous, when routine practice decisions seem inappropriate or inadequate. EBP is about *how* to approach practice thinking and decision-making in uncertain situations; it is not about *what* to think or decide.

Step 1: Specifying an answerable practice question.

Step 2: Identifying the best evidence for answering that question.

Step 3: Critically appraising the evidence and its applicability to the question/problem.

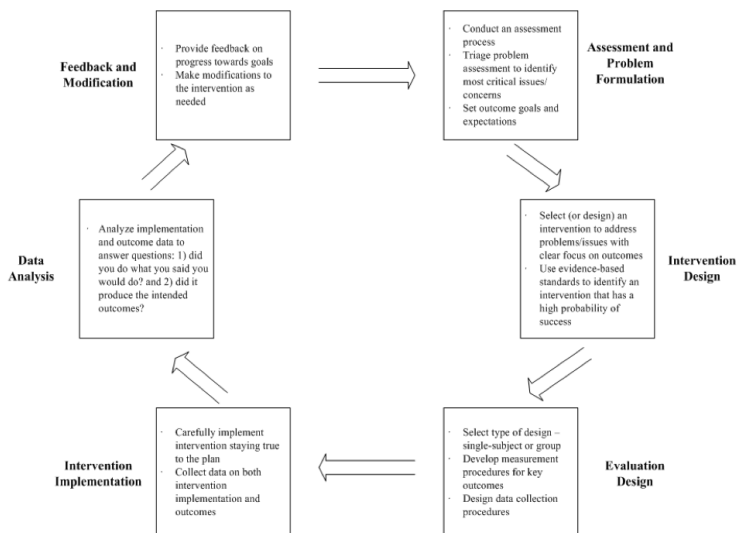
Step 4: Integrating results from the critical appraisal with practice expertise and the client's or client system's unique circumstances.

Step 5: Taking appropriate actions based on this critical appraisal of evidence.

Step 6: Monitoring and evaluating outcomes of (a) the practice decision/intervention and (b) effectiveness and efficiency of the EBP process (steps 1-5).

We examine each of these steps in greater detail in Module 2. For now, let's look at how these steps might fit into the social work problem-solving process overall—regardless of the “micro” to “macro” level of intervening in which a social work is engaged. The Figure 2-2 problem-solving process diagram is provided for us by Dr. Jerry Bean (unpublished), and adjunct instructor with the Ohio State University College of Social Work.

Figure 2-2. Social work problem-solving process and engaging with evidence.



As you can see, engaging with evidence occurs at multiple points in the problem-solving process—from initial assessment of the problem to be addressed, to developing solutions or interventions, to evaluating the impact of those interventions. Good social work interventions are informed by evidence, appropriate for the social work populations to whom they will be delivered (regarding

diversity characteristics, past history and experiences), feasible given the resources available (time, skills, space, and other resources), acceptable to the social work professional (in terms of values, beliefs, ethics), and acceptable to the recipient clients.

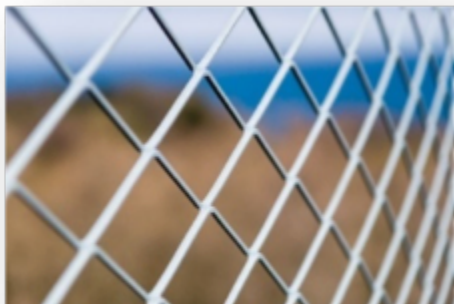
Critics of Evidence-Based Practice

Before considering criticism of EBP it is important to repeat that the process is not engaged with routine practice matters. EBP is engaged when a practice situation does not fit the typical or routine. For example, a substance abuse treatment program may routinely provide (evidence supported) cognitive behavioral therapy (CBT) to clients whose goal is to become free of substance misuse. There is no need to engage a search for evidence in this scenario, only to provide clients with their options and the evidence behind them. However, what happens when the social worker encounters a client who experiences significant cognitive limitations resulting from traumatic brain injury (TBI) following an accident, sports-related, or military service injury? The social worker in this scenario could engage in the EBP process to determine the intervention strategies that have the greatest likelihood of successful outcomes for this client experiencing co-occurring problems. However, you should be aware that there are critics who do not entirely embrace the EBP process as a professional standard practice (see Gibbs & Gambrill, 2002).

External constraints. One limitation experienced with strict adherence to the EBP method is concern over a lack of information as to how practitioners account for constraints and opportunities associated with policies or service delivery system structures (Haynes, Devereaux, & Guyatt, 2002). For the most part, the EBP method emphasizes the decision-making process that takes place between the practitioner and the client (or patient). However, practice decisions and decision-making processes are heavily

influenced by the context in which they take place. This remains an important practical concern that should be accounted for in both the sphere of practitioner experience and sphere of client circumstances/preferences.

For example, there exists considerable evidence to support sober housing as a desirable placement for individuals in recovery from a substance use disorder. However, there also exist constraints in many communities that include a lack of sufficient sober housing units, sober housing not being adapted for persons with co-occurring mental or behavioral health concerns, and restrictions on eligibility for persons with an incarceration history. Or, for example, a social worker may practice within an agency dedicated to a specific practice philosophy; implementing an innovative approach may not be supported by the agency, regardless of the evidence supporting its use. Some addiction recovery programs, for example, are founded on a philosophy that does not support the use of prescribed medications to support behavioral counseling or therapy (medication assisted therapy, or MAT). A social worker in such a program would be constrained to making a referral to another provider for clients who would benefit from or desire MAT. Thus, practice decisions are influenced by boundaries imposed from the contexts where they happen—the EBP process may not be adaptable to some of these constraints, barriers, or boundaries.



Access to Empirical Evidence. Another possible limitation of the EBP model is that it is heavily dependent on the degree to which empirical evidence might exist related to the practice questions at hand. This issue has several inter-connected parts.

It is difficult and time consuming to locate evidence. The search for relevant evidence can be time and labor intensive. However, degree of difficulty in locating evidence does not excuse a failure to search for evidence. Related arguments reported by Gibbs and Gambrill (2002) include concerns expressed about implementing EBP when practitioners no longer have access to academic libraries at the institutions where they were educated. The counter-argument to this concern: technology has opened access to interacting with a great deal of information, globally.



Being able to search effectively is not the problem it once was; being able to search efficiently is the greater problem—there is often too much information that needs to be sorted and winnowed.

A related argument concerns high caseloads (Gibbs & Gambrill, 2002) and that a practitioner may not be paid for the time spent in this aspect of practice since it is not “face time” spent with clients. Social workers should be advocating for these activities to be reimbursable as client-services. Regardless, our professional code of ethics requires social workers “...to search for practice-related research findings and to share what is found with clients (including nothing)” (Gibbs & Gambrill, 2002, p. 463). It is a professional activity engaged on behalf of, and perhaps with, clients.

Which relates to another of the concerns addressed by Gibbs and Gambrill (2002): EBP as a hinderance to therapeutic alliance or rapport in the practitioner-client relationship. As a counter-argument, this should not be the case if the practitioner engages appropriately in the process. What could be more important to clients than learning about the best alternatives for addressing their concerns? As noted above, the search for evidence can be engage

with clients, not only for them. Engaging with evidence is also a professional development activity that may relate to maintaining professional licensure, at least in some states. And, the more you engage in these activities, the more efficient you are likely to become in the process of locating and analyzing this kind of information. It is no longer considered best practice to simply rely on practice traditions (Gibbs & Gambrill, 2002), nor is the search limited to what is available in brick-and-mortar libraries.



© A. Begun, 2015

*Evidence is based on **aggregated data**, not individuals.* This observation means that empirical evidence is typically presented about a group of individuals or cases, not about individuals—and the client or client system with whom the social worker practices is an individual. So, this observation about aggregated data is mostly true, but not necessarily a liability if the information is properly evaluated.

As you may recall from our prior course, social and behavioral

science develops theory and evidence based on observations of “samples” made up of individuals who (hopefully) represent the population of interest. The knowledge developed is generalizable from the sample to the population. The disadvantage in aggregating information about a sample of individuals is that we lose specificity about what is going on for any one individual. This loss of individual specificity increases as a function of increasing individual variation (diversity or heterogeneity). So, the criticism is valid up to a point. Evidence from aggregated data gives us some initial good guesses about what to expect, but is not predictive of every individual in the population. Our initial good guesses based on aggregated data improve the more representative the sample is of the population—that the studies included representative diversity. For example, if evidence about the effectiveness of a combined medication/behavioral intervention for substance use disorders is based on a sample of men treated through their Employee Assistance Programs (EAPs), that evidence tells us relatively little about what to expect from the intervention with women or with individuals incarcerated for substance-related offenses—not only might these individuals differ in life circumstances, they may differ in other important ways, such as race/ethnicity, age, and severity of the problem being treated.

The EBP process needs to include attention to representativeness of the samples in the reported studies and how well those studies represent the circumstances of the individual clients with whom we are working. And, the EBP process needs to include an on-going monitoring mechanism so that we can be attuned to points where our individual client experiences diverge from the expected.



Evidence is based on controlled experimental conditions, not real-world conditions. This criticism is partially accurate. Efficacy studies involve carefully controlled experimental conditions that reduce variance as means of enhancing internal validity—the very variability that we see in real-world practice conditions (risking external validity). The rationale is that these approaches ensure that conclusions about the study results accurately reflect the impact of the intervention itself, and not the influence of other explanatory variables. For example, up until the 2000's, a great deal of breast-cancer research was focused on the population of post-menopausal women. Not only was this the largest group of persons diagnosed with breast cancer, introducing younger, pre-menopausal women and men with breast cancer meant that intervention study results were confounded by these other factors—including them in the studies would make it difficult to determine what was working. However, this also meant that there was little evidence available

to inform practice with pre-menopausal women and men who contracted breast cancer. Furthermore, many intervention studies early on were conducted in centers where practitioner-investigators were breast cancer specialists, well-prepared to provide treatment with a high degree of fidelity to the treatment protocols being tested.

For this reason, the next step in the knowledge building process about intervention involves effectiveness studies—testing those efficacy conclusions under more diverse, real-world conditions. As previously noted, this work includes more diverse populations of clients and more diverse practitioners, working under less artificially controlled conditions. Consider, for example, the evolution of Motivational Interviewing (MI), originally developed for addressing alcohol use disorders and now applied across many different physical and behavioral/psychological health conditions (Rubak, Sandbæk, Lauritzen, & Christensen, 2005). When MI was first studied, the supporting evidence was based on intervention provided by practitioners specifically and highly trained in the approach. As evidence for its efficacy and effectiveness expanded, a wider range of practitioners began applying the approach. The originators of MI created a certification process for training practitioners (Motivational Interviewing Network of Trainers, or MINT, certification) as a means of reducing variability in its application (enhancing fidelity), but there is no requirement that trainers have this certification (see <http://www.motivationalinterviewing.org/> for more information). The rarified practice conditions of the initial intervention studies can be equated with a sort of virtual world, somewhat divorced from real-world practice conditions.



The evidence is not necessarily about social work interventions. There should be no disciplinary boundaries placed on the search for evidence related to a particular practice problem—perhaps the relevant practice questions have been tackled by psychology, medicine, nursing, public health, criminal justice, education, occupational therapy, or another profession. We can tap into that potentially rich, diverse knowledge base to inform social work intervention. Many arenas in which social workers practice are interdisciplinary fields, areas such as:

- substance misuse
- gerontology
- developmental disabilities
- corrections
- health care
- education
- mental health

As such, social workers intervene as members of teams where the interventions are not exclusively “social work” interventions.



Chapter Summary

In uncertain or ambiguous circumstances, social workers might have the good fortune to be able to engage in the evidence-based practice (EBP) process, applying a strong evidence base alongside practice wisdom and client preferences to making practice decisions about interventions. However, at other times we might need to consider engaging in evidence-informed practices (EIPs) or evidence-based practices (EBPs). The point is to make a strong effort to identify and assess the available evidence, thereby helping

clients become informed participants in making decisions about intervention plans. Another point is that social work professionals need to contribute to developing evidence, especially where we encounter significant information gaps.



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=26>

Module 1 Chapter 3: Practice Evaluation as Evidence

The right-hand side of the evidence-intervention-evidence figure from Chapter 1 (Figure 1-1) is the focus of this chapter.



In Chapter 2 we looked at evidence-informed practice decisions. In this chapter, we introduce information about evaluating practice, what other disciplines call data-based or data-driven decision making: using data and evaluation research methods to make social work practice accountable and to inform practice improvement efforts.

In this chapter you will learn:

- basic principles related to four evaluation formats (needs assessment, outcome, process, cost-effectiveness)
- distinctions between practice, program, and policy evaluation
- how evaluation and intervention research compare.

Why Evaluate?

In short, social work professionals engage in evaluation of practice as an accountability issue. We are accountable to clients, programs, funders, policy decision-makers, and the profession to ensure that we are delivering the best possible services, that the services we deliver achieve the promised benefits, and that the resources dedicated to our services are well-spent. This has previously been covered in our discussions regarding standards presented in the Social Work Code of Ethics. Of particular relevance to this discussion is the Standard 5.02 concerning evaluation and research (p. 27). Social workers are expected to evaluate policies, programs, and practice interventions, as well as facilitate research that contributes to the development of knowledge.

What is Evaluation?

Throughout the remainder of our course *Research and Statistics for Understanding Social Work Intervention* we examine methods for evaluating intervention efforts. A framework for understanding different approaches to evaluation is helpful, beginning with the nature of the evaluation research questions and exploring how these relate to different forms or approaches to evaluation.

Evaluation Questions. By now you recognize that research designs and methodologies are driven by the nature of the research questions being asked. This is equally true in the evaluation research arena. Here is a sample of the kinds of questions asked in evaluating social work practice at different levels:

- Did client behavior change to a significant degree and in the desired direction?
- Were gains associated with intervention sustained over time?
- Are there unintended negative consequences associated with

the intervention?

- To what extent are principles of diversity awareness integrated into practitioner behaviors and practitioner supervision?
- How satisfied are clients with various aspects of the delivered intervention?
- Is the intervention's cost/benefit ratio favorable compared to other intervention options?
- Are some people deriving more benefit than others from the intervention?
- Is there a more cost-efficient way to achieve similar gains from the intervention?

Evaluation Formats. Because evaluation questions differ, social workers employ varied formats for engaging in evaluation. Here is a description of four major forms of evaluation research: needs assessment, outcome evaluation, process evaluation, and cost-effectiveness evaluation.

Needs assessment. The aim of **needs assessment** is to answer questions related to the scope of a problem or need and where gaps exist in efforts to address the problem or need. For example, school social workers may want to know about the problem of bullying that occurs in a school district. They might engage in a needs assessment to determine the nature and extent of the problem, what is needed to eradicate the problem, and how the problem is being addressed across the district. Where they detect sizeable gaps between need and services provided, social workers can develop targeted responses. The needs assessment might also indicate that different responses need to be launched in different circumstances, such as: elementary, middle, and high school levels; or, parents, teachers, administrators, peers, and mental health professionals in the district; or, different neighborhood schools across the district. Needs assessment is often concerned with the discrepancy between what is needed and what is accessed in services, not only what is offered. As proponent of social justice, social workers are also concerned with identifying and addressing

disparities (differential gaps) based on income, race/ethnicity, gender/gender identity, sexual orientation, age, national origin, symptom severity, geographical location (e.g., urban, suburban, rural disparities), and other aspects of human diversity. This represents an important extension of what you learned in our earlier course, *Research and Statistics for Understanding Social Work Problems and Diverse Populations*. The gap between two sides or groups is sometimes monumental.



©Audrey Begun

Outcome evaluation. Evaluating practice outcomes happens at multiple levels: individual cases, programs, and policy. Social work professionals work with clients or client systems to achieve specific change goals and objectives. For example, this might be reducing a person's alcohol consumption or tobacco use, a couple having fewer arguments, improving student attendance throughout a school, reducing violence in a community, or breaking a gender or race based "glass ceiling" in an institution. Regardless of the level of intervention, social work professionals evaluate the impact of their practices and intervention efforts. This type of research activity is called **outcome evaluation**. When outcome evaluation is directed to understanding the impact of practices on specific clients or client systems, it is called **practice evaluation**.

Evaluating the outcomes of interventions also happens at the

aggregate level of programs. Social workers engaged in **program evaluation** look at the impact of an intervention program on the group of clients or client systems it serves. Rather than providing feedback about an individual client or client system, the feedback concerns multiple clients engaged in the intervention program. For example, social workers might wish to evaluate the extent to which child health goals (outcomes) were achieved with an intervention program for empowering parents to eliminate their young children's exposure to third-hand smoke. The background for this work is described in an article explaining that third hand smoke is the residue remaining on skin, clothing, hair, upholstery, carpeting, and other surfaces; it differs from first- or second-hand smoke exposure because the individuals are not exposed by smoking themselves or breathing the smoke someone else produces. Young children come into close contact with contaminated surfaces when being held by caregivers, riding in vehicles, or crawling and toddling around the home where smoking has occurred, leaving residue behind (Begun, Barnhart, Gregoire, & Shepperd, 2014). Outcome oriented program evaluation would be directed toward assessing the impact of an intervention delivered to a group of parents with young children at risk of exposure to third-hand smoke at home, in transportation, from relatives, or in child care settings.



Policy evaluation has a lot in common with program evaluation, because policy is a form of intervention. Policy evaluation data are based on intervention effects experienced by many individuals,

neighborhoods, communities, or programs/institutions taken together, not tracking what happens with one client system or a single program at a time. For example, communities may gather a great deal of evaluation data about the impact on drug overdose deaths related to policies supporting first-responders, family members, friends, and bystanders being able to deliver opioid overdose reversal medications (naloxone) when first encountering someone suspected of experiencing opioid overdose. “As an antidote to opioid overdoses, naloxone has proven to be a valuable tool in combating overdose deaths and associated morbidity” (Kerensky & Walley, 2017, p. 6). Policy evaluation can answer the question of how much impact such a policy change can make. Policy evaluation also answers questions such as: who should be provided with naloxone rescue kits; how naloxone rescue kit prescribing education might alter opioid prescribing behavior; whether different naloxone formulations, doses, and delivery methods provide similar results and how do their costs compare; how what happens after overdose rescue might keep people safe and link them to services to prevent future overdose events; and, how local, state, and federal laws affect this policy’s implementation (see Kerensky & Walley, 2017). These factors help determine if the impact of a policy is simply a drop in the bucket or a flood of change.



Process evaluation. **Process evaluation** is less concerned with questions about outcomes than with questions about how an intervention or program is implemented. Why evaluating process matters is clear if you think about fidelity examples previously discussed (e.g., the Duluth model for community response to domestic violence). Process evaluation matters in determining what practitioners really do when intervening and what clients or client systems experience during an intervention. It also matters in terms of understanding the “means to the end,” beyond simply observing the end results. Process evaluation also examines the way an intervention or program is supported by agency administrators, agency activities, and distribution of resources—the context of the intervention—and possible efficiencies or inefficiencies in how an intervention is delivered.

“Process evaluations involve monitoring and measuring variables such as communication flow, decision-making protocols, staff workload, client record keeping, program supports, staff training, and worker-client activities. Indeed, the entire sequence of activities that a program undertakes to achieve benefits for program clients or consumers is open to the scrutiny of process evaluations” (Grinell & Unrau, 2014, p. 662).

For example, despite child welfare caseworkers’ recognition of the critically important role in child development for early identification of young children’s mental health problems and needs, they also encounter difficulties that present significant barriers to effectively doing so (Hoffman et al., 2016). Through process evaluation, the investigators identified barriers that included differences in how workers and parents perceived the children’s behavioral problems, a lack of age-appropriate mental health services being available, inconsistencies with their caseworker roles and training/preparation to assess and address these problems, and a lack of standardized tools and procedures.

at the level of a felony and for individuals with severe psychiatric disorders. While cost savings were realized by taxpayers, complicating the picture was the fact that the budget where gains were situated (criminal justice) is separate from the budget where the costs were incurred (mental health system).



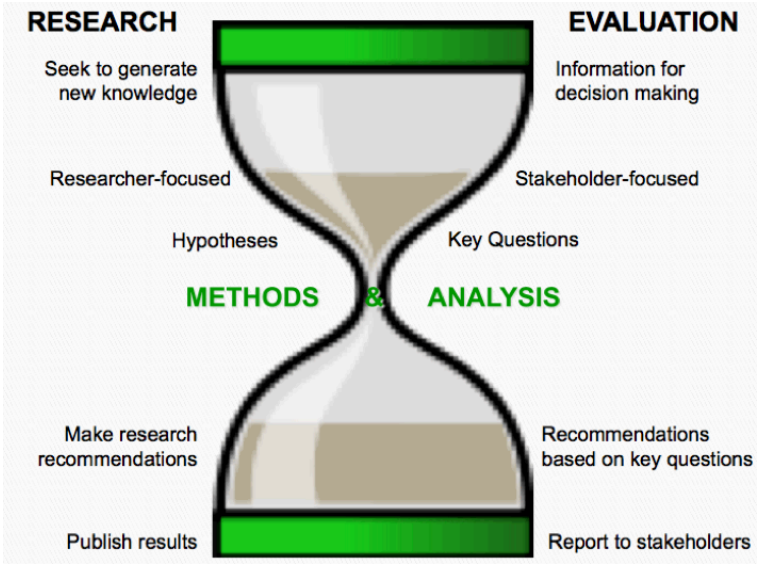
How Evaluation and Intervention Research Compare

The goals, objectives, and methods of **evaluation research** and **intervention research** often appear to be very similar. In both cases, systematic research procedures are applied to answer questions about an intervention. However, there exist important differences between evaluation and research to consider, important because they have implications for how investigators and evaluators approach the pursuit of evidence.

Differences begin with the nature of the research questions being asked. Evaluation researchers pursue specific knowledge,

intervention researchers pursue generalizable knowledge. In evaluation, the goal is to inform leader or administrator decisions about a program, or to inform an individual practitioner’s intervention decisions about work with specific clients. The aim of practice or program evaluation is to determine the worth of an intervention to their agency, their clients, and their stakeholders. Intervention researchers, on the other hand, have as their goal the production of knowledge or the advancing of theory for programs and practitioners more generally—not a specific program or practitioner. This difference translates into differences in how the research process is approached in evaluation compared to intervention science. Figure 3-1 depicts the differences in approach, methodology, analysis, and reporting between evaluation and intervention research (LaVelle, 2010).

Figure 3-1. Differences between intervention and evaluation research.





Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=30>

Chapter Summary

In this chapter, you were introduced to why evaluation is important in social work, extending what you learned in the prior course about the relationship of empirical evidence to social work practice. You also learned about the nature of evaluation questions and how these relate to evaluation research—an extension of what you learned in the prior course concerning the relationship between research questions and research approaches. In this chapter you were introduced to four different formats for evaluation (needs assessment, outcome evaluation, process evaluation, and cost-effectiveness evaluation), and you learned to distinguish between evaluation and intervention research.

Module 1 Chapter 4: Ethics Related to Intervention Research

In our prior course, you learned a great deal about research integrity and ethical issues that might arise when conducting research. All those integrity and ethical topics apply to research related to understanding social work intervention. Several additional ethical concerns are also relevant to our new course topic. In this chapter you will learn:

- distinctions between social work intervention research and practice
- four ethics topics specific to intervention research.

Is It Research or Practice?

The Belmont Report (1978) that you read about in our earlier course warrants revisiting here. The report distinguished between research and practice in this way: practice involves interventions:

“designed solely to enhance the well-being of an individual patient or client and that have a reasonable expectation of success” (p. 3).

Research was defined in the Belmont Report as:

“...an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge” (p. 3).

These two types of activity are clearly contrasted—their goals are markedly different. Practice goals relate to benefits for the individuals being served; research goals are served by the individuals who participate.

The Belmont Report authors acknowledged that the boundaries between research and practice are sometimes blurred, especially when innovative, radically new, untested practices are employed. Applying an “experimental” intervention does not automatically mean that it is research; however, the report’s authors advised that such interventions should “be made the object of formal research at an early stage in order to determine whether they are safe and effective” (p. 4).

Critically important is that the client understand the distinction between what activities are part of a research study or experiment and what activities are part of receiving services. Consent for treatment or other social work services is different from consent to participate in research, and any confusion needs to be clearly untangled for clients who might be involved with both. For example, they need to understand if they are completing an assessment tool because it will inform their care plan, or if they are completing it as data for a research study.

investigating an innovative intervention, investigators do not know all the potential risks—the intervention is still an unknown. They can make informed guesses, especially by consult the existing literature about similar interventions. But until they deliver the innovative intervention, no one knows what might happen under all circumstances. This explains why many interventions are first studied on a very small scale, putting as few people as possible at risk. This type of initial intervention research is called an **efficacy study**. It is conducted under carefully controlled experimental conditions, limiting the range of variability among study participants, and limiting the extent of risk exposure to relatively few participants. The goal in efficacy studies is to improve the assessment of potential risks and benefits associated with the intervention; it is not yet a goal to determine how effective the intervention might be under real-world conditions. That would be an **effectiveness study**, involving larger numbers of participants, a more diverse group of individuals receiving the intervention, and delivery of the intervention under more real-world condition. The risk/benefit ratio knowledge gained from efficacy studies can inform future studies about the innovation.



Consent to Participate. The Belmont Report and the NASW Code of Ethics emphasis on autonomy and self-determination apply to all research involving human participants. Guidelines and templates for informed consent procedures under many different research scenarios are available. Intervention research, however, sometimes involves special ethical concerns related to consent. Four that warrant attention are: clients' (or patients') dependence on the investigators to receive needed services, establishing participants' capacity to consent, who consents in macro-level intervention research, and incentive payments,

Dependence for services. Ethical conduct of research requires that a person's involvement is voluntary—this is part of self-determination and autonomy. A person's decision should be free from coercion. However, investigators may not be aware of subtle forms of coercion that might influence a potential study participant's decision. Individuals who depend on an agency, program, or practitioner for essential services may believe that consenting to participate in an offered research study is required to continue receiving good service or that it will create some form of preferential treatment. They may not feel completely free to say “No” to a request for participants, despite investigator assurances of the study's voluntary nature. This is no different from students or employees who are fearful of negative consequences from refusing to participate in an instructor's or employer's study, or who consent because they believe there will be advantages in their grades or job benefits, separate from the study benefits.



These examples represent forms of coercion to which intervention research investigators must attend. A best practice is to make sure that the person engaging in soliciting study participation and the informed consent process with potential study participants is not also involved with any aspect of providing them services. This also helps minimize the potential for or appearance of a **conflict of interest** where practitioners might have a vested interest in the outcomes of a study—to the point their behavior either intentionally or unintentionally skews study results.

Capacity to consent. Social workers often serve individuals whose capacity for engaging in informed, considered choices is limited. For example, young children cannot provide informed consent; we must rely on a combination of parental consent and the child assenting to participate. Other examples include individuals who experience cognitive or intellectual impairment resulting from dementia, brain injury, disease, developmental disabilities, chronic substance misuse, taking certain medications, or other causes. Furthermore, it is important for social workers to understand that decision-making is powerfully affected by crisis situations—people often make decisions during a crisis or when experiencing trauma that they might not make under their normal, daily-living conditions. These

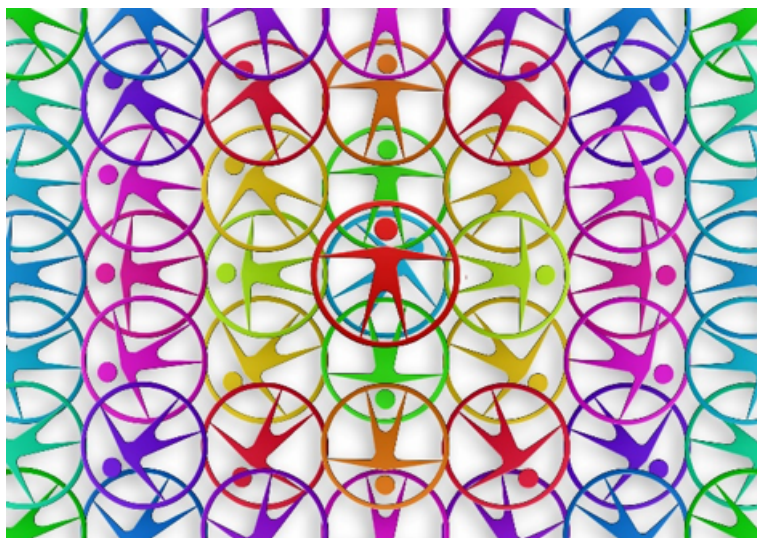
are examples of potentially vulnerable populations who may require additional considerations be undertaken to ensure their protection as participants in intervention research studies.



Who consents to macro-level intervention research. While it is usually clear who should provide consent for an individual to participate in research, who should consent on behalf of “macro” level groups is not necessarily evident. Consider, for example, an innovative program or policy might have significant impact on a community’s experience. State and local policies legislating the distribution of certain substances (tobacco, alcohol, and marijuana) offer good examples of this point. When policy decision makers in some states reduced the legal drinking age from 21 to 18 years during the 1970s, there was a significant impact on high schools—many high school seniors could now legally possess and consume alcohol as could any other adult. Enforcement of underage drinking policies among friends of those 18 years old quickly overwhelmed the high schools’ ability to address other discipline concerns (see Begun, 1980). If an experiment were conducted regarding implementation of such a policy, who should be involved in providing consent? This is true for more recent changes in state

or community policies regarding marijuana possession or distribution (medical or recreational) and restricting access to tobacco products for 18- to 21-year old emerging adults.

These policy “experiments” potentially have powerful effects on the experiences of individuals, families, communities, schools and other institutions—none of whom provided consent to have these “experiments” conducted in their midst. We rely on elected and appointed officials to make these decisions and evaluate their impact. These decisions are often politicized, and the evidence for their support or contradiction often lags far behind when policy changes are adopted.



Participant incentives. Study participants may be offered specific benefits as a means of motivating (incentivizing) them to participate in a research study. True of all research, but especially true of intervention studies where there may be risks associated with participation, investigators need to ensure that the offered options are not **coercive incentives**. In other words, that the person is not

accepting excessive risk simply because of the magnitude of the incentive. This may be difficult to assess, however, because what is motivating to one person may be coercive to another. For example, a \$15 meal voucher means something different to a person who is hungry and lacks resources for the next meal compared to a person who experiences a great deal of food security. A best practice is to conduct focus groups or survey measures, or consult the research literature, concerning what would be a motivating but not coercive amount and form of incentive.





Consider each of the following scenarios, considering what amount of money or other incentives would influence YOU to become a volunteer participant in each of the following types of intervention research, knowing that you could be in the “experimental” group (no amount might be your answer, but consider what level would become coercive for YOU).

1. An influenza immunization prevention study that had the potential to make you sick for up to two weeks with the flu if it does not work.
2. An alcohol consumption study designed to measure how an one-time informational drinking guideline intervention affects how much alcohol it takes for you to feel “buzzed” in a laboratory setting.
3. A group intervention study where members are coached to deliver positive, morale boosting statements to one another as a means of increasing amount of time spent studying and decrease procrastination on completing assignments.
4. A study of severely restricted calorie intake for one month on ability to perform complex cognitive tasks related to performing well in college.

Experimental and Control Groups. Two significant potential ethical issues surround experimental studies designed with a **control group**.

Non-treatment control groups. First, a study might recruit a group of individuals who experience the problem for which an innovative

intervention is being tested—overwhelming anxiety/panic attacks, for example. What if half of these individuals are provided with the experimental innovative intervention and the other half receive no intervention (a “non-treatment” control group)? While this study design meets scientific criteria for rigor, how do we as social work professionals feel about closing the opportunity for care to a group of individuals who came seeking services—are we okay with asking them to make an informed decision based on a 50:50 chance of receiving intervention? Despite this design’s scientific integrity it is important to consider the ethical concerns.

Instead, investigators could design the study to compare the innovative intervention with a **treatment-as-usual (TAU) condition**, rather than comparing it with a non-intervention control group. Or, instead of not providing any intervention to the control group, they could provide intervention later, once the study has demonstrated that it is safe and has positive effects. For this reason, individuals on a wait list for services are often engaged as a “control” group. However, it would be unethical to create a wait list simply as a means of meeting the demands of a study design.



At-risk participants. Imagine a prevention intervention study that recruits individuals at risk of developing a certain problem, but who do not yet experience that problem. Investigators do not know which of these individuals will go on to develop the problem—some most likely will not, despite having risk factors. How do we feel about the ethics of exposing all these individuals to the research burden and potential risks of the experimental intervention? Some individuals might be helped, but others were exposed to risk and burden without there being an advantage gained because they would not have developed the problem anyway.

For example, consider a study where investigators tested a school-based, group intervention for students at risk of developing an alcohol use disorder based on having been caught engaging in underage alcohol use. A portion of the students will progress to an alcohol use disorder without intervention, but *only* a portion will do

so, and no one is able to accurately predict which ones will do so. Are we comfortable with exposing all these students to the burden and risks—such as using their class or after-school time for the intervention and data collection sessions, and the stigma associated with being in the identified at-risk group? We lack a reliable crystal ball to predict which students have a favorable or unfavorable risk/benefit ratio.



What Happens at Study's End. Imagine the best possible scenario for a social work intervention study: the innovative intervention was found to be safe, more effective than previously existing options, and more cost-efficient, too. Now imagine what happens to individuals who were receiving the superior intervention when the study is complete. Agencies and programs may not be ready to adopt the innovation immediately—it will take time to generate the resources needed to implement the new intervention as a routine practice. What does this mean for individuals who were benefitting from the innovative intervention during the study? What

arrangements are the investigators ethically responsible for making to ensure that study participants continue to progress toward their goals, without interruption or the disruption of having to switch to a different intervention or interventionist?

This is not an issue if the studied intervention is naturally terminated as individuals reach their goals—for example, a smoking cessation intervention might naturally be withdrawn through client termination processes as each individual reaches a predetermined milestone (e.g., six months tobacco free). It is an issue when the intervention is about maintenance of intervention goals when a chronic condition is involved—for example, maintaining job and housing security among persons with schizophrenia. A transition plan needs to be developed and implemented, preferably before the study begins. This might be practiced in medication assisted treatment (MAT) studies testing the effectiveness of medication supported recovery from alcohol or other substance use disorders—along with behavioral interventions and counseling. Participants for whom the experimental medications were working need a pathway for access being maintained on those medications while the Federal Drug Administration (FDA) process progresses on the road toward approved use with the general population.



Chapter Summary

In this chapter, you learned about several ethical concerns that might arise in conjunction with intervention research—concerns over and above those that occur with all research involving human participants. First, we explored the important distinction between research and practice. Next, we examined concerns related to the risk/benefit ratio for intervention research and how knowledge from efficacy and effectiveness studies relates to risk/benefit decisions. Then we analyzed several intervention research consent issues, including potential participants' dependence on services and how this might influence consent decisions, potential participants' capacity to consent, who provides consent for macro-level intervention research, and the coercive potential of some incentive payment plans. The ethical concerns that arise with study designs involving non-intervention control groups and how investigators need to ensure continued care following an intervention study's conclusion were also addressed.

Module 1 Conclusions

In this module you first reviewed key principles from the earlier course, *Research and Statistics for Understanding Social Work Problems and Diverse Populations*, as they apply to the current course, *Research and Statistics for Understanding Social Work Intervention*. The readings then introduced concepts related to understanding social work intervention and the kinds of evidence used to inform and evaluate social work intervention. Information about the evidence-based practice (EBP) model was presented, along with several points of criticism, and EBP was contrasted with evidence-informed practice and evidence-based practices (EBPs). Concepts related to practice and program evaluation were introduced, and you read about different types of evaluation and the kinds of information they provide. Finally, you re-visited issues of research ethics, this time in the context of ethics specific to intervention research. You are now well-prepared to engage with the next module's content: identifying and using evidence to inform intervention.

Module 1 Key Terms and Definitions

aggregated data: data are combined from multiple sources (study participants or measures) in a summary form, rather than being presented on individual cases or measures.

beneficence principle: a principle directed toward ensuring that research participants are treated in an ethical manner, protected from harm, and what is in their best interests is promoted; one of three key principles identified in the Belmont Report.

coercive incentives: influencing a person to make a specific decision or behave in a certain manner that the person would not have ordinarily made or done, especially of concern if the decision or behavior might be detrimental to the person.

conflict of interest: situation where a person in an official role (e.g., research investigator, practitioner, supervisor) might derive personal benefit from their professional decisions or actions.

control group: in an experiment, the group used as a comparison or benchmark for groups receiving an intervention or condition of study interest.

cost-effectiveness evaluation: research analyzing the costs of an intervention relative to the resulting benefits or observed outcomes.

efficacy study: intervention research conducted under ideal, controlled conditions with select study participants (in contrast to effectiveness studies).

effectiveness study: intervention research conducted under real-world conditions with relatively diverse study participants (in contrast to efficacy studies).

evaluation research: a form of applied research aimed at determining the worth of engaging in a specific intervention (practice, program, or policy).

evidence-based practices (EBPs): interventions that have a body of evidence supporting their use or application in professional practice.

evidence-based practice (EBP): a specific process for practice decision-making that incorporates practice evidence with other sources of information.

evidence-informed practice: practices and interventions developed on the basis of information from research evidence.

iatrogenic effect: harm or illness resulting from applying an intervention.

intervention: taking action to improve an undesirable condition or situation.

intervention research: research studies designed to answer generalizable questions concerning the effects, safety, effectiveness, cost-effectiveness, and/or implementation of an intentional change strategy or approach.

needs assessment: research effort directed toward identifying unmet needs of a group/population or gaps in a service delivery system.

outcome evaluation: a systematic research approach applied to determine the effects (on specific outcomes) of intervening.

practice evaluation: a systematic research approach applied by practitioners to determine the effects of intervening with a particular client or client system.

process evaluation: a systematic research approach applied to determine how an intervention was actually delivered.

program evaluation: a systematic research approach applied to determine the effects of specific programs for clients/consumers/participants or other stakeholder groups.

policy evaluation: a systematic research approach applied to determine the impact or effects of an intervention at the level of policy.

treatment-as-usual (TAU) condition: an experimental control condition where the comparison group is offered the usual type/

level of intervention (as opposed to no intervention), to be compared with the intervention innovation of interest.

Module 1 References

- Begun, A. L. (1980). Social policy evaluation: An example from drinking age legislation. *Evaluation and Program Planning: An International Journal*, 3(3), 165-170.
- Begun, A.L., Barnhart, S.M., Gregoire, T.K., & Shepherd, E.G. (2014). If mothers had their say: Research-informed intervention design for empowering mothers to establish smoke-free homes. *Social Work in Health Care*, 53(5), 446-459.
- Belmont Report, The. (1978). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects research*. Prepared by The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (April 18). Washington, DC: Department of Health, Education, and Welfare. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>
- Briar, S. (1974). Effective social work intervention in direct practice: Implications for education. In M.D. Glickens, (Ed.), *Toward effective social work practice*, (pp. 7-20). NY: MSS Information Corporation.
- Blenkner, M., Bloom, M., & Nielsen, M. (1971). A research and demonstration project of protective services. *Social Casework*, 52, 483-499.
- Gambrill, E., & Gibbs, L. (2017). *Critical thinking for helping professionals: A skills-based workbook*, 4th edition. NY: Oxford University Press.
- Gibbs, L.E. (2003). *Evidence-based practice for the helping professions: A practical guide with integrated multimedia*. Pacific Grove, CA: Brooks/Cole and Thomson Learning.
- Gibbs, L. (2005). Applying research to making life-affecting judgments and decisions. Keynote address presented to the Seventeenth National Symposium on Doctoral Research in Social Work, The Ohio State University College of Social Work (April

- 16). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.5360&rep=rep1&type=pdf>
- Gibbs, L., & Gambrill, E. (2002). Evidence-based practice: Counterarguments to objections. *Research on Social Work Practice*, 12(3), 452-476.
- Haynes, R.B., Devereaux, P.J., & Guyatt, G.H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *Evidence-based Medicine*, 7, 36-38.
- Hoffman, J.A., Bunger, A.C., Robertson, H.A., Cao, Y., & West, K.Y. (2016). Child welfare caseworkers' perspectives on the challenges of addressing mental health problems in early childhood. *Children and Youth Services Review*, 65, 148-155.
- LaVelle, J. (2010). John LaVelle on describing evaluation. Blog posting to AEA365, A tip-a-day by and for evaluators (February 26). Retrieved from <http://aea365.org/blog/john-lavelle-on-describing-evaluation/>
- Miron, J.A. (2015). Good intentions don't always lead to good results. *The New York Times*, (May 12). Retrieved from <https://www.nytimes.com/roomfordebate/2015/05/12/should-prices-at-nail-salons-and-elsewhere-reflect-their-true-cost/good-intentions-dont-always-lead-to-good-results>
- Psychiatric Times. (2007). Mental health courts reduce incarceration, save money. (July) Retrieved from <http://www.psychiatrictimes.com/schizoaffective/mental-health-courts-reduce-incarceration-save-money>
- Rubak, S., Sandbæk, A., Lauritzen, T., & Christensen, B. (2005). Motivational interviewing: A systematic review and meta-analysis. *British Journal of General Practice*, 55(513), 305-312.
- Shepard, M.F., & Pence, E. (Eds.) (1999). *Coordinating community responses to domestic violence: Lessons from Duluth and beyond*. Thousand Oaks, CA: Sage Publications.
- Strauss, S.E., Glasziou, P. Richardson, W.S., & Haynes, R.. (2019). *Evidence-based medicine: How to practice and teach EBM, 5th edition*. Edinburgh: Elsevier.
- Sundell, K., & Olsson, T. (2017). *Social intervention research*. Oxford

Bibliographies, retrieved from
[http://oxfordbibliographiesonline.com/view/document/
obo-9780195389678/obo-9780195389678-0254.xml#](http://oxfordbibliographiesonline.com/view/document/obo-9780195389678/obo-9780195389678-0254.xml#)

MODULE 2

Module 2 Introduction

This module emphasizes the “how to” aspects of the social work profession’s movement to engage with evidence for informing practice: a movement “fueled by our desire to use the tools of science, rather than tradition or opinion, to identify policies and practices that are effective” (Gutierrez, 2011, p. xv). This module explores how social work professionals identify, access, assess, and utilize evidence in making practice decisions. The skill set of this module builds on what was learned in our previous course about reviewing empirical literature, focusing specifically on intervention evidence. We work within a “how to” framework of problem identification, assessment and measurement, and matching to the intervention as facets of the evidence-based practice (EBP) process introduced in Module 1 of our current course:

Step 1: specify an answerable practice question

Step 2: identify the best evidence for answering that question

Step 3: Critically appraise the evidence and its applicability to the question/problem

Step 4: Integrate the results from the critical appraisal with practice expertise and the client’s/clients’/client system’s unique circumstances

Step 5: Take appropriate actions based on this critical appraisal of evidence

Step 6: Monitor and evaluate outcomes of (a) the practice decision/intervention and (b) effectiveness and efficiency of the EBP process (steps 1-4).

Reading Objectives

This module begins with a review of what was learned in our first course Module 2 about research questions and empirical literature;

here you will learn to apply these concepts to understanding interventions. After engaging with these reading materials and learning resources, you should be able to describe social work practice activities associated with each of 6 steps in the EBP process. This includes being able to:

- Develop appropriate practice questions to guide the search for evidence;
- Explain the relationship of theory to intervention and logic models;
- Identify approaches and resources for acquiring evidence to answer intervention-related questions, including systematic, scoping, meta-analysis, and literature reviews;
- List criteria to consider in critically analyzing evidence;
- Specify an intervention plan based on evidence (including alternate plans);
- Identify implementation, monitoring, and evaluation as key steps in the EBP process;
- Define key terms related to social work research and research ethics.

Module 2 Chapter 1: Course Review and Overview

This brief chapter provides a context for our second course in the two-course sequence, *Research & Statistics for Understanding Social Work Intervention*. We begin with a brief review of key concepts learned in our first course (*Research & Statistics for Understanding Social Work Problems and Diverse Populations*), concepts with direct application to our new course learning objectives.

In this chapter you:

- briefly revisit key concepts relevant to the new course, *Research & Statistics for Understanding Social Work Intervention*.

Review of Key Concepts

A great deal of content was covered in the first course, much of which is directly relevant to understanding social work interventions. Here are several key topics revisited in the new course.

- The translational science framework included basic research, intervention research, and implementation science. Our earlier course emphasized basic research (epidemiology, etiology, and theory) about diverse populations, social work problems, and social phenomena. The current course emphasizes the intervention research aspects of translational science—applying basic research and evidence-informed theory in designing interventions, conducting efficacy and effectiveness studies, and engaging with implementation

science.

- In our earlier course you learned about social work research questions: that they need to be specific, feasible to study, and relevant. You learned about the funnel that runs from a general topic, to research questions, and ends in specific hypotheses (for quantitative study approaches). You also read about the potential for bias in the research questions that are asked. These same principles apply to social work research questions that lead to a better understanding of interventions—how interventions are studied flows directly from what is asked in the research questions.
- Much of our earlier course focused on “background” questions; our present course emphasizes “foreground” questions—questions about best practices and approaches to screening, assessing, and intervening in social work practice at all levels of intervention (including prevention).
- You have learned to distinguish between three major types of research questions that have powerful implications for research design, methodology, and analysis approaches: exploratory, descriptive, and explanatory questions. A great deal of emphasis in our current course is on explanatory questions where the effects or impact of interventions are what need to be examined (explained). We do, however, also address exploratory and descriptive questions in the context of understanding social work interventions.
- The significant role played by theory remains a very important topic in our understanding of interventions. The design of any intervention is typically informed by theory, although this is not always the case. Theory also helps inform social work practitioners about who might benefit (most or least) from a specific intervention, and about the mechanisms of change involved in how interventions have their impact.
- A great deal of attention in Module 2 of our earlier course was dedicated to understanding empirical literature—what it is (and is not), strategies for locating empirical literature, and how one

might analyze empirical literature to answer questions about diverse populations, social work problems, and social phenomena. That content remains highly relevant to the search and analysis of empirical literature to inform social work practice decisions and understand interventions.

- The topic of research ethics was woven throughout our earlier course. Research ethics continues to be highly relevant and important throughout our current course—including the topics we previously explored, as well as several new topics specific to understanding interventions.
- Content you learned regarding study methods (participants, measures, procedures, data collection approaches), data and statistical analysis, and presenting evidence applies to the current course content. Additional topics and skills related to these content areas are applied to understanding intervention throughout the new course.

Feel free to consult the [first coursebook](#) to refresh your knowledge and understanding of specific content as it is re-encountered—references to the relevant modules and chapters are presented in the present coursebook. The interactive format of the present coursebook is the same as the past coursebook and should seem familiar to you.

Module 2 Chapter 2: Formulating Practice Evidence Questions

“The search question is the engine that drives” a search for and review of evidence to inform practice (Bronson & Davis, 2012, p. 16). Thus, attention to the process by which practice questions are formulated warrants attention.

In this chapter you learn:

- the COPES framework for developing strong practice questions to inform the search for evidence,
- the PICO framework for developing practice questions,
- a working example where COPES was applied.

Question Formulation

The questions that drive a social worker’s search for evidence to inform practice decisions differ somewhat from the kind of questions that drive social workers to understand diverse populations, social work problems, and social phenomena. Two frameworks for helping develop questions for evidence-based practice are prevalent in the literature: **COPES** and **PICO**.



Using the COPES Framework to Formulate the Practice Question.

The COPES framework refers to engaging in a **C**lient **O**riented **P**ractical **E**vidence **S**earch. In this framework, the client system might be an individual client, individuals in a relationship, a family, a group, an agency or institution, a neighborhood or community, or larger social system.

Well-formulated COPES practice questions, according to Gibbs (2003) have four features:

1. *Client type and problem.* The practitioner to specifically identifies the client system, what problem the client system experiences, and the practice context in which intervention would or could occur. This is at the heart of the **C**lient **O**riented approach—the COPES question is central to the welfare of the client system and those whose lives are affected by the client. For example, a social worker in a hospital setting being able to address an individual's probable alcohol use disorder is important to that person, as well as to that person's family, friends, co-workers, and others (including the general public who share the road with this person when driving). The problem is complicated by the fact that feedback handled poorly leads to a person becoming become angry and resistant to seeking treatment for the suspected alcohol problem.
2. *What you might do.* This involves identifying a possible strategy for intervening around the identified practice problem. This includes whether the approach will be a specific treatment intervention, acting to prevent a problem, measuring to assess a problem, screening to assess risk, surveying clients, or engaging in some other specific action. Considering our alcohol use disorder example from step 1, the strategy might be to engage in screening for an alcohol problem and to provide feedback about the screening results using a

Screening, Brief Intervention, and Referral to Treatment (SBIRT) approach—SBIRT is based on principles of screening and feedback using a motivational interviewing (MI) approach.

3. *Alternative course(s) of action.* This involves the practitioner identifying at least one main alternative course of action for addressing the identified problem. This might possibly include doing nothing, observing and monitoring over time, or applying a specific intervention protocol. In our alcohol use disorder example, the alternative might be to monitor the problem over time if the screening result indicates a low likelihood of a substance use problem. Or, it might be to refer the individual for specialized assessment and treatment for an alcohol/substance use disorder if the screening result indicates a moderate or high likelihood of an alcohol/substance use problem.
4. *What you want to accomplish.* This feature specifies what will be accomplished with the intervention the evidence leads you to select. In other words, what outcome(s) should be expected if the intervention is successful—defining the goal of intervening in this way. This is directly related to the **P**actical aspects of the COPES question—not only does it have practical relevance to the client, it has relevance to social workers who encounter this problem with frequency in their routine practice. In our alcohol problem example, the goal would be for the person to engage in alcohol treatment if the drinking problem is serious enough or for the person to cut down on drinking if the problem does not warrant treatment but is causing problems for the person (and significant others).

Furthermore, the COPES question leads the practitioner's **E**vidence **S**earch. It contains information to guide the selection of key terms for an electronic search of literature and other appropriate sources of evidence. To complete our example, the COPES question could be phrased as follows:

If a person seen in the hospital setting is suspected of having an alcohol use disorder, would the SBIRT or Johnson Model Intervention be more likely to result in either becoming engaged in treatment for an alcohol use disorder (if warranted) or decreasing problematic alcohol use (if treatment is not warranted)?

Searching for evidence around this question led the social worker to support adopting an SBIRT approach within the medical/hospital setting. (The Johnson Model is similar to what you see on the television show, *Intervention*. The evidence does not consistently support applying this confrontational approach.)

In addition to specifying the four features of a well-constructed COPES question, Gibbs (2003) identified five general types of COPES questions:

- effectiveness
- prevention
- assessment
- description, and
- risk.

He provided examples of each in a table with the prior four features of well-constructed COPES questions across the top and these five question types down the left side. This format (adapted from Gibbs, 2003) guides practitioners in moving toward a well-specified COPES question.

| | client type and problem ^[1] | what you might do ^[2] | alternative course of action ^[3] | what you want to accomplish ^[4] |
|---------------|--|-------------------------------------|---|--|
| effectiveness | | | | |
| prevention | | | | |
| assessment | | | | |
| description | | | | |
| risk | | | | |

Here are examples Gibbs (2003) provided, edited for format (see originals at <http://www.evidence.brookscole.com/copse.html> and <http://www.evidence.brookscole.com/moredetails.html>). Note that [1]through [4]refer to each of the four features of well-formulated COPES questions running along the top of the table.

- *Effectiveness*. Effectiveness questions ask about the direct effects of exposure to an intervention for a specified type of client or population. Gibbs' (2003) first example:

"If disoriented aged persons who reside in a nursing home^[1]are given Reality Orientation Therapy^[2]or Validation Therapy^[3]which will result in better orientation to time, place, person^[4]?"

Gibbs' (2003) second example:

"Among adult criminal offenders on probation,^[1]will a group cognitive intervention program^[2] compared to no such program^[3] result in a lower recidivism (re-offense) rate^[4]?"



Prevention. The prevention type of question is also an effectiveness question, but it concerns interventions intended to avoid the problem occurring rather than treat it after the fact. Gibbs' (2003) first example:

"If sexually active high school students at high risk for pregnancy^[1] are exposed to Baby-Think-It-Over^[2] or to didactic material on proper use of birth control methods,^[3] will the former have fewer pregnancies during an academic year and better knowledge of birth control methods?^[4]"

Gibbs' (2003) second example is a little different from the previous examples in that the social worker does not yet know which specific intervention to search:

"Among Hmong and Asians less than sixteen years old,^[1] which gang prevention program^[2,3] will most effectively prevent them from joining the gang^[4]?"



- **Assessment.** Social workers engage in assessment at multiple points in the process of intervening with client systems. Assessment involves applying standardized procedures or measures with the goal of determining whether the client system demonstrates a particular problem or strength. Assessment is also used to determine whether an intervention has achieved its goals. The literature search can help identify valid and reliable approaches to screening and assessment. Gibbs' (2003) first example:

"If aged residents of a nursing home who may be depressed or may have Alzheimer's Disease or dementia^[1] are administered depression screening tests^[2] or short mental status examination tests^[3] which measure will be the briefest, most inexpensive, valid and reliable screening test to discriminate between depression and dementia?^[4]"

Gibbs' (2004) second example also presumes that the social worker has little pre-existing knowledge about the best assessment tools:

"To detect children with shaken baby syndrome^[4], which

assessment or diagnostic procedure will provide the most valid and reliable determination^[2,3] that the child has been injured by shaking^[1]?”

Gibbs (2003) offered a third example of an assessment question, one which is actually about screening, which is a step that precedes a more detailed, in-depth assessment.

“For hospitalized aged persons suspected of having dementia,^[1] which rapid assessment measure^[2,3] will most quickly, reliably, and validly identify demented patients^[4]?”

And, Gibbs offered an example that looks at assessment of an interdisciplinary team’s functioning, rather than assessing specific client systems—he classified this example as descriptive, but it seems to fit as an assessment question (see what you decide):

“If members of a hospital team who are concerned about team functioning^[1] take the Preliminary Checklist (Clinical) Team Effectiveness test^[2] or take the Interdisciplinary Team Weekly Inventory^[3] which measure will most reliably and validly reflect the team’s ability to accomplish tasks^[4]?”



- **Description.** This is what Gibbs (2003) said about description questions:

“Description questions most often concern surveys of client needs or client satisfaction, but can include any kind of effort that involves observations of clients within a sample and generalizations made from that sample. Description Questions

can include qualitative studies that often seek an in-depth understanding for client experiences and perceptions” (retrieved from <http://www.evidence.brookscole.com/moredetails.html>)

Gibbs (2003) offered a description question to consider: “Among children in integrated children’s homes or orphanages,^[1] at what age^[2,3] do they first begin to see themselves as being either of white or black race^[4]?”

Another example that Gibbs (2003) presented starts out sounding like an effectiveness or assessment type question, but becomes a description question in the end: “If family members of persons diagnosed with aphasia meet in a support group^[1] and receive a short client satisfaction questionnaire for all support group participants^[2,3] which will the clients list as their area or areas of greatest and least satisfaction^[4]?”



- Risk. Risk is another type of assessment question that attempts to predict the future. In medical terms, this is often called the prognosis. It is about the probability of a specific consequence happening and often includes specification of that consequence and the time frame under consideration. Gibbs’ (2003) first example “If crisis line callers to a battered women shelter^[1] are administered a risk assessment scale by telephone^[2] or we rely on practical judgment unaided by a risk assessment scale^[3] then will the risk assessment scale have higher reliability and predictive validity for violent behavior^[4]?” Gibbs’ (2003) second example is:

“For abused or neglected children placed in foster care by a

protective service worker,^[1] which risk assessment measure will provide the greatest predictive accuracy^[2,3] to predict re-abuse when children are placed back into their homes with their families^[4]?” A third example provided by Gibbs (2003) is: “Among chronically mentally ill clients,^[1] is the Brief Psychiatric Rating Scale^[2] compared to other suicide prevention instruments^[3] as accurate at predicting actual suicide or suicidal behaviors^[4]?”



A final note about the examples presented by Gibbs is warranted: these were published prior to a point in time when social work professionals were conscientiously addressing the language that we use to describe people and populations. If he were writing today, Gibbs would likely revise many of these COPES questions to eliminate labelling language (e.g., “disoriented aged persons,” “adult criminal offenders,” and “chronically mentally ill”). It is a sign of the times that we address this issue of language use at all.

Using the PICO Framework to Formulate the Practice Question.

The PICO framework comes from literature on evidence-based practice in medicine and nursing. This is demonstrated in its full name, the Patient, Intervention, Comparison, Outcome framework. Social workers refer to clients, client systems, or consumers of services; medical professions refer to patients. The PICO framework results in the comparison of two intervention options.

- *Patient* (or client or consumer) refers to specifically identifying the characteristics of the population around which the search for evidence is directed.
- *Intervention* requires specification of the interventions or other practice activities (e.g., type of assessment or clinical test, treatment, prevention, or policy) under consideration.
- *Comparison* is the act of comparing the intervention options identified in the prior step.
- *Outcome* refers to the goal or goals to be achieved by intervening.

Here is an example of a PICO practice question mapped with the numbered COPES steps[1-4].

“If an elder residing in a nursing home[1] participates in a pet therapy program[2] or attends an adult day program[3] which intervention will result in lower depression[4]? (Heltzer, n.d., p. 2)

As you can see, the resulting practice question is structured just as it would be using the COPES approach.

Where COPES and PICO Questions Fit

As you may recall from our earlier course, there are two general categories of questions that practitioners might ask: background and foreground questions. COPES and PICO questions fit under the heading of foreground questions because the practitioner seeks evidence to inform professional practice decision-making. COPES and PICO questions are complex forms of foreground questions. The COPES framework was developed on the PICO structure, but designed for use in non-medical contexts (Gibbs, 2003), which explains why they appear to be so similar. The main reason for sharing the PICO framework in a social work course is because social workers often work on interdisciplinary teams with practitioners trained in medical models. Knowing how to engage in COPES prepares social workers to function with colleagues trained in PICO, and vice versa.

Working Example: Addressing High School Dropout

Here is an example of a practice question that we can follow through the next chapters. The example details steps in evidence-based practice decision-making described in a chapter by Kelly and Franklin (2011).



Background

The Forest Grove (Illinois) community was charged by the school board to address a problem with too-high rates of high school dropout:

“...we can’t keep doing the same things we’ve been doing and expecting a new result. It’s time to make some changes in our dropout interventions...” (Kelly & Franklin, 2011, p. 141).

A local newspaper reported a dramatic increase in dropout from the community’s two high schools, particularly among the African American and Hispanic student populations. The schools were being accused of “pushing out” students as a means of improving scores related to meeting standards imposed by the 2002 *No Child Left Behind* policy. District data indicated that the rate of students failing to complete high school in four years at Forest Grove North (FGN) and Forest Grove South (FGS) had doubled from 12% to 24% over the past 10 years, and the dropout rate among African American and Hispanic students was disproportionately high: 35%. First, the team generated a general question:

What are best practices for preventing student dropout from high school?

This question was too general to direct the practical, client-oriented search for evidence (the school district/board being the client). As a next step, a social worker on the team explored literature concerning two background questions, one describing the problem seriousness and the other describing factors contributing to observed disparities in student dropout:

What are the outcomes of the problem?

How are poor and minority youth at risk of dropping out?

This general literature review indicated that students who drop out, never returning to complete a general high school equivalency education, are at risk of earning less, having poorer job prospects, poorer health, and incarceration as adults (Kelly & Franklin, 2011). As far as disparities were concerned, the evidence indicated nation-wide concerns with higher dropout rates in urban (compared to suburban) communities, and among ethnic minority and male students. Much of this was attributed to lower socioeconomic status/income in communities with high dropout rates. These two questions and the descriptive evidence were followed up with the team developing a specific practice question. Engaging in the COPES process resulted in the team developing their practice question as:

“If students deemed at risk for dropping out of high school are given a specific school-based dropout intervention program or standard school-based intervention, such as counseling and academic advising, will the school-based dropout program produce better outcomes, specifically helping youths stay in school and graduate on time? (Kelly & Franklin, 2011, p. 147).

In the next chapter, we learn how the team went about identifying evidence to inform their practice decisions.

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=92>

Chapter Summary

In this chapter you learned two similar approaches to developing practice questions that guide the search for evidence to inform intervention: PICO and COPES. The steps in the process were demonstrated with examples. This leads to our next topic: where practitioners might seek evidence that will help them address their practice questions.

Module 2 Chapter 3: Logic Models and Theory Informed Intervention

In our earlier course you learned about the relationship between theory and research. In this course about understanding social work interventions the relationship between theory and intervention takes center stage. The design of an intervention is typically informed by theory and evidence, and the research about that intervention further develops theory and evidence. Over time, with repetitions of this cycle, social workers have sufficient evidence to inform their practice decisions. The logic model is an approach to systematically defining an intervention and relating the theory and evidence base to its development.

In this chapter, you learn:

- how theory and logic models inform intervention design

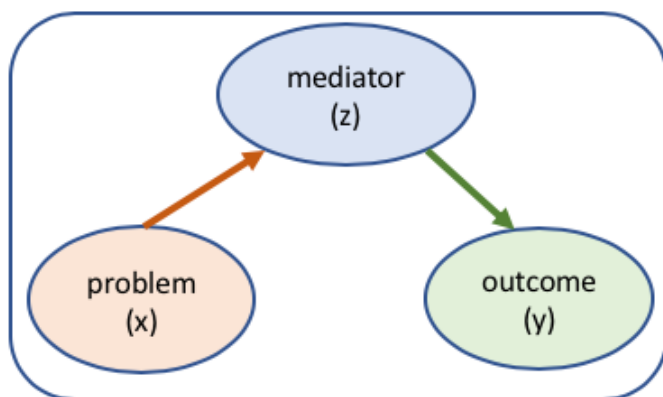
The Role of Theory in Intervention Design

The PICO or COPES process helped define the practice problem or question. The next step in the process of designing an intervention is to apply theory in mapping the ***mechanism of change***. Evidence about mechanisms of change answers questions of how change occurs or is facilitated rather than restricting focus on what changes occur. In other words, mechanisms of change research is about change processes, not just change outcomes. For example, a body of literature suggests that

In our earlier course you learned to understand social work

problems and social phenomena, as they occur across diverse populations, in terms of theory and evidence. This is the change theory and evidence applied by intervention planners. Theory helps planners identify “targets for change”—leverage points where it makes the most sense to intervene and how intervening at those points might lead to the desired change outcomes (Fraser, Richman, Galinsky, & Day, 2009). Practical targets for change are limited to factors that can be modified through intervention; it is impractical to target phenomena that cannot be modified through intervention (e.g., exposure to past events cannot be undone, but reactions/responses to them might be modifiable). In the empirical literature, these factors are often identified as **mediators**.

This diagram shows the rationale involved: the outcome (y) is considered a direct effect of the problem (x), but the relationship between them is influenced or controlled by a mediating process (z). This mediator becomes a prime target for change—by changing it (z), the problem (x) might have a different impact or outcome (y).



For example, empirical evidence tells us that a specific biological process in the brain (accumulation of amyloid plaques) is a direct cause of Alzheimer’s disease. The problem (x) in this case is amyloid

plaque formation in the hippocampus and thalamus regions and causes—or at least heavily influences—the outcome, which is Alzheimer’s Disease (y). Recent evidence identified a mediator of this relationship (z)—lack of sleep causes an accumulation of beta-amyloid proteins that “clump” together forming plaques in the human brain (Shokri-Kojori et al., 2018). While we are not able to change the neurobiology whereby plaque formation (x) causes Alzheimer’s Disease (y), we might intervene around the mediator, lack of sleep (z) to help prevent or delay the onset of Alzheimer’s disease due to beta-amyloid plaque formation. Evidence about lack of sleep as a mediator points to a potentially modifiable intervention target: interventions to promote good “sleep hygiene” practices. This is suggested because beta-amyloid increased by about 5% with study participants losing one night’s sleep—it is a cellular waste product that clears away during sleep. It is unknown whether or to what extent a night of rest can reverse the increase in beta-amyloid that accumulated during the sleep deprivation period (NIAAA, 2018).



Ideally, intervention plans are shaped by causal evidence, not merely correlational evidence. Systematic intervention planning and design proceed from this phase to the creation of a logic model.

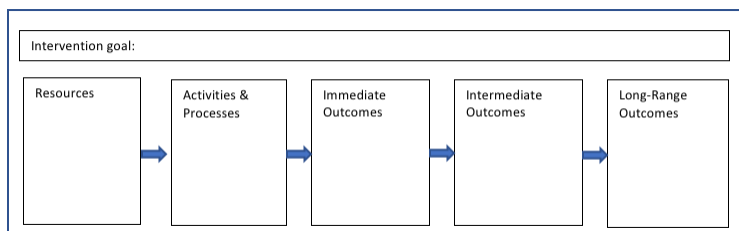
Logic Models Explained

A **logic model** is a conceptual map for a planned intervention, one

that shows the logical connections between **inputs** and **outputs** in the intervention process and the theory of change (mechanisms of change) underlying the intervention plan (Fraser et al., 2009; Harris, 2010). Inputs are those core elements and resources necessary for implementing the planned intervention. These inputs include personnel, time/effort, space, finances, supervisory, technology, materials, and other resources necessary to deliver the intervention (Harris, 2010). Outputs, on the other hand, are the changes anticipated to result from the planned intervention. The outputs include initial/immediate products or outcomes (**proximal** outcomes), intermediate impacts or outcomes, and long-range/final outcomes (**distal** outcomes) associated with the planned intervention (Kapp & Anderson, 2010). The logic model serves the following purposes (adapted from Harris, 2010, p. 56):

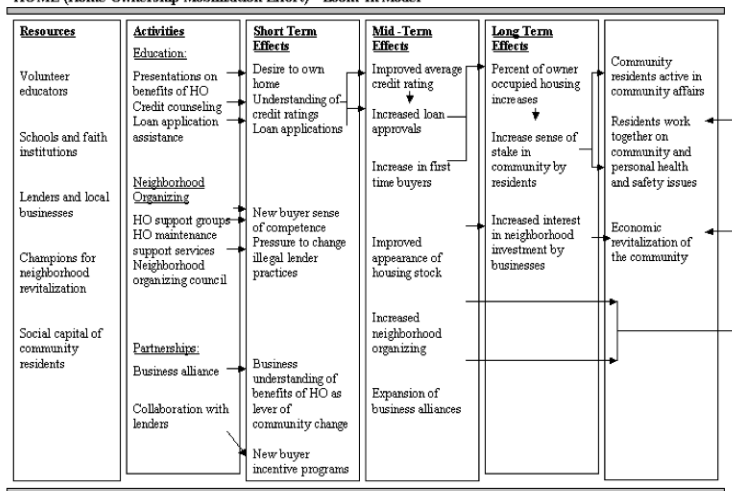
- mapping an intervention during planning and evaluation phases;
- documenting the reasoning behind the intervention activities;
- “a tool to facilitate stakeholder insight and reflection”;
- “a tool to inform monitoring and the development of benchmarks” for evaluating the intervention.

Here is a sample template for an intervention or program logic model. Inputs and Outputs/Outcomes were previously described. Activities are those change-producing processes involved—the crucial intervention components informed by evidence. Activities and processes might be separated into “staff activities” and “program processes” to distinguish between what is delivered (staff activities) and what clients do (program processes) for a more nuanced logic model (Kapp & Anderson, 2010). Note that a separate logic model should be developed for each intervention goal if multiple goals are targets of a single intervention, program, or policy.



Here is an example of a completed logic model from the HOME intervention (Home Ownership Mobilization Effort) in the United Kingdom, an intervention to improve communities through increased home ownership (retrieved from <https://ctb.ku.edu/en/table-of-contents/overview/models-for-community-health-and-development/logic-model-development/example>).

HOME (Home Ownership Mobilization Effort) - Zoom-In Model



Stop and Think



Take a moment to complete the following activity.

Scenario: A county wishes to intervene to reduce severe health disparities between its urban and suburban populations. Their search for evidence led them to an intervention plan that increases safe, appealing spaces and opportunities for physical activity and provides affordable, appealing access to healthful food. Thinking about how their logic model might be constructed, answer the following multiple-choice questions.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=95>

Chapter Summary

In this brief chapter you were introduced to the logic model for planning and evaluating interventions. This is an approach that

clearly identifies the logic in designing an intervention to achieve specified outcomes based on the intervention “inputs,” activities, and processes. A logic model is founded on theory and evidence informing the choices of inputs, activities, and processes, as well as determining the likely short- and long-range impacts of the intervention. In our next chapter, we consider approaches to locating that evidence.

Module 2 Chapter 4: Locating Intervention Evidence

This chapter is about identifying approaches and resources for evidence used in making social work practice decisions. As you learned in earlier modules, evidence-based practice incorporates practitioner experience and expertise, client preferences and circumstances, and the best possible research evidence available. The aim of this chapter is to answer the question: “If social workers are committed to using the best available evidence, where do they locate it?” General approaches and resources for locating evidence to answer practice questions are discussed.

In this chapter you read about:

- limitations of literature reviews
- systematic reviews
- scoping reviews
- meta-analysis reviews
- locating individual research studies.

Limitations of Literature Reviews

The introduction to an empirical article often includes a literature review conducted by the authors. In addition, literature reviews sometimes appear in the published literature. These may provide important information for practitioners to consider, however they are seldom sufficient by themselves. Bronson and Davis (2012) summarized the problems facing social work practitioners who rely on a literature review to inform their practice:

- a great deal of the research literature lacks sufficient

intervention detail to inform practitioners about *how* to implement the best practices reported in the articles;

- outcome studies often present the practitioner with contradictory results;
- authors' reviews often lack sufficient background concerning underlying assumptions and biases that may influence the credibility and relevance for practice decision-making, particularly within specific practice settings or with specific types of clients.

A literature review is only as thorough as the search for literature; an extensive review is more informative than a search limited by factors such as:

- the discipline where the literature appears, ignoring literature from other professions and disciplines;
- the time frame covered in the search (e.g., limited to just a few years);
- the nation where the work was conducted (e.g., an “Americentric” search might miss important contributions from other countries);
- only reviewing what is easily accessed from a library.

Generic literature reviews also may be conducted in an unsystematic fashion, leaving a reader uncertain as to the extensiveness of the coverage. This leads to discussing the qualities of more systematically conducted types of reviews: systematic, scoping, and meta-analysis reviews.

Systematic Reviews

The topic of **systematic reviews** was briefly introduced in our earlier course. In this chapter, you are provided with greater detail about

these particularly useful tools for practitioners in many disciplines. Systematic review is a methodical process of summarizing evidence that results in a product: the systematic review report. What the systematic review method provides is:

- *“An attempt to include all relevant research, including published and unpublished studies and those with contradictory findings;*
- *A systematic method of collecting data from the existing studies to allow for meaningful synthesis;*
- *Explicitly stated inclusion and exclusion criteria for identifying relevant research; and*
- *Transparent search methods to allow for replication”*(Bronson & Davis, 2012, p. 15).

The application of systematic review methodology is how systematic reviews differ from general literature reviews. Systematic review reports are published in various dissemination outlets, including:

- professional journals;
- government agencies and government resources, such as:
 - various institutes of the National Institutes of Health/NIH (National Institute of Mental Health/NIMH, National Institute on Alcohol Abuse and Alcoholism/NIAAA, National Institute on Drug Abuse/NIDA, National Institute of Child Health and Human Development/NICHHD, National Institute on Aging/NIA, National Institute of Environmental Health Sciences/NIEHS, and National Institute on Minority Health and Health Disparities/NIMHD),
 - Substance Abuse and Mental Health Services Administration/SAMHSA,
 - Agency for Healthcare Research and Quality (AHRQ),
 - Office of Justice Programs/OJP,
 - Office of Juvenile Justice and Delinquency Prevention,

- Child Welfare Information Gateway at childwelfare.gov/topics/responding/child-protection/evidence/,
- National Registry of Evidence-based Programs and Practices, NREPP at <https://www.samhsa.gov/nrepp>,
- U.S. Department of Education's What Works Clearinghouse/WWC at <https://ies.ed.gov/ncee/wwc/>and,
- Centers for Disease Control and Prevention/CDC;
- United States Preventive Services Taskforce (USPSTF) at www.uspreventiveservicestaskforce.org;
- Campbell Collaboration at www.campbellcollaboration.org;
- Cochrane Collaboration database at <http://us.cochrane.org> and the Cochrane Community Database of Abstracts of Reviews of Effects (DARE);
- the patient information resource, Informed Health Online, at www.informedhealth.org;
- PubMed database of systematic reviews at the U.S. National Library of Medicine in NIH at <https://www.ncbi.nlm.nih.gov/pubmed/>.

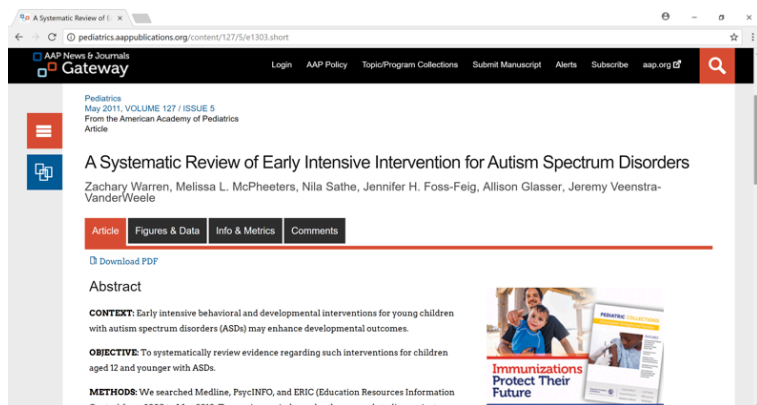
Many of these sources address medical practice questions but may also include behavioral health questions of interest to social work practitioners. The Campbell Collaboration may be of particular interest to social work practice and policy (<http://www.campbellcollaboration.org>), including umbrella topics such as:

- crime and justice
- disability
- education
- international development
- social welfare
- knowledge translation and implementation

One way to locate systematic reviews about a specific practice

topic or question is to place the phrase “systematic review” in your electronic search terms. Let’s consider a practice example where a social worker needed to find evidence to inform practice decisions concerning the practice question: what is the best intervention for promoting development in young children with autism? The social worker wanted to inform a family about options for early and intensive intervention to improve outcomes for their 4-year old child with autism. So, the social worker entered the search terms: “systematic review parenting education intervention” into Google Scholar. The result appears in Figure 3-1:

Figure 3-1. Result of a systematic review search for autism interventions

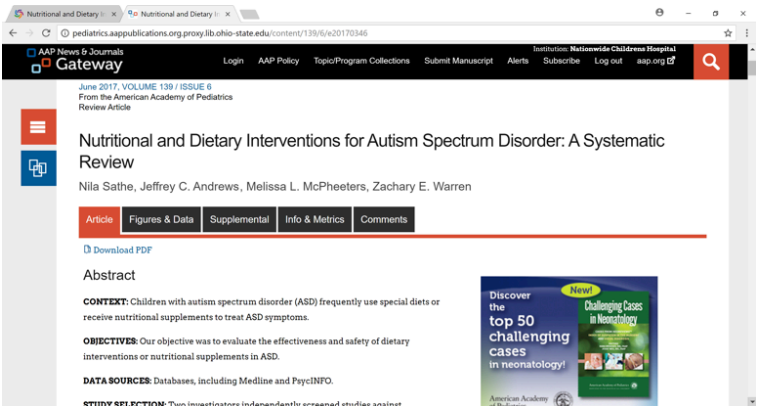


Reading the abstract for this review indicated that the objective was a good match for the practice question, and the social worker learned generally how the review was conducted (methods). While this looked promising, the results and the conclusions of the published review were somewhat disappointing because the strength of the evidence the reviewers identified was generally low. Two intervention approaches showed benefits in some aspects of development for some children (the Lovaas-based intervention and the Early Start Denver Model), but which children did and did not

benefit was poorly understood (Warren et al, 2011); parent intervention showed some short-term gains, but long-term durability was unclear. So, the social worker did not yet have a strong evidence-based answer to the practice question and turned back to the search for systematic reviews related to autism.

The next article that came up reviewed nutritional and dietary intervention strategies (Figure 3-2).

Figure 3-2. Next result of a systematic review search for autism interventions



In reading the abstract, the social worker was not as certain about the goodness-of-fit with the practice questions regarding promoting development in young children with autism but continued reading through the contents of the abstract. Again, the strength of the evidence was low, the results were inconsistent and were not demonstrated to persist over time (Sathe, Andrews, McPheeters, & Warren, 2017). The social worker still had little evidence to inform practice decisions.

Next the social worker considered the possibility that perhaps the search question was too broad and general. Autism is a complex problem with many variants and theories about etiology and factors that influence its course, thus it has many possible avenues for

intervention. The social worker then considered what the family described as what they were most interested in pursuing with this child at this time. They were interested in behavioral approaches rather than medication. The next reference that the social worker located was the Systematic Review of Early Intensive Behavioral Interventions for Children with Autism (Howlin, Magiati, & Charman, 2009). This review applied inclusion criteria that the study's population had to be children under the age of six years, which made it a good fit for the 4-year old family member. The review results indicated that there exists evidence to support the effectiveness of the Early Intensive Behavioral Intervention (EIBI) approach initially developed at UCLA and offered through various programs elsewhere. The review also specified that the observed effectiveness was not consistent across all children, and observed effects typically appeared within the first 12 months of intervention. The results of the review were similar to those for a similar systematic review conducted by another team, and this confirmation served to enhance the social worker's confidence in the results. This information gave the social worker a starting place to seek local programs based on this approach in referring the family, and a timeline for evaluating the impact of the intervention (12 months). However, the social worker might also want to extend the literature search to see whether more recent evidence exists about this or other relevant interventions.



The results of the EIBI autism review were based on over 640 published studies, making the single review highly useful to a practitioner who does not have time to locate, read, critically analyze, and draw conclusions from so many pieces. Furthermore, individual practitioners may find it difficult to be systematic in the process of critically reviewing many pieces of literature. It helps to understand what goes into a systematic review to know how to trust the authors' critical review process in place of one's own critical analysis.

Process of a Systematic Review

Authors of a research skills workbook delineated 12 stages in the systematic review process they called “meta-synthesis” (Corcoran & Secret, 2013). Here is an overview of what they outlined, integrated with content from other resources about systematic reviews (Bronson & Davis, 2012; Higgins & Green, 2011; Littell, Corcoran, & Pillai, 2008).

- *Develop specific, answerable research question.* The project begins with clearly formulated objectives and the research question is specific. The COPES or PICO approach you learned in Chapter 2 might be applied at this stage.
- *Form a review team.* The team should include a diverse skill set, covering a range (substantive on the content area and relevant methodological/technical expertise).
- *Determine explicit inclusion/exclusion criteria.* These criteria concerning which reports to include in the review should be specific about the conditions, circumstances, populations, types of intervention, intervention settings, outcome indicators, and study design/approach.
- *Develop a protocol.* This is a set of written guidelines about procedures and methods the team will use in the review

process. This is done prescriptively (before the work begins), it is not meant to be a post hoc description after the work is completed.

- *Implement a reproducible strategy for study identification.* This strategy is defined in the protocol and incorporates the inclusion/exclusion criteria that were established in prior stages. The strategy needs to result in identification of all relevant studies, including those that might be unpublished. The review should include outcomes that were adverse, beneficial, or “no observed difference” on the meaningful (primary) outcomes.
- *Screen titles and abstracts.* The team applies their pre-determined criteria and protocol to determine which of the previously identified studies are potentially relevant to the review.
- *Retrieve potentially relevant study reports.* The team utilizes strong library and electronic search approaches to gain access to the reports (including unpublished reports) identified as potentially relevant. A strong system of organizing the secured reports is important at this stage, as it can reduce duplicated effort.
- *Team members judge each study.* As in the case of coding qualitative data, two or more individuals should independently determine whether each report represents an eligible study—based on the predetermined eligibility criteria. Reviewers document their decisions, and discrepancies are resolved by the team.
- *Data from eligible studies are extracted.* For each study deemed eligible, data about the study are copied into standardized forms (created as part of the protocol). As in the judging phase, this is done by at least two team members, and disagreements are resolved by the team.
- *Rate study quality.* For each study deemed eligible, at least two raters critically appraise its quality (scientific integrity and design adequacy) and its qualities (sampling, data collection

methods, results, and conclusions). As before, this is documented on standardized forms and disagreements are resolved by the team.

- *Describe key features of included studies.* Tables are used to document descriptions of these elements, created as part of the review protocol. Common categories are: author's/authors' purpose(s)/study aims; study design and approach; sample information; measures/data collection; data analysis/results; conclusions.
- *Synthesize results.* Team members work with the data accumulated in the prior steps to draw conclusions, answering the initial research question.



Strength of Evidence in a Systematic Review

Systematic reviews differ, to some extent, in terms of eligibility criteria concerning strength of evidence (scientific rigor) of the studies included. Study replication, where different investigations lead to the same observed results, increases confidence in reliability of the intervention study findings. Also, interventions found to be effective in different contexts (with different populations, in

different settings, and/or by different service providers) are considered especially meritorious in terms of strength of the evidence.

In many instances, the only studies included in a systematic review are randomized control trials (RCTs) since these are believed to represent the strongest evidence possible (internal validity). This is the case for the Cochrane review system, for example. A table describing levels of evidence summarizes a prevailing belief system concerning the strength of evidence different studies present (adapted from Ackley, Swan, Ladwig, & Tucker, 2008, p. 7 and the Winona State University library <https://libguides.winona.edu/c.php?g=11614&p=61584>). In this framework, the “best” evidence for informing practice is associated with the top of the hierarchy (Level 1), the “weakest” evidence is at the bottom (Level 7), progressing from “green means go” toward “orange means caution.” This table follows the decision rules presented in the Oxford guidelines concerning levels of evidence (Centre for Evidence-Based Medicine, <https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>).

| Level of Evidence | Description |
|-------------------|--|
| Level 1 | Evidence from a systematic review or meta-analysis of all relevant RCTs or evidence-based clinical practice guidelines based on systematic reviews of RCTs or at least three RCTs of good quality that have similar results. |
| Level 2 | Evidence obtained from at least one well-designed RCT. |
| Level 3 | Evidence obtained from well-designed controlled trials without randomization (random assignment to conditions; quasi-experimental designs). |
| Level 4 | Evidence from well-designed case-control or cohort studies. |
| Level 5 | Evidence from systematic reviews of descriptive and qualitative studies (meta-synthesis) |
| Level 6 | Evidence from a single descriptive or qualitative study. |
| Level 7 | Evidence from the opinion of authority or experts. |

Scoping Reviews

An alternative in line with the spirit and methodology of systematic reviews is called a **scoping review**. Scoping reviews are conducted

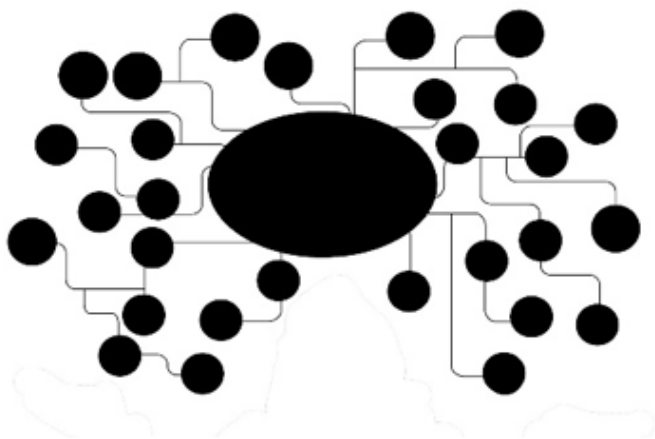
with the aim of “mapping” the literature in an area (Arksey & O’Malley, 2005). The result of the scoping effort is an analysis of the main sources and types of evidence available on a topic or about an intervention approach. Sometimes a scoping review serves as a prelude to conducting a formal systematic review, but there are other reasons for them being conducted: they help identify gaps in existing literature, summarize a wide-ranging body of evidence, and provide decision-makers with a summary of relevant literature (Arksey & O’Malley, 2005).

“The scoping review has become an increasingly popular approach for synthesizing research evidence” (Pham et al., 2014, p. 371). This approach is used when the existing literature in an area is wide ranging, complex, and diverse in terms of study approaches, methods, and levels of evidence available. Scoping reviews apply systematic procedures and review criteria, resulting in a “map” depicting an area of research—particularly an emerging area where the approaches to gathering evidence remain diverse.

Conducting a scoping review follows the stages and methodological rigor previously described for conducting a systematic review, the major differences being:

- the research question tends to be broader, less well-defined in a scoping review;
- greater “tolerance” allowed in the inclusion/exclusion criteria related to quality or level of evidence—this allows reviewers to include evidence of greater diversity (Arksey & O’Malley, 2005).

In addition to the label “scoping review,” these are sometimes called scoping studies or systematic mapping of evidence/literature.



Meta-Analysis Reviews

Meta-analysis is an interesting hybrid approach to evidence, combining elements of a systematic review with the statistical analytic approaches used in generating results from data. In reviewing literature, investigators often find that there exist multiple published studies on a topic, but that conclusions cannot be confidently drawn from the results because the individual studies were each conducted with relatively small sample sizes. The statistical analytic approach comes into play when investigators statistically combine the smaller samples from these individual studies into a single, larger sample. This increases the investigators' power to potentially detect relatively small effects in the data that the smaller samples could not detect with confidence.

The difficulty in conducting meta-analysis lies in gaining access to the data needed for the combined analysis. Meta-analysis investigators do not necessarily need access to original data provided by study participants to the original investigators. However, they do need access to certain descriptive statistics and a

clear understanding of how the variables of interest were measured in each study.



An example meta-analysis report comes from Kennedy et al., (2016), addressing the question of whether parent-child interaction therapy (PCIT) when utilized with families engaged in or at risk of engaging in child physical abuse is effective in preventing future episodes of child maltreatment. PCIT literature indicated that it was a best practice with other types of families. The article's authors described the steps taken in conducting this meta-analysis.

- They first conducted a systematic review of literature concerning the use of parent-child interaction therapy (PCIT) specifically with families who either engaged in or were identified as being at risk of physical child abuse. Limiting this review to studies that applied an experimental or quasi-experimental design resulted in their identifying six relevant studies.
- Next, they identified the outcome variable(s) of interest that these studies had in common. The investigators focused on variables that depicted three parent-level outcomes: physical

abuse recurrence, child abuse potential, and parenting stress (not child-level outcomes).

- The investigators determined the **effect size** of PCIT intervention compared to treatment as usual conditions, to see if the statistically significant differences were meaningful. They found that the size of the effect produced by using PCIT rather than the tested alternatives varied considerably across the different individual studies. When analyzed together, the effect size was medium for reducing physical abuse recurrence, small for reducing child abuse potential and parenting stress across the reported studies. This makes it reasonable to consider as a practice option, at least in terms of reducing physical abuse recurrence.

As you can see, the meta-analytic approach is a potentially powerful tool for synthesizing literature on a specific topic. It helps sort out the variability that often exists in individual study results. It follows a specific systematic methodology and is reproducible when well-described, which meets our criteria for empirical literature.



Locating Individual Reports

Conducting a search of literature that will include individual reports and articles is an important adjunct to existing systematic, scoping, and meta-analysis reviews. The handbook on conducting systematic reviews presented by the Cochrane Collaboration (<https://handbook-5-1.cochrane.org/>) describes how to set up a search for relevant studies that can form a basis for conducting a general literature search (they call it a narrative review as opposed to a systematic review). An important point made in their handbook is the distinction (and tension) between achieving **sensitivity** and **precision**.

- Sensitivity refers to greater degrees of comprehensiveness of a search—casting a wide net, so to speak. Operationally, they define sensitivity as “the number of relevant reports identified divided by the total number of relevant reports in existence,” (section 6.4.4) although the latter is like a population—its dimensions can never really be known. The idea here is that the larger this ratio, the greater the sensitivity, and the more the search is inclusive. However, in casting this wider net, the greater the number of irrelevant reports that are also included in the search results.



- Precision, on the other hand, refers to the fewest number of irrelevant results—hitting the center of the target, so to speak. Operationally, they define precision as “the number of relevant reports identified divided by the total number of reports

identified” (section 6.4.4). In other words, precision is about reducing how much wasted effort went into identifying reports that are irrelevant. It is about being efficient.



In practical terms, the sensitivity and precision of your search are dependent on the search terms you use to guide your search. Sensitivity increases with the use of more synonyms, related terms, and variations in spelling. For example, searching for articles about social work interventions for responding to natural disasters might include at least the following search terms:

- natural disaster
- natural disasters
- disaster distress
- disaster relief
- disaster services

A more sensitive, lower precision list might include:

- traumatic events
- traumatic stress events
- tornado relief
- hurricane relief
- earthquake relief

- earthquake disaster recovery
- tsunami relief
- FEMA responses
- Red Cross responses
- ASPCA responses

Even more sensitive, possibly less precise would be naming specific events, such as 911, Hurricane Sandy, Hurricane Maria, Hurricane Katrina, Hurricane Florence, the Oklahoma tornadoes, the San Francisco earthquake of 1989, Mount St. Helens eruption, and other named events. If you wished to include international disaster relief efforts, you might include the December 26th tsunami, the 2018 North Sulawesi tsunami, and the 2011 earthquake and tsunami in Japan.



The content learned in Module 2 (Chapter 3) from our prior course concerning how and where to locate empirical evidence applies to the pursuit of evidence about interventions, as well. You learned about:

- Different types of journals. In addition to topic specific journals (e.g., *Journal of Social Work Practice in the Addictions*, *Child and Family Social Work*, *Journal of Gerontological Social Work*, *Affilia—Journal of Women and Social Work*, *Social Work in Health Care*, *Health and Social Work*, *Social Work in Public Health*, *Social Work with Groups*, *Social Work in Mental Health*), the profession has numerous generic journals and journals in allied disciplines to consult—just a sample are listed here.
 - *Social Work*
 - *Journal of the Society for Social Work and Research*
 - *Social Work Research*
 - *Research on Social Work Practice*
 - *Journal of Social Work*
 - *The British Journal of Social Work*
 - *Journal of Social Work Practice*
 - *Clinical Social Work Journal*
 - *Social Work Today*
 - *Social Service Review*
 - *International Social Work*
 - *Social Policy and Administration*
 - *Journal of Orthopsychiatry*
 - *Journal of Marriage and Family*
 - *Family Relations*
 - *Trauma, Violence and Abuse*
 - *American Journal of Community Psychology*
 - *Journal of Social Policy*
- Different types of abstracting and indexing databases. A sample are listed here.
 - MEDLINE/PubMed
 - PsycINFO
 - Campbell Collaboration
 - Cochrane Collaboration.
- Government sites (reproduced from your Module 2 Chapter 3

readings).

- Agency for Healthcare Research and Quality (AHRQ) at <https://www.ahrq.gov/>
- Bureau of Justice Statistics (BJS) at <https://www.bjs.gov/>
- Census Bureau at <https://www.census.gov>
- Centers for Disease Control and Prevention (CDC) at <https://www.cdc.gov>
- Child Welfare Information Gateway at <https://www.childwelfare.gov>
- Children's Bureau/Administration for Children & Families at <https://www.acf.hhs.gov>
- Forum on Child and Family Statistics at <https://www.childstats.gov>
- National Institutes of Health (NIH) at <https://www.nih.gov>, including:
 - National Institute on Aging (NIA at <https://www.nia.nih.gov>
 - National Institute on Alcohol Abuse and Alcoholism (NIAAA) at <https://www.niaaa.nih.gov>
 - National Institute of Child Health and Human Development (NICHD) at <https://www.nichd.nih.gov>
 - National Institute on Drug Abuse (NIDA) at <https://www.nida.nih.gov>
 - National Institute of Environmental Health Sciences at <https://www.niehs.nih.gov>
 - National Institute of Mental Health (NIMH) at <https://www.nimh.nih.gov>
 - National Institute on Minority Health and Health Disparities at <https://www.nimhd.nih.gov>
- National Institute of Justice (NIJ) at <https://www.nij.gov>
- Substance Abuse and Mental Health Services Administration (SAMHSA) at <https://www.samhsa.gov/>
- United States Agency for International Development at <https://usaid.gov>

Finally, a helpful way to build a literature search is to look at citations used in articles that you found to be of interest/relevance. On the flip side, you can locate a citation index that lists articles which cited an article that you found to be of interest/relevance.

Working Example: Addressing High School Dropout

Returning to the example about best practices for preventing high school dropout from chapter 2, the school district team examined literature about the risk factors for dropping out. They found that the reasons for dropping out fit into three categories: individual, family, and school-related reasons (Kelly & Franklin, 2011). The team first turned to online clearinghouses, five of them, as well as recent textbooks, and article databases (e.g., EBSCO, PsycInfo, and Social Work Abstracts). The team specified their keyword search terms used in the article searches (e.g., effective school dropout prevention programs, school dropout and prevention, and others). Their effectiveness review criteria were:

1. *The program being studied had an experimental or quasi-experimental design with a control or comparison group and a sample size that allowed for statistical power.*
2. *The studies had been conducted with students that had similar demographics (low to middle-income socioeconomic status [SES] white, African American, and Hispanic) and could be generalized to the specific FGN and FGS student population.*
3. *Findings from the programs in the study showed sustained treatment effects after a minimum of a one-year follow-up.*
4. *The study had materials that could be easily accessed for implementation via Web site or contact person with clear instructions on how to receive training related to the intervention (Kelly & Franklin, 2011, p. 149).*

In the end, the team identified 12 interventions that met their criteria.



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=96>

Chapter Summary

In this chapter you examined four types of literature review that practitioners use to inform their practice decisions based on their practice questions: general, systematic, scoping, and meta-analysis reviews. You also reviewed information from our earlier course about locating research evidence and study reports—a discussion expanded to improve our understanding social work interventions. You learned about one model rating levels of evidence, as well as learning about the tension that exists between sensitivity and precision in search strategies. The material was further developed through an example of the process put into practice. The final

steps in the EBP process inform our next chapter: integrating the acquired information, implementing an intervention plan, and evaluating the intervention process and outcomes.

Module 2 Chapter 5:

Integrating, Implementing, and Evaluating

By this point in the process, a social worker will have devoted a great deal of energy and effort toward generating a pool of evidence-supported materials which can help inform practice decisions. In reviewing and evaluating these materials, it is important to keep in mind the lessons learned in our earlier course concerning the problem with pseudoscience. As a reminder, the risk of relying on pseudoscience increases with authors making assumptions not based on evidence, failing to acknowledge contradictory evidence, relying on “shaky” evidence (poorly designed studies or purely anecdotal evidence), obscuring facts with artificially constructed lingo, over-interpreting the implications of study results, and/or circumventing the peer review process (Lilienfeld, Lynn, & Lohr, 2015; Thyer & Pignotti, 2015). At this point in the evidence-based practice process, social workers are challenged with the need to integrate what they learned from the literature to make evidence-informed practice decisions. Then they will implement the intervention decision and evaluate the process and outcomes.

In this chapter you:

- learn practices for reviewing evidence presented in the empirical literature,
- learn practices for critically appraising evidence presented in the literature,
- are introduced to steps of integrating evidence, monitoring, and evaluating in practice.

Reviewing and Critiquing the Located Evidence

Drawing from what you learned in our earlier course Module 2 Chapter 4, the same principles about how to review an empirical article or report apply. Practicing social work in the “Information Age” has both advantages and disadvantages. On one hand, we have access to vast amounts of information (only some of which is evidence), much of it being available at low cost and at high rates of speed. Unfortunately:

“the fabulous Age of Information we’re living in doesn’t guarantee that we can make the most informed choices...the information may be out there, but it still takes work to find it, and think about it” (Ropeik, n.d.).

This section is about appraising the located evidence to improve practice decisions. We begin with a brief reminder about how to review empirical articles (from our prior course) and expand on these topics for applying evidence to informing practice decisions.

Review titles and keywords

The title of an empirical article should help a reviewer determine if an article’s relevance to addressing the intervention question at hand. The title may or may not refer to the specific intervention, social work problem or phenomenon, population, or outcomes studied. This initial review is not going to be conclusive regarding an article’s relevance, but it can help weed out some irrelevant pieces (improve precision).

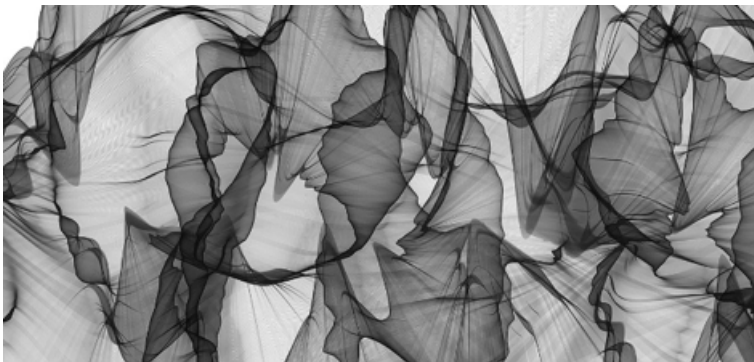
In addition to the title, many published works also have a list of 3 to 5 **keywords** that the authors selected to help individuals search for their work electronically. Keywords might or might not appear

in the title but can help inform a reviewer about the elements that the authors felt were most relevant to describe the work.



Review the abstract.

The abstract provides a reviewer with a summary of the study: aims, approach/design, methodology (participants, measures, and procedures), main results/findings, and conclusions/implications of the findings. Ideally, the abstract provides enough information for you to determine whether it is sufficiently relevant to pursue the full article. The abstract alone does not provide sufficient information for a reviewer to evaluate the evidence—that evaluation requires reviewing the full article.



Review the article.

If an empirical report survives the screening applied in review of the title, keywords, and abstract, it is time to acquire the full report and review its contents. In reviewing articles for relevance to understanding interventions, this is a summary of what to look for in the separate sections of an article.

Introduction: The introduction should inform you about the background and significance of what was studied, the state of knowledge and gaps in what is known, and the rationale for engaging in the study. You should come away with an understanding of the research questions, and if there are hypotheses, what these might be. After reviewing the introduction, you should have a better idea of whether the remainder of the article is relevant for your purposes.

Methodology: The methodology should describe the research approach adopted by the investigators—this should follow logically from the research questions presented in the introduction. If the study is qualitative, you should know what tradition was followed. If the study is quantitative, you should know what design was applied. If the study is mixed-methods, you should understand the approach adopted. Then, you should be able to develop a clear understanding of what was done at each step of the study—how participants were recruited and, if random assignment to conditions was involved, how this was accomplished; what variables were studied and how each was measured; and, the data collection procedures employed. Note that intervention or evaluation research reports need to provide sufficient details about the intervention that you can evaluate its characteristics, replicate its delivery (if you so choose), and understand how fidelity/integrity was addressed in the study.

Results/Findings: This section should explain how data were analyzed and what findings resulted from the analyses.

Discussion/Conclusions/Implications: This is the place where authors are expected to tie together the results with the

introduction, showing what the study contributed to knowledge and what answers emerged to the research questions. Authors should address any limitations of the study and show implications of the study results for practice and/or future research on the topic.

References: An article's reference list contributes to important things to the analysis. First, it allows a reader to determine whether the literature behind the study was well-covered and up-to-date. Second, it offers potentially relevant articles to seek in conducting one's own review of the literature.

Analyzing what occurred—Critiquing study methods.

Once the review of contents is completed, it is time to critically analyze what was presented in an article. This is where the reviewer makes decisions about the strength of the evidence presented based on the research methods applied by the investigators. In critiquing the methods, here are some points to consider.

Participants. You should consider the appropriateness and adequacy of the study “sample” in terms of the research approach, study design, and strength of evidence arguments. Authors should present descriptive information about the “sample” in terms of numbers and proportions reflecting categorical variables (like gender or race/ethnicity) and distribution on scale/continuous variables (like age).

The participant response rate (in quantitative studies) might also be calculated—this is the number of participants enrolled in the study divided by the number of persons eligible to be enrolled (the “pool”), multiplied by 100%. Very low response rates make a study vulnerable to selection bias—the few persons who elected to participate might not represent the population. In a qualitative study, generalizability is not a goal, but information about the participants should provide an indication of how robust the results

might be. Finally, regardless of study approach, authors typically make evident that the study was reviewed by an Institutional Review Board for the inclusion of human participants.

Not only does information about a study's participants help you evaluate the strength of the evidence, it helps you consider the relevance of the "sample" to the social work practice problem/decision you are facing. Consider, for example:

- the nature of the pool or population represented
- inclusion and exclusion criteria—what they were and their implications
- adequacy of the "sample" to represent the "target" population (numbers and representativeness) as a generalizability/external validity issue (quantitative studies)
- diversity in the "sample" and inclusiveness (qualitative studies)
- whose presence might have been excluded (intentionally or unintentionally)
- attrition/drop out from the study (if longitudinal) or from the intervention
- relevance of the "sample" to the clients for whom you are seeking intervention information
- potential ethical concerns should you wish to replicate the study or intervention.



Intervention details. Intervention and evaluation reports differ from other forms of empirical literature in that they need to describe key aspects/elements of the intervention being studied. These features (adapted and expanded from a list presented by Grinnell & Unrau, 2014) need to be appraised in the process of reviewing such an article:

- intervention aim—what were interventionists attempting to change, what were they attempting to achieve, what were the measurable objectives of the intervention? Were these elements appropriate to the logic model and/or theory underlying the intervention?
- intervention context—where, when, and under what conditions was the intervention delivered? How do these factors likely influence the intervention implementation and outcomes? Are these factors replicable?
- change agent—who delivered the intervention (change agent) and what were the characteristics of the change agent(s) involved? How do these factors likely influence the intervention's implementation and outcomes? Are these factors replicable?
- intervention elements—what were the key elements of the intervention, how do these relate to theory, logic model, empirical literature?
- intervention fidelity—to what extent did intervention implementation actually reflect the intervention protocol? How did the investigators assess fidelity (or intervention integrity)?
- inclusiveness—to what extent is the intervention culturally and otherwise inclusive, sensitive, appropriate? How did investigators ensure or assess inclusiveness (cultural competence) of the intervention and its delivery?

Measurement and data collection. In assessing strength of evidence, it is important to consider how data were collected and

how data collection frames the evidence. A great deal depends on the study approach—qualitative, quantitative, or mixed methods. Qualitative data collection procedures should be clearly identified by the investigators, including a description of what was asked of participants or what was being observed, and how the data were handled and coded. Practices related to inter-observer or inter-coder reliability should be reported, as well. Your job is to assess whether the variables were measured in a reliable, valid, and unbiased manner—particularly the outcome variables in the case of intervention or evaluation research.

Quantitative approaches require clear descriptions of the variables and how each was measured. You learned about measurement principles in our earlier course (Module 3, Chapter 5)—validity and reliability, in particular. These **psychometric properties** of measures used help determine the quality and strength of the data collected. Authors typically report this information for quantitative data collection tools previously published in the research literature; they also may summarize literature concerning how their measures were known to perform with the specific type of study participants involved in the study—for example, different ages, diagnoses, races/ethnicities, or other characteristics. On the other hand, investigators may create or modify existing tools and instruments for their own study, and psychometric information may not be available. This is an important consideration in your analysis of a study’s methodology—it does not mean that the study is not valid, just that its strength may be unknown.

A note concerning administrative and secondary data analysis is warranted here: it is important to evaluate, for yourself, how adequately you believe the variables of interest were indicated by the data used. For example, primary data in a study of who participated in prisoner visitation (parents, siblings, partners/spouses, children, and other family members) lacked consistency regarding the variable for “parent.” Administrative records included various terms, such as parent, mother, father, mother-in-law, step-

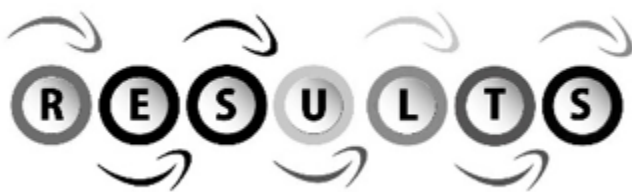
father, foster mother, and others. The investigators made decisions concerning how to manage inconsistencies in data recording and readers need to decide for themselves if they agree with the decisions made (Begun, Hodge, & Early, 2017).



Study procedures. Sometimes study procedures are described in the participants and measures sub-sections, and sometimes there is a separate sub-section where they are discussed. Study procedures content describes activities in which the investigators and study participants engaged during the study. In a quantitative experimental study, the methods utilized to assign study participants to different experimental conditions might be described here (i.e., the randomization approach used). Additionally, procedures used in handling data are usually described. In a quantitative study, investigators may report how they scored certain measures and what evidence from the literature informs their scoring approach. Regardless of the study's research approach or whether procedures are described in a separate sub-section, you should come away with a detailed understanding of how the study was executed. As a result, you should be sufficiently informed about the study's execution to be able to critically analyze the strength of the evidence developed from the methods applied.

Analyzing what was found—Critiquing results.

The structure and format of the results section varies markedly for different research approaches (qualitative, quantitative, and mixed methods). Regardless, critical review of how results were determined from the data is an important step. In the prior “article review” step you would have noted the actual results. In this step you are assessing the appropriateness and adequacy of the data analysis procedures and report of findings. To some extent, you can rely on the peer reviewers to have addressed these issues. This is a major advantage of limiting your search to peer reviewed manuscripts. However, as your familiarity with qualitative and quantitative data analysis matures, you will be able to engage in this assessment more fully. At the very least, you can assess how the reported results relate to the original research aims, questions, and/or hypotheses. You can also assess descriptive analyses for what they tell you about the data and how those results relate to your practice questions. You can also become informed about any problems the investigators encountered (and reported) regarding data analysis, how these problems were addressed, and how they might influence the strength of evidence and conclusions drawn from the data.



Analyzing what was concluded—Critiquing the

discussion.

The discussion section of an article (review or empirical) presents the authors' interpretation of what was found. In your critique of the manuscript you need to assess their interpretation—determining the extent to which you agree with their interpretation and how the study fits with the other pieces of evidence you have assembled and reviewed, how well it relates to the previous existing literature—did the results confirm or contradict the literature, or were the results ambiguous so that no strong conclusions could be drawn? Here are several points to consider in your analysis:

- the extent to which the conclusions are appropriate based on the study approach/design, participants, measures, procedures, data obtained, and data analyses performed—recommendations need to be supported by the evidence;
- the extent to which alternative explanations (competing hypotheses) fit the evidence, rather than or in addition to the interpretation offered by the authors—assessing what else could explain the observed outcome and whether the study design was strong enough to conclusively determine that the observed intervention outcomes actually were due to the intervention and not due to other factors;
- conclusions are not over-reaching the evidence—the conclusions are based on the strength of the evidence developed;
- the extent to which you believe the authors identified relevant methodological, analytic, or data quality limitations in the study, and how you believe the limitations affect the study's strength and relevance;
- implications of the study results for practice and/or future research that you believe the authors might have missed reporting.

Critiquing other relevant pieces.

As previously noted, you might review an article's reference list to evaluate its adequacy, appropriateness, and strength. One feature of significance is the extent to which the references are up-to-date. This does not mean that older references are not important; it simply means paying attention to whether new evidence is integrated into the manuscript.

Another piece of evidence to seek concerns **errata** or other corrections to the article that may appear in published literature subsequent to the article first appearing. The word *errata* (Latin) refers to errors that appear in print, and journals sometimes publish corrections. Often the corrections are minor—perhaps a number was incorrectly reported. Sometimes, however, the errors discovered have major implications for interpreting the evidence. And, unfortunately, journals are sometimes faced with the necessity to retract an article because of research integrity/misconduct concerns. A 2012 review of over 2,000 retracted research articles in biomedical and life-sciences research indexed in PubMed reported that over 67% were retracted because of research misconduct (fraud), almost 10% because of plagiarism, 21% because of significant errors, and about 14% because they had previously been published elsewhere (Fang, Steen & Casadevall, 2012). The journal published a correction in 2013—a table depicting the most frequently cited retracted articles contained errors. The corrected article (Fang, Steen, & Casadevall, 2013) presented a corrected table.



Proc Natl Acad Sci U S A. 2013 Jan 15; 110(3): 1137.

PMCID: PMC3549137

Published online 2012 Dec 17. doi: [10.1073/pnas.1220549110](https://doi.org/10.1073/pnas.1220549110)

Correction for Fang et al., Misconduct accounts for the majority of retracted scientific publications

[Copyright and License information](#) • [Disclaimer](#)

This corrects the article "[Misconduct accounts for the majority of retracted scientific publications](#)" in volume 109 on page 17028.

MEDICAL SCIENCES Correction for "Misconduct accounts for the majority of retracted scientific publications," by Ferrie C. Fang, R. Grant Steen, and Arturo Casadevall, which appeared in issue 42, October 16, 2012, of *Proc Natl Acad Sci USA* (109:17028–17033; first published October 1, 2012; 10.1073/pnas.1212247109).

The authors note that [Table 3](#) appeared incorrectly. The corrected table appears below.

Assessing Appropriateness

As you complete this phase of the EBP process, you are taking into consideration all of the evidence that you were able to locate, assemble, review, and critique to make an informed practice decision. One dimension of analysis for social work professionals to consider is the degree to which a study is appropriate to include in the decision-making process—this is over-and-above the critique of its quality and strength of evidence. A study could be very strong on these dimensions but not have relevance to the practice question at hand. For example, there may exist a great deal of evidence concerning interventions to prevent unplanned pregnancies among older adolescents and young adults, but this evidence might not be relevant to preventing pregnancy among younger adolescents (aged 11–15 years). Thus, it is important for the social work practitioner to consider how well a study's participants represented the clients for whom the studied intervention is being considered. Bronson and Davis (2012) explain:

“In other words, social workers have an obligation to know which interventions or programs are supported by rigorous research and to share that information with clients. However, research knowledge provides only the starting place for selecting the intervention of choice. The practitioner must also consider the similarities between the subjects of the research and the client seeking services, the acceptability and appropriateness of the intervention for the client, and the client’s ability to participate in the intervention” (p. 5).

Along these same lines, it is important to consider the outcomes specified in the reviewed studies and how well these outcomes align with the intervention goals prompting the search for evidence. For example, a systematic review of medication assisted treatment (MAT) identified and reviewed 40 studies—observing that they reported on a diverse set of outcome measures. Results were mixed in terms of recommending MAT over other treatments and the outcomes differed on the basis of which medications were administered in the treatment protocol (Maglione et al., 2018). In other words, some of the reviewed studies would be less relevant than others, depending on the practice concerns being addressed by the practitioner and client.

Another dimension of appropriateness that needs to be taken into consideration is feasibility. Feasibility is about the practical realities involved—the likelihood that the intervention can be implemented with fidelity in the context where the practice question arose. For example, a form of mental health or addiction treatment may not be feasible to implement with persons being released from incarceration or with persons at risk of homelessness with a high degree of fidelity to the intervention protocol because of the uncertainties and fluctuations in the life structures and contexts these individuals might experience. Or, it may not be feasible to implement these intervention protocols if there are too few trained professionals to deliver the intervention in a particular community. Political and cultural context are also important considerations,

depending on characteristics of a particular intervention and how those features intersect with the target population or community—this issue relates to the acceptability aspect of an intervention’s appropriateness.

Working Example: Addressing High School Dropout

Considering the 12 interventions that the team identified in the previous step, they felt that 12 intervention options were too many. They re-engaged with the COPES framework, deciding to “favor the approach that best allowed them to change their focus and emphasis without having to totally revamp all the organizational structure or staffing” (Kelly & Franklin, 2011, p. 151). In other words, finding the most effective option that could most practically be adopted. Based on this further analysis, the team decided on one intervention as being the most feasible for the district to implement: Quantum Opportunities Program (QOP). They provided this information to the school board in their final report.



Implementing, Monitoring, and Evaluating

Let's review the six steps of the evidence-based practice (EBP) process specified in the introduction to this module:

Step 1: specifying an answerable practice question

Step 2: identifying the best evidence for answering that question

Step 3: Critically appraising the evidence and its applicability to the question/problem

Step 4: Integrating the results from the critical appraisal with practice expertise and the client's or client system's unique circumstances

Step 5: Taking appropriate actions based on this critical appraisal of evidence

Step 6: Monitoring and evaluating outcomes of (a) the practice decision/intervention and (b) effectiveness and efficiency of the EBP process (steps 1-4).

Subsequent chapters prepared you to develop answerable practice questions, and to identify and critically appraise the relevant evidence. The next step is for the practitioner and client(s) together to determine the most appropriate evidence-based intervention for their situation—what is appropriate to THIS practice situation. The process recognizes that different tools (intervention options) often exist for promoting change and that the same tool is not always the best tool for different individuals or circumstances. Consider the saying (often attributed to Maslow, 1966), “if all you have is a hammer, everything looks like a nail” and how inappropriate it might be to apply that hammer to open a piggy bank. Thus, client values, preferences, and attributes, along with practitioner experience and circumstances, are part of the social work EBP process—what works for some might not be the best choice for others under all circumstances.



For example, an intervention designed to address homelessness might not be the best choice for individuals who also experience co-occurring mental health or substance use problems. One-third of persons who are homeless experience a serious mental disorder (e.g., schizophrenia, bipolar disorder) and at least half of persons who are homeless have substance misuse problems or a substance use disorder (Padgett, Stanhope, & Henwood, 2011). Whether these mental health and substance-related problems came first or they came after the homeless situation, they tend to worsen and complicate (exacerbate) the problem of homelessness. “Usual” intervention strategies fall short in meeting the needs of this distinct population of individuals experiencing or at risk of homelessness (Padgett, Stanhope, & Henwood, 2011). Thus, a search for evidence about alternative best practices begins with the stated research question including these features if they are part of the encountered practice concern.

Once the prime-choice intervention has been selected, the next step is to implement that intervention with the greatest possible degree of fidelity and integrity. Implementation may require additional training or certification by the professional delivering

the intervention. Instead, it may be necessary to collaborate with or make a referral to another service provider to ensure proper implementation.

Critical to this implementation process is the application of monitoring and evaluation processes, as well. This means that the implementation process will be monitored for fidelity and outcomes will be evaluated along the way. Together with the client(s), measurable goals and objectives need to be identified. The evidence collected from monitoring and evaluation will form part of a feedback loop to inform modifications or changes in the intervention strategy. It may even become necessary to switch from the initially selected intervention plan (Plan A) to an alternative that is also supported by the search for evidence (Plan B or Plan C). These six steps apply to social work practice at any level—individuals, families, groups, communities, institutions, and policy. In future modules you will learn more about the evaluation process and what to consider when analyzing a study's methods and results.





Take a moment to complete the following activity.

1. Create a checklist of information you want to include in reviewing an empirical article about an intervention study.
2. Create a second checklist of points to analyze in your critique of the reviewed article, including points related to appropriateness.
3. Keep these tools handy as you begin to locate, review, and critically analyze intervention literature.

Chapter Summary

In this chapter you learned how the final steps in the evidence-based practice (EBP) process are fulfilled. You learned to review intervention and evaluation literature for strength of evidence and how empirical literature informs practice decisions. You learned steps and issues in critically analyzing the empirical literature, as well. Finally, you were introduced to issues related to implementing an evidence-informed practice decision and the importance of monitoring and evaluating the intervention implemented.

Module 2 Conclusions

This module reviewed a number of topics and issues from our prior course related to research questions, the role of theory, locating empirical evidence, and analyzing the evidence. These topics were applied to understanding social work interventions, the theme of our current course. The new content in this module explored the COPES and PICO frameworks for developing practice questions, the role of theory and logic models in planning interventions, locating empirical evidence to inform practice (literature reviews, systematic reviews, scoping reviews, meta-analysis, and individual articles), and how to critically evaluate the evidence you have located. In addition, the final steps of the EBP process were introduced: selecting the intervention strategy, implementing the intervention, and monitoring/evaluating the intervention. This leads to our next modules: how to evaluate interventions.

Module 2 Key Terms and Definitions

COPES framework: a system for helping practitioners frame clear practice questions to guide their search for evidence.

distal outcomes: effects or impact of an intervention that appear in the long-term or long-range time frame.

effect size: a quantitative (statistical) indicator of the size or magnitude of an observed difference, helping interpret how meaningful a difference might be.

errata: means of correcting errors that appeared in print (published errors).

inputs: the resources committed to an intervention or change effort.

keywords: search terms associated with a manuscript to assist those seeking to locate information on the topics covered.

logic model: a “map” of an intervention plan that includes the various inputs and outputs anticipated.

mechanism of change: the underlying processes leading to a change outcome, often identified as mediators of influence.

mediators: variables or factors that influence or determine the path between an “input” and “outcome” (see mechanism of change).

meta-analysis: form of systematic statistical analysis allowing data from multiple studies to be combined into a single analysis to assess an intervention’s impact common to each of the included studies.

outputs: the results, effects, or impacts of an intervention effort.

PICO framework: a system for helping practitioners frame clear practice questions to guide their search for evidence, common in medical professions.

precision (of search): a concept from information science related to the proportion of items retrieved from a search strategy (query).

proximal outcomes: effects or impact of an intervention that appear in the short-term or immediate time frame.

psychometric properties: results of analysis concerning the validity and reliability characteristics of a measurement instrument, indicative of how valid and reliable the instrument is for measuring the intended constructs; psychometric properties may vary with different populations being measured.

scoping review: a form of systematic review that applies less stringent (precise) criteria, typically applied when an area or topic is emerging and insufficiently developed for a formal systematic or meta-analysis review to be conducted.

sensitivity (of search): a concept from information science related to the inclusiveness or breadth in range of items retrieved from a search strategy (query).

systematic review: a summary of evidence presented in literature that is conducted systematically, applying unbiased criteria for locating, reviewing, and critiquing the literature and integrating the analysis results.

Module 2 References

- Ackley, B.J., Swan, B.A., Ladwig, G.B., & Tucker, S. (2008). Evidence-based nursing care guidelines: Medical-surgical interventions. St. Louis, MO: Mosby/Elsevier.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research*, 8(1), 19-32.
- Begun, A.L., Hodge, A.I., & Early, T.J. (2017). A family systems perspective in prisoner reentry. In S. Stojkovic, (Ed.), *Prisoner reentry: Critical issues and policy directions*. NY: Palgrave Macmillan.
- Bronson, D.E. & Davis, T. S. (2012). *Finding and evaluating evidence: Systematic reviews and evidence-based practice*. NY: Oxford University Press, Pocket Guides to Social Work Research Methods.
- Corcoran, J., & Secret, M. (2013). *Social work research skills workbook: A step-by-step guide to conducting agency-based research*. NY: Oxford University Press.
- Fang, F.C., Steen, R.G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028-17033.
- Fang, F.C., Steen, R.G., & Casadevall, A. (2013). Correction for Fang et al., Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 110(3), 1137.
- Fraser, J.W., Richman, J.M., Galinsky, M.J., & Day, S.H. (2009). *Intervention research: Developing social programs*. NY: Oxford University Press, Pocket Guides to Social Work Research Methods.
- Gutierrez, L. (2011). Foreward. In M. Roberts-DeGennaro & S. J. Fogel, (Eds.), *Using evidence to inform practice for community and organizational change*, (pp. xv-xxv). Chicago: Lyceum.

- Harris, M. (2010). *Evaluating public and community health programs*. San Francisco: Jossey-Bass.
- Heltzer, K. (n.d.) Finding research evidence, student activity 2. Retrieved from https://www.d.umn.edu/~kheltzer/sw5095sp08/handouts/develop_research_to_practice_q.pdf
- Higgins, J.P.T., & Green, S. (editors). (2011). *Cochrane handbook for systematic reviews of interventions*, version 5.1.0[updated March 2011]. The Cochrane Collaboration. Retrieved from <https://handbook-5-1.cochrane.org/>.
- Kapp, S.A., & Anderson, G.R. (2010). *Agency-based program evaluation: Lessons from practice*. Los Angeles: Sage.
- Kelly, M.S., & Franklin, C. (2011). Bringing the village together to prevent high school dropout: Using the process of evidence-based practice. In M. Roberts-DeGennaro & S. J. Fogel, (Eds.), *Using evidence to inform practice for community and organizational change*, (pp. 140-156). Chicago: Lyceum.
- Kennedy, S.C., Kim, J.S., Tripodi, S.J., Brown, S.M., & Gowdy, G. (2016). Does Parent-Child Interaction Therapy Reduce Future Physical Abuse? A Meta-Analysis. *Research on Social Work Practice*, 26(2), 147-156.
- Lilienfeld, S.O., Lynn, S.J., & Lohr, J.M. (2015). *Science and pseudoscience in clinical psychology*. NY: Guilford Press.
- Littell, J.H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. NY: Oxford University Press, Pocket Guides to Social Work Research Methods.
- Maglione, M. A., Raaen, L., Chen, C., Axhar, G.S., Shahidinia, N...Hempel, S. (2018). Effects of medication assisted treatment (MAT) for opioid use disorder on functional outcomes: A systematic review. *Journal of Substance Abuse Treatment*, 89, 28-51.
- Maslow, A.H. (1966). *The psychology of science: A reconnaissance*. NY: Harper & Row.
- National Institute on Alcohol Abuse and Alcoholism (NIAAA). (2018). News release: Lack of sleep may be linked to risk factor for Alzheimer's disease. Preliminary NIH study shows increased levels

- of beta-amyloid. (April 13). <https://www.nih.gov/news-events/lack-sleep-may-be-linked-risk-factor-alzheimers-disease>
- Padgett, D.K., Stanhope, V., & Henwood, B.F. (2011). Housing-First services for homeless adults with co-occurring disorders. In M. Roberts-DeGennaro & S. J. Fogel, (Eds.), *Using evidence to inform practice for community and organizational change*, (pp. 81-98). Chicago: Lyceum.
- Pham, M.T., Rajić, A., Greig, J.D., Sargeant, J.M., Papdopoulos, A., & McEwen, S.A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, 5, 371-385.
- Ropeik, D. (n.d.) How the information age makes it harder to make informed choices about risk (or anything else). Retrieved from <http://bigthink.com/risk-reason-and-reality/how-the-information-age-makes-it-harder-to-make-informed-choices-about-risk-or-anything>
- Sathe, N., Andrews, J.C., McPheeters, M.L., & Warren, Z.E. (2017). Nutritional and dietary interventions for autism spectrum disorder: A systematic review. *Pediatrics*, 139(6), e20170346.
- Shokri-Kojori, E., Wang, G.J., Wiers, C.E., Demiral, S.B., Guo, M...Volkow, N. (2018). β -amyloid accumulation in the human brain after one night of sleep deprivation. *Proceedings of the national Academy of Sciences*. Retrieved from <http://www.pnas.org/content/early/2018/03/29/1721694115.full>.
- Thyer, B.A., & Pignotti, M. (2015). *Science and pseudoscience in social work practice*. NY: Springer.
- Warren, Z., McPheeters, M.L., Sathe, N., Foss-Feig, J.H., Glasser, A., Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, 127(5), e1303-1311.

MODULE 3

Module 3 Introduction

This module introduces research designs and methods used in addressing research questions concerning intervention effectiveness and in evaluating practice at multiple levels of social work intervention. We expand on our prior course content about study approaches as we explore quantitative, qualitative, mixed, and community-engaged research approaches to understanding interventions. Related to these research approaches, we compare study design options, measurement and data collection strategies, and participant recruitment issues appropriate for answering social work intervention research and evaluation questions. Regarding study design, we address strengths and limitations of specific study designs (including experimental and quasi-experimental designs). In exploring intervention and evaluation study participant issues, we examine recruitment and retention strategies, sample size issues, and techniques for random assignment to intervention conditions. Ethical concerns related to intervention and evaluation study participants were covered in Module 1 of the current course and directly relate to these discussions. Specific to measurement and data collection strategies, we examine issues such as using clinical assessment tools as research measures, identifying measurement tools and their psychometric properties (reliability, validity), and cultural competence related to measurement approaches. Intervention fidelity is an integral aspect of this module.

READING OBJECTIVES

After engaging with these reading materials and learning resources, you should be able to:

- Describe and critique different approaches and designs

appropriate for addressing intervention and evaluation research questions, including addressing issues of intervention fidelity/integrity;

- Recognize measurement issues and strategies for ensuring a strong measurement and data collection plan related to the study design and variables;
- Explain issues in participant recruitment (and retention), including sampling strategies, sample size, and random assignment in intervention research;
- Identify issues of cultural appropriateness and relevance related to study approach, participant recruitment, and measurement plans in intervention research;
- Define key terms related to the design and implementation of social work intervention and evaluation research.

Module 3 Chapter 1: Overview of Intervention/ Evaluation Research Approaches

In our prior course, you learned how the nature of an investigator's research question dictates the type of study approach and design that might be applied to achieve the study aims. Intervention research typically asks questions related to the outcomes of an intervention effort or approach. However, questions also arise concerning implementation of interventions, separate from understanding their outcomes. Practical, philosophical, and scientific factors contribute to investigators' intervention study approach and design decisions.

In this chapter you learn:

- how content from our earlier course about study approaches and designs relate to intervention research;
- additional approaches to intervention research (participatory research; formative, process, outcome, and cost-related evaluation research)
- intervention research strategies for addressing intervention fidelity and internal validity concerns.

Review and Expansion: Study Approaches

In our earlier course you became familiar with the ways that research questions lead to research approach and methods.

Intervention and evaluation research are not different: the question dictates the approach. In the earlier course, you also became familiar with the philosophical, conceptual and practical aspects of different approaches to social work research: qualitative, quantitative, and mixed methods. These methods are used in research for evaluating practice and understanding interventions, as well. The primary emphasis in this module revolves around quantitative research designs for practice evaluation and understanding interventions. However, taking a few moments to examine qualitative and mixed methods in these applications is worthwhile. Additionally, we introduce forms of participatory research—something we did not discuss regarding efforts to understand social work problems and diverse populations. Participatory research is an approach rich in social work tradition.

Qualitative methods in intervention & evaluation research.

The research questions asked by social workers about interventions often lend themselves to qualitative study approaches. Here are 5 examples.

- Early in the process of developing an intervention, social workers might simply wish to create a rich description of the intervention, the contexts in which it is being delivered, or the clients' experience with the intervention. This type of information is going to be critically important in developing a standardized protocol which others can use in delivering the intervention, too. Remember that qualitative methods are ideally suited for answering *exploratory* and *descriptive* questions.
- Qualitative methods are well-suited to exploring different experiences related to diversity—the results retain individuality

arising from heterogeneity rather than homogenizing across individuals to achieve a “normative” picture.

- Qualitative methods are often used to assess the degree to which the delivery of an intervention adheres to the procedures and protocol originally designed and empirically tested. This is known as an **intervention fidelity** issue (see the section below on the topic of process evaluation).
- Intervention outcomes are sometimes evaluated using qualitative approaches. For example, investigators wanted to learn from adult day service participants what they viewed as the impact of the program on their own lives (Dabelko-Schoeny & King, 2010). The value of such information is not limited to evaluating this one program. Evaluators are informed about important evaluation variables to consider in their own efforts to study interventions delivered to older adults—variables beyond the typical administrative criteria of concern. The study participants identified social connections, empowering relationships with staff, and enjoyment of activities as important evaluation criteria.
- Assessing the need for intervention (needs assessment) is often performed with qualitative approaches, especially focus groups, open-ended surveys, and GIS mapping.
- Qualitative approaches are an integral aspect of mixed-methods approaches.

Qualitative approaches often involve in-depth data from relatively few individuals, seeking to understand their individual experiences with an intervention. As such, these study approaches are relatively sensitive to nuanced individual differences—differences in experience that might be attributed to cultural, clinical, or other demographic diversity. This is true, however, only to the extent that diversity is represented among study participants, and individuals cannot be presumed to represent groups or populations.



Quantitative methods in intervention & evaluation research.

Many intervention and evaluation research questions are quantitative in nature, leading investigators to adopt quantitative approaches or to integrate quantitative approaches in mixed methods research. In these instances, “how much” or “how many” questions are being asked, questions such as:

- how much change was associated with intervention;
- how many individuals experienced change/achieved change goals;
- how much change was achieved in relation to the resources applied;
- what trends in numbers were observed.

Many study designs detailed in Chapter 2 reflect the philosophical roots of quantitative research, particularly those designed to zero in on causal inferences about intervention—the explanatory research

designs. Quantitative approaches are also used in descriptive and exploratory intervention and evaluation studies. By nature, quantitative studies tend to aggregate data provided by individuals, and in this way are very different from qualitative studies. Quantitative studies seek to describe what happens “on average” rather than describing individual experiences with the intervention—you learned about central tendency and variation in our earlier course (Module 4). Differences in experience related to demographic, cultural, or clinical diversity might be quantitatively assessed by comparing how the intervention was experienced by different groups (e.g., those who differ on certain demographic or clinical variables). However, data for the groups are treated in the aggregate (across individuals) with quantitative approaches.

Mixed methods in intervention & evaluation research.

Qualitative and quantitative approaches are very helpful in evaluation and intervention research as part of a mixed-methods strategy for investigating the research questions. In addition to the examples previously discussed, integrating qualitative and quantitative approaches in intervention and evaluation research is often done as means of enriching the results derived from one or the other approach. Here are 3 scenarios to consider.

- Investigators wish to use a two-phase approach in studying or evaluating an intervention. First, they adopt a qualitative approach to inform the design of a quantitative study, then they implement the quantitative study as a second phase. The qualitative phase might help inform any aspect of the quantitative study design, including participant recruitment and retention, measurement and data collection, and presenting study results.

- Investigators use a two-phase approach in studying or evaluating an intervention. First, they implement a quantitative study. Then, they use a qualitative approach to explore the appropriateness and adequacy of how they interpret their quantitative study results.
- Investigators combine qualitative and quantitative approaches in a single intervention or evaluation study, allowing them to answer different kinds of questions about the intervention.

For example, a team of investigators applied a mixed methods approach in evaluating outcomes of an intensive experiential learning experience designed to prepare BSW and MSW students to engage effectively in clinical supervision (Fisher, Simmons, & Allen, 2016). BSW students provided quantitative data in response to an online survey, and MSW students provided qualitative self-assessment data. The quantitative data answered a research question about how students felt about supervision, whereas the qualitative data were analyzed for demonstrated development in critical thinking about clinical issues. The investigators concluded that their experiential learning intervention contributed to the outcomes of forming stronger supervisory alliance, BSW student satisfaction with their supervisor, and MSW students thinking about supervision as being more than an administrative task.



Cross-Sectional & Longitudinal Study Designs.

You are familiar with the distinction between cross-sectional and longitudinal study designs from our earlier course. In that course, we looked at these designs in terms of understanding diverse populations, social work problems, and social phenomena. Here we address how the distinction relates to the conduct of research to understand social work interventions.

- A **cross-sectional study** involves data collection at just one point in time. In a program evaluation, for example, the agency might look at some outcome variable at the point when participants complete an intervention or program. Or, perhaps an agency surveys all clients at a single point in time to assess their level of need for a potential new service the agency might offer. Because the data are collected from each person at only one point in time, these are both cross-sectional studies. In terms of intervention studies, one measurement point obviously needs to be after the intervention for investigators to draw inferences about the intervention. As you will see in the discussion of intervention study designs, there exist considerable limitations to using only one single measurement to evaluate an intervention (see post-only designs in Chapter 2).
- A **longitudinal study** involves data collection at two or more points in time. A great deal of intervention and evaluation research is conducted using longitudinal designs—answering questions about what changes might be associated with the intervention being delivered. For example, in program evaluation, an agency might compare how clients were functioning on certain variables at the time of discharge compared to their level of functioning at intake to the program. Because the same information is collected from each individual at two points in time (pre-intervention and post-

intervention), this is a longitudinal design.

- Distinguishing cross-section and longitudinal in studies of systems beyond the individual person can become confusing. When social workers intervene with individuals or families or small groups, that longitudinal study involves the same individuals or members at different points in time is evident—perhaps measuring individuals before, immediately after, and months after intervention (this is called **follow-up**). However, if an intervention is conducted in a community, a state, or across the nation, the data might not be collected from the same individual persons at each point in time—the unit of analysis is what matters here. For example, if the longitudinal study's unit of analysis is the 50 states, District of Columbia, and 5 inhabited territories of the United States, data are repeatedly collected at that level (states, DC, and territories), perhaps not from the same individual persons in each of those communities.



Formative, Process, and Outcome Evaluation

Practice and program evaluation are important aspects of social work practice. It would be nice if we could simply rely on our own

sense of what works and what does not. However, social workers are only human and, as we learned in our earlier course, human memory and decisions are vulnerable to bias. Sources of bias include recency, confirmation, and social desirability biases.

- *Recency bias* occurs when we place higher emphasis on what has just happened (recently) than on what might have happened in the more distant past. In other words, a social worker might make a casual practice evaluation based on one or two exceptionally good or exceptionally bad recent outcomes rather than a longer, larger history of outcomes and systematic evidence.
- *Confirmation bias* occurs when we focus on outcomes that reinforce what we believed, feared, or hoped would happen and de-emphasize alternative events or interpretations that might contradict those beliefs, fears, or hopes.
- Social desirability bias by practitioners occurs when practice decisions are influenced by a desire to be viewed favorably by others—that could be clients, colleagues, supervisors, or others. In other words, a practice decision might be based on “popular” rather than “best” practices, and casual evaluation of those practices might be skewed to create a favorable impression.

In all three of these forms of bias, the problem is not necessarily intentional, but does result in a lack of sufficient attention to evidence in monitoring one’s practices. For example, relying solely on qualitative comments volunteered by consumers (anecdotal evidence) is subject to a **selection bias**—individuals with strong opinions or a desire to support the social workers who helped them are more likely to volunteer than the general population of those served.

Thus, it is incumbent on social work professionals to engage in practice evaluation that is as free of bias as possible. The choice of systematic evaluation approach is dictated by the evaluation

research question being asked. According to the Centers for Disease Control and Prevention (CDC), there are four most common types of intervention or program evaluation: formative, process, outcome, and impact evaluation (<https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf>). Here, we consider these as three types, combining impact and outcome evaluation into a single category, and we consider an additional category, as well: cost evaluation.

Formative Evaluation.

Formative evaluation is emphasized during the early stages of developing or implementing a social work intervention, as well as following process or outcome evaluation as changes to a program or intervention strategy are considered. The aim of formative evaluation is to understand the context of an intervention, define the intervention, and evaluate feasibility of adopting a proposed intervention or change in the intervention (Trochim & Donnelly, 2007). For example, a needs assessment might be conducted to determine whether the intervention or program is needed, calculate how large the unmet need is, and/or specify where/for whom the unmet need exists. Needs assessment might also include conducting an inventory of services that exist to meet the identified need and where/why a gap exists (Engel & Schutt, 2013). Formative evaluation is used to help shape an intervention, program, or policy.



Process Evaluation.

Investigating how an intervention is delivered or a program operates is the purpose behind **process evaluation** (Engel & Schutt, 2013). The concept of intervention fidelity was previously introduced. Fidelity is a major point of process evaluation but is not the only point. We know that the greater the degree of fidelity in delivery of an intervention, the more applicable the previous evidence about that intervention becomes in reliably predicting intervention outcomes. As fidelity in the intervention's delivery drifts or wanes, previous evidence becomes less reliable and less useful in making practice decisions. Addressing this important issue is why many interventions with an evidence base supporting their adoption are **manualized**, providing detailed manuals for how to implement the intervention with fidelity and integrity. For example, the Parent-Child Interaction Therapy for Traumatized Children (PCIT-TC) treatment protocol is manualized and training certification is available for practitioners to learn the evidence-based skills involved (<https://pcit.ucdavis.edu/>). This strategy increases practitioners' adherence to the protocol.

Process evaluation, sometimes called implementation evaluation and sometimes referred to as program monitoring, helps investigators determine the extent to which fidelity has been

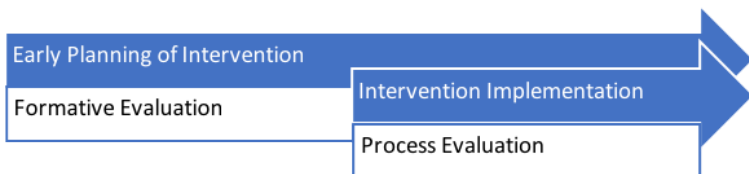
preserved. But, process evaluation serves other purposes, as well. For example, according to King, Morris and Fitz-Gibbon (1987), process evaluation helps:

- document details about the intervention that might help explain outcome evaluation results,
- keep programs accountable (delivering what they claim to deliver),
- inform planned modifications and changes to the intervention based on evidence.

Process evaluation also helps investigators determine where the facilitators and barriers to implementing an intervention might operate and can help interpret outcomes/results from the intervention, as well. Process evaluation efforts addresses the following:

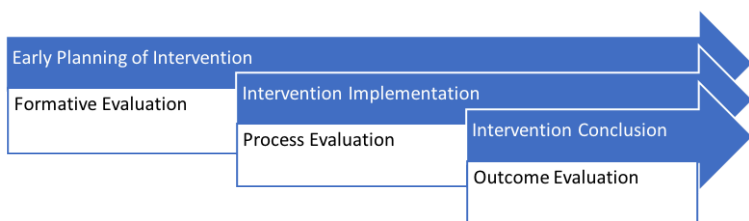
- Who delivered the intervention
- Who received the intervention
- What was (or was not) done during the intervention
- When intervention activities occurred
- Where intervention activities occurred
- How the intervention was delivered
- What facilitated implementation with fidelity/integrity
- What presented as barriers to implementation with fidelity/integrity

For these reasons, many authors consider process evaluation to be a type of formative evaluation.



Outcome and Impact Evaluation.

The aim of **outcome or impact evaluation** is to determine effects of the intervention. Many authors refer to this as a type of **summative evaluation**, distinguishing it from formative evaluation: its purpose is to understand the effects of an intervention once it has been delivered. The effects of interest usually include the extent to which intervention goals or objectives were achieved. An important factor to evaluate concerns positive and negative “side effects”—those unintended outcomes associated with the intervention. These might include unintended impact of the intervention participants or impacts on significant others, those delivering the intervention, the program/agency/institutions involved, and others. While impact evaluation, as described by the CDC, is about policy and funding decisions and longer-term changes, we can include it as a form of outcome evaluation since the questions answered are about achieving intervention objectives. Outcome evaluation is based on the elements presented in the logic model created at the outset of intervention planning.



Cost-Related Evaluation.

Social workers are frequently faced with efficiency questions related to the interventions we deliver—thus, **cost-related evaluation** is part of our professional accountability responsibilities. For example, once an agency has applied the evidence-based practice (EBP) process to select the best-fitting program options for addressing an identified practice concern, program planning is enhanced by information concerning which of the options is most cost-effective. Here are some types of questions addressed in cost-related evaluation.

cost analysis: How much does it cost to deliver/implement the intervention with fidelity and integrity? This type of analysis typically analyzes monetary costs, converting inputs into their financial impact (e.g., space resources would be converted into cost per square foot, staffing costs would include salary, training, and benefits costs, materials and technology costs might include depreciation).

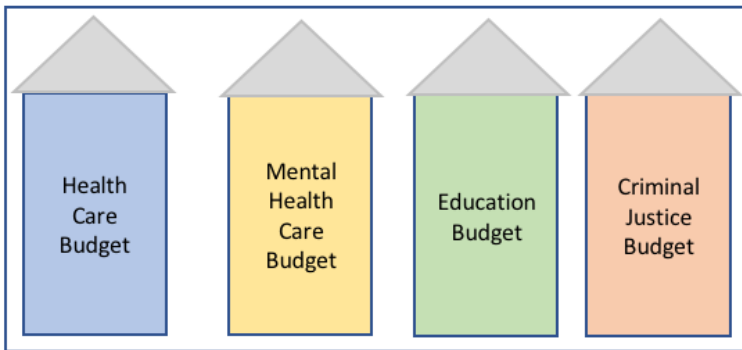
- **cost-benefit:** What are the inputs and outputs associated with the intervention? This type of analysis involves placing a monetary value on each element of input (resources) and each

of the outputs. For example, preventing incarceration would be converted to the dollars saved on jail/prison costs; and, perhaps, including the individuals' ability to keep their jobs and homes which could be lost with incarceration, as well as preventing family members needing public assistance and/or children being placed in foster care if their family member is incarcerated.

- *cost-effectiveness*: What is the ratio of cost units (numerator) to outcome units (denominator) associated with delivering an intervention. Outcomes are tied to the intervention goals rather than monetary units. For example, medical interventions are often analyzed in terms of **DALYs** (disability-adjusted life years)—units designed to indicate “disease burden,” calculated to represent the number of years lost to illness, disability, or premature death (morbidity and mortality). Outcomes might also be numbers of “cases,” such as deaths or hospitalizations related to suicide attempts, drug overdose events, students dropping out from high school, children reunited with their families (family reunification), reports of child maltreatment, persons un- or under-employed, and many more examples. Costs are typically presented as monetary units estimated from a costs analysis. (See <http://www.who.int/heli/economics/costeffanalysis/en/>).
- *cost-utility*: A comparison of cost-effectiveness for two or more intervention options, designed to help decision-makers make informed choices between the options.

Two of the greatest challenges with these kinds of evaluation are (1) ensuring that all relevant inputs and outputs are included in the analysis, and (2) realistically converting non-monetary costs and benefits into monetary units to standardize comparisons. An additional challenge has to do with budget structures: the gains might be realized in a different budget than where the costs are borne. For example, implementing a mental health or substance misuse treatment program in jails and prisons costs those facilities;

the benefits are realized in budgets outside those facilities—schools, workplaces, medical facilities, family services, and mental health programs in the community. Thus, it is challenging to make decisions based on these analyses when constituents are situated in different systems operating with “siloe” budgets where there is little or no sharing across systems.



An Additional Point.

An intervention or evaluation effort does not necessarily need to be limited to one type. As in the case of mixed-methods approaches, it is sometimes helpful to engage in multiple evaluation efforts with a single intervention or program. A team of investigators described how they used formative, process, and outcome evaluation all in the pursuit of understanding a single preventive public health intervention called VERB, designed to increase physical activity among youth (Berkowitz et al., 2008). Their formative evaluation efforts allowed the team to assess the intervention's appropriateness for the target audience and to test different messages. The process evaluation addressed fidelity of the intervention during implementation. And, the outcome evaluation led the team to draw conclusions concerning the intervention's

effects on the target audience. The various forms of evaluation utilized qualitative and quantitative approaches.

Participatory Research Approaches

One contrast previously noted between qualitative and quantitative research is the nature of the investigator's role. Every effort is made to minimize investigator influence on the data collection and analysis processes in quantitative research. Qualitative research, on the other hand, recognizes the investigator as an integral part of the research process. **Participatory research** fits into this latter category.

“Participant observation is a method in which natural social processes are studied as they happen (in the field, rather than in the laboratory) and left relatively undisturbed. It is a means of seeing the social world as the research subjects see it, in its totality, and of understanding subjects’ interpretations of that world” (Engel & Schutt, 2013, p. 276).

This quote describes **naturalistic observation** very well. The difference with **participatory observation** is that the investigator is embedded in the group, neighborhood, community, institution, or other entity under study. Participatory observation is one approach used by anthropologists to understand cultures from an embedded rather than outsider perspective. For example, this is how Jane Goodall learned about chimpanzee culture in Tanzania: she became accepted as part of the group she observed, allowing her to describe the members’ behaviors and social relationships, her own experiences as a member of the group, and the theories she derived from 55 years of this work. In social work, the participant approach may be used to answer the research questions of the type we explored in our earlier course: understanding diverse populations, social work problems, or social phenomena. The investigator might

be a natural member of the group, where the role as group member precedes the role as observer. This is where the term indigenous membership applies: naturally belonging to the group. (The term “indigenous people” describes the native, naturally occurring inhabitants of a place or region.) It is sometimes difficult to determine how the indigenous member’s observations and conclusions might be influenced by his or her position within the group—for example, the experience might be different for men and women, members of different ages, or leaders. Thus, the conclusions need to be confirmed by a diverse membership.

Participant observers are sometimes “adopted” members of the group, where the role of observer precedes their role as group member. It is somewhat more difficult to determine if evidence collected under these circumstances reflects a fully accurate description of the members’ experience unless the evidence and conclusions have been cross-checked by the group’s indigenous members. Turning back to our example with Jane Goodall, she was accepted into the chimpanzee troop in many ways, but not in others—she could not experience being a birth mother to members of the group, for example.

Sometimes investigators are more actively engaged in the life of the group being observed. As previously noted, participant observation is about the processes being left relatively undisturbed (Engel & Schutt, 2013, p. 276). However, participant observers might be more actively engaged in change efforts, documenting the change process from “inside” the group promoting change. These instances are called **participatory action research (PAR)**, where the investigator is an embedded member of the group, joining them in making a concerted effort to influence change. PAR involves three intersecting roles: participation in the group, engaging with the action process (planning and implementing interventions), and conducting research about the group’s action process (see Figure 2-1, adapted from Chevalier & Buckles, 2013, p. 10).

Figure 2-1. Venn diagram of participatory action research roles.



For example, Pyles (2015) described the experience of engaging in participatory action research with rural organizations and rural disaster survivors in Haiti following the January 12, 2010 earthquake. The PAR aimed to promote local organizations' capacity to engage in education and advocacy and to secure much-needed resources for their rural communities (Pyles, 2015, p. 630). According to the author, rural Haitian communities have a history of experience with exploitative research where outsiders conduct investigations without the input or participation of community members, and where little or no capacity-building action occurs based on study results and recommendations. Pyles also raised the point that, "there are multiple barriers impeding the participation of marginalized people" in community building efforts, making PAR approaches even more important for these groups (2015, p. 634).

The term **community-based participatory research (CBPR)** refers to collaborative partnerships between members of a community (e.g., a group, neighborhood, or organization) and researchers throughout the entire research process. CBPR partners (internal

and external members) all contribute their expertise to the process, throughout the process, and share in all steps of decision-making. Stakeholder members of the community (or organization) are involved as active, equal partners in the research process, co-learning by all members of the collaboration is emphasized, and it represents a strengths-focused approach (Harris, 2010; Holkup, Tripp-Reier, Salois, & Weinert, 2004). CBPR is relevant in our efforts to understand social work interventions since the process can result in interventions that are culturally appropriate, feasible, acceptable, and applicable for the community since they emerged from within that community. Furthermore, it is a community empowerment approach whereby self-determination plays a key role and the community is left with new skills for self-study, evaluation, and understanding the change process (Harris, 2010). These characteristics of CBPR help define the approach.

- (a) recognizing the community as a unit of identity,*
- (b) building on the strengths and resources of the community,*
- (c) promoting colearning among research partners,*
- (d) achieving a balance between research and action that mutually benefits both science and the community,*
- (e) emphasizing the relevance of community-defined problems,*
- (f) employing a cyclical and iterative process to develop and maintain community/ research partnerships,*
- (g) disseminating knowledge gained from the CBPR project to and by all involved partners, and*
- (h) requiring long-term commitment on the part of all partners (Holkup, Tripp-Reier, Salois, & Weinert, 2004, p. 2).*

Quinn et al (2017) published a case study of CBPR practices being employed with youth at risk of homelessness and exposure to violence. The authors cited a “paucity of evidence-based, developmentally appropriate interventions” to address the mental health needs of youth exposed to violence (p. 3). The CBPR process

helped determine the acceptability of a person-centered trauma therapy approach called narrative exposure therapy (NET). The results of three pilot projects combined to inform the design of a randomized controlled trial (RCT) to study the impact of the NET intervention. The three pilot projects engaged researchers and members of the population to be served (youth at risk of homelessness and exposure to violence). The authors of the case study article discussed some of the challenges of working with youth in the CBPR process and research process. Adapted from Quinn et al (2017), these included:

- Compliance with federal regulations for research involving minors (defined as “children” in the policies). Compounding this challenge was the vulnerable status of the youth due to their homeless status, and the frequency with which many of the youth were not engaged with any adults who had legal authority to provide consent for them to participate.
- The team was interdisciplinary, which brings many advantages. However, it also presented challenges regarding different perspectives about how to engage in the varied research processes of participant recruitment and retention, measurement, and intervention.
- Logistics of conducting focus groups with this vulnerable population. Youth encounter difficulties with participating predictably, and for this vulnerable population the practical difficulties are compounded. They experience complex and often competing demands on their schedules, “including school obligations, court, group or other agency appointments, or childcare,” as well as managing public transportation schedules and other barriers (p. 11). Furthermore, members of the group may have pre-existing relationships and social network ties that can impinge on their comfort with openly sharing their experiences or perspectives in the group setting. They may also have skepticism and reservations about sharing with the adults leading the focus group sessions.

Awareness of these challenges can help CBPR teams develop solutions to overcome the barriers. The CBPR process, while time and resource intensive, can result in appropriate intervention designs for under-served populations where existing evidence is not available to guide intervention planning.



A somewhat different approach engages members of the community as consultants regarding interventions with which they may be engaged, rather than a fully CBPR approach. This adapted consultation approach presents an important option for ensuring that interventions are appropriate and acceptable for serving the community. However, community members are less integrally involved in the action-related aspects of defining and implementing the intervention, or in the conduct of the implementation research. An example of this important community-as-consultant approach involved a series of six focus group sessions conducted with parents, teachers, and school stakeholders discussing teen pregnancy prevention among high-school aged Latino youth (Johnson-Motoyama et al., 2016). The investigating team reported recommendations and requests from these community members

concerning the important role played by parents and potential impact of parent education efforts in preventing teen pregnancy within this population. The community members also identified the importance of comprehensive, empowering, tailored programming that addresses self-respect, responsibility, and “realities,” and incorporates peer role models. They concluded that local school communities have an important role to play in planning for interventions that are “responsive to the community’s cultural values, beliefs, and preferences, as well as the school’s capacity and teacher preferences” (p. 513). Thus, the constituencies involved in this project served as consultants rather than CBPR collaborators. However, the resulting intervention plans could be more culturally appropriate and relevant than intervention plans developed by “outsiders” alone.



One main limitation to conducting CBPR work is the immense amount of time and effort involved in developing strong working collaborative relationships—relationships that can stand the test of

time. Collaborative relationships are often built from a series of “quick wins” or small successes over time, where the partners learn about each other, learn to trust each other, and learn to work together effectively.

Chapter Summary

This chapter began with a review of concepts from our earlier course: qualitative, quantitative, mixed-methods, cross-sectional and longitudinal approaches. Expanded content about approach came next: formative, process, outcome, and cost evaluation approaches were connected to the kinds of intervention questions social workers might ask, and participatory research approaches were introduced. Issues of cultural relevance were explored, as well. This discussion of approach leads to an expanded discussion of quantitative study design strategies, which is the topic of our next chapter.

Stop and Think



following activity.

Take a moment to complete the



*An interactive or media element has been excluded
from this version of the text. You can view it online*

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=170>

Module 3 Chapter 2:

Quantitative Design Strategies

Previously, you reviewed different approaches to intervention and evaluation research and learned about a level of evidence (evidence hierarchy) model. These approaches and frameworks relate to how studies are designed to answer different types of research questions. The design strategies differ in the degree to which they address **internal validity** concerns—the ability to conclude that changes or differences can be attributed to an intervention rather than to other causes. Designs for quantitative intervention research are the focus of this chapter.

In this chapter, you learn about:

- pre-experimental, quasi-experimental, and experimental designs in relation to strength of evidence and internal validity concerns;
- how quantitative and single-system study designs are diagrammed;
- examples where different designs were used in social work intervention or evaluation research.

Addressing Internal Validity Concerns via Study Design Strategies

The study designs we examine in this chapter differ in terms of their capacity to address specific types of internal validity concerns. As a reminder of what you learned in our previous course, improving

internal validity is about increasing investigator confidence that the outcomes observed in an experimental study are as believed, and are due to the experimental variables being studied. In the case of intervention research, the experimental variable being studied is the intervention.

Three general internal validity challenges are important to consider addressing with intervention research.

- Were there actually changes that occurred with the intervention (comparing participants' status pre-intervention to their post-intervention status)?
- Did observed changes persist over time (comparing participants' post-intervention status to their status at some later post-intervention follow-up time)?
- Is the observed change most likely due to the intervention itself?

Let's consider how these three questions might relate to the kind of study design choices that investigators make.

Types of Intervention Research Designs

The intervention study design options can be loosely categorized into three types: pre-experimental, quasi-experimental, and experimental. **Pre-experimental designs** do not include comparison (control) groups, only the one group of interest. **Quasi-experimental designs** do include comparison (control) groups, but individuals are not randomly assigned to these groups—the groups are created naturally. **Experimental designs** include random assignment of study participants to the groups being compared in the study. These three types of designs differ in terms of their attempts to control as many alternative explanation factors as possible—the extent to which they address internal validity concerns.

Introducing short-hand for intervention study design diagrams.

It can become quite cumbersome to describe all the elements of an intervention study's design in words and complete sentences; sometimes it is easier to use a diagram instead. This type of short-hand quickly communicates important information about a study design. It can become somewhat confusing, like understanding how the diagram of a football play drawn on a chalkboard translates into reality.



The first thing to know about this intervention study design short-hand is the symbols used and what each symbol means.

- **X** is used to designate that an intervention is being administered.
- **[X]** designates that an “intervention” or event naturally occurred, rather than one imposed by the investigator (for example, a change in policy, a natural disaster, or trauma-inducing crisis event).

- **O** is used to designate that an observation is being made (data are collected).
- **Subscript numbers** are used to designate which intervention or observation is relevant at that point. For example, X_1 might refer to the new intervention that is being tested and X_2 might refer to the second condition where the usual intervention is delivered. And, O_1 might refer to the first observation period (maybe before the intervention), O_2 to the second observation period (maybe after the intervention), and O_3 to a third observation period (maybe after a longer-term follow-up period).
- **R** is used to designate that individual elements were randomly assigned to the different intervention conditions. (Important reminder: This **random assignment** is not about **random selection** of a sample to represent a population; it is about using a randomization strategy for placing participants into the different groups in the experimental design).

Costs & Benefits of Various Design Strategies.

Before we get into discussing the specific strategies that might be adopted for intervention research, it is important to understand that every design has its advantages and its disadvantages. There is no such thing as a single, perfect design to which all intervention research should adhere. Investigators are faced with a set of choices that must be carefully weighed. As we explore their available options, one feature that will become apparent is that some designs are more “costly” to implement than others. This “cost” term is being used broadly here: it is not simply a matter of dollars, though that is an important, practical consideration.

- First, a design may “cost” more in terms of more data collection points. There are significant costs associated with

each time investigators have to collect data from participants—time, space, effort, materials, reimbursement or incentive payments, data entry, and more. For this reason, longitudinal studies often are more costly than cross-sectional studies.

- Second, “costs” increase with higher numbers of study participants. Greater study numbers “cost” dollars (e.g., advertising expenses, reimbursement or incentive payments, cost of materials used in the study, data entry), but also “cost” more in terms of a greater commitment in time and effort from study staff members, and in terms of greater numbers of persons being exposed to the potential risks associated with the intervention being studied.
- Third, some longitudinal study designs “cost” more in terms of the potential for higher rates of participant drop-out from the study over time. Each person who quits a study before it is completed increases the amount of wasted resources, since their data are incomplete (and possibly unusable), and that person may need to be replaced at duplicate cost.



Ten Typical Evaluation/Intervention Study

Designs

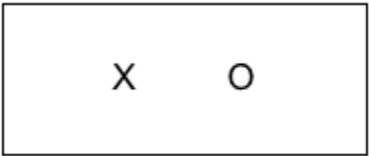
This section presents 10 general study designs that typically appear in intervention and evaluation research. These examples are presented in a general order of increasing ability to address internal validity concerns, but this is offset by increasing costs in resources, participant numbers, and participant burden to implement. Many variations on these general designs appear in the literature; these are 10 general strategies.

#1: Case study.

Original, novel, or new interventions are sometimes delivered under unusual or unique circumstances. In these instances, when very little is known about intervening under those circumstances, knowledge is extended by sharing a **case study**. Eventually, when several case studies can be reviewed together, a clearer picture might emerge about intervening around that condition. At that point, theories and interventions can more systematically be tested. Case studies are considered pre-experimental designs.

An example happened with an adolescent named Jeanna Giese at Milwaukee's Children's Hospital of Wisconsin. In 2004, Jeanna was bitten by a bat and three weeks later was diagnosed with full-blown rabies when it was too late to administer a vaccine. At the time, no treatments were known to be successful for rabies once it has developed; the rabies vaccine only works before the disease symptoms develop. Until this case, full-blown rabies was reported to be 100% fatal. The hospital staff implemented an innovative, theory and evidence-informed treatment plan which became known as the "Milwaukee Protocol." The case study design can be diagrammed the following way, where X represents the "Milwaukee

Protocol” intervention and O represents the observed outcomes of the treatment delivered in this case.



Jeanna Giese became the first person in the world known to survive full-blown rabies. The intervention team published the case study, and a handful of individuals around the world have been successfully treated with this protocol—rabies continues to be a highly fatal, global concern. The Milwaukee Protocol is considered somewhat controversial, largely because so few others have survived full-blown rabies even with this intervention being administered; some authors argue that this single case was successful because of unique characteristics of the patient, not because of the intervention protocol’s characteristics (Jackson, 2013).

This argument reflects a major drawback of case studies, which are pre-experimental designs: the unique sample/small sample size means that individual characteristics or differences can have a powerful influence on the outcomes observed. Thus, the internal validity problem is that the outcomes might be explained by some factors other than the intervention being studied. In addition, with a very small sample size the study results cannot be generalized to the larger population of individuals experiencing the targeted problem—this is an **external validity** argument. (We call this an “N of 1” study, where the symbol N refers to the sample size.) The important message here is that case studies are the beginning of a knowledge building trajectory, they are not the end; they inform future research and, possibly, inform practice under circumstances where uncertainty is high with very new problems or solutions. And,

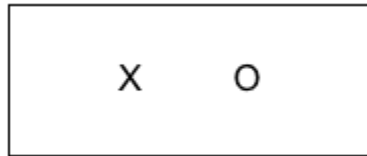
just in case you are curious: although some permanent neurological consequences remained, Jeanna Giese completed a college education, was married in 2014, and in 2016 became the mother of twins.



#2: Post-intervention only.

Looking a great deal like the case study design is a simple pre-experimental design where the number of individuals providing data after an intervention is greater than the single or very small number in the case study design. For example, a social work continuing education training session (the intervention) might collect data from training participants at the end of the session to see what they learned during the training event. The trainers might ask participants to rate how much they learned about each topic covered in the training (nothing, a little, some, a lot, very much) or they might present participants with a quiz to test their post-training knowledge of content taught in the session. The **post-only**

design diagram is the same as what we saw with the single case study; the only difference is that the sample size is greater—it includes everyone who completed the evaluation form at the end of the training session rather than just a single case.



The post-intervention only design is cross-sectional in nature (only one data collection point with each participant). This design strategy is extremely vulnerable to internal validity threats. The investigator does not know if the group's knowledge changed compared to before the training session: participants quizzed on knowledge may already have known the material before the training; or, a perception of how much they learned may not accurately depict how much they learned. The study design does not inform the investigators if the participants' learning persisted over time after completing the training. The investigators also do not have a high level of confidence that the training session was the most likely cause of any changes observed—they cannot rule out other possible explanations.

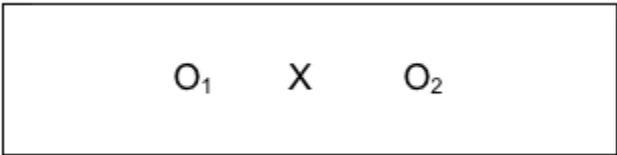
In response to the internal validity threat concerning ability to detect change with an intervention, an investigator might ask study participants to compare themselves before and after the intervention took place. That would still be a simple post- only design because there is only one time point for data collection: post-intervention. This kind of retrospective approach is vulnerable to bias because it relies on an individual's ability to accurately recall the past and make a valid comparison to the present, a comparison that hopefully is not influenced by their present state-of-mind. It helps to remember what you learned from SWK 3401 about the

unreliability of the information individuals remember and how memories become influenced by later information and experiences.

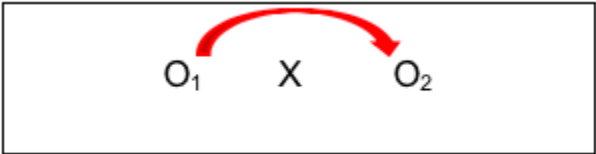


#3: Pre-/Post- Intervention Comparison.

A wiser choice in terms of internal validity would be to directly compare data collected at the two points in time: pre-intervention and post-intervention. This **pre-/post-design** remains a pre-experimental design because it lacks a comparison (control) group. Because the data are collected from the same individuals at two time points, this strategy is considered longitudinal in nature. This type of pre-/post-intervention design allows us to directly identify change where observed differences on a specific outcome variable might be attributed to the intervention. A simple pre-/post- study design could be diagrammed like this:



Here we see an intervention X (perhaps our social work in-service training example), where data were still collected after the intervention (perhaps a knowledge and skills quiz). However, the investigators also collected the same information prior to the intervention. This allowed them to compare data for the two observation periods, pre- and post- intervention. See the arrow added to the diagram that shows this comparison:

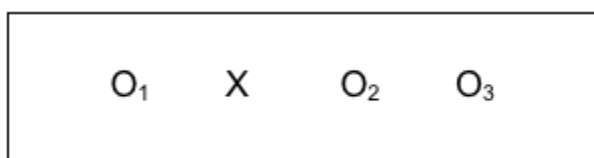


While this pre-/post- intervention design is stronger than a post-only design, it also is a bit more “costly” to implement since there is an added data collection point. It imposes an additional burden on participants, and in some situations, it simply might not be possible to collect that pre-intervention data. This design strategy still suffers from the other two internal validity concerns: we do not know if any observed changes persisted over time, and we do not have the highest level of confidence that the changes observed can be attributed to the intervention itself.

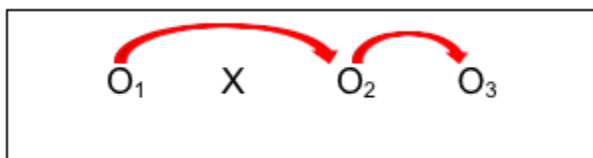


#4: *Pre-/Post-/Follow-Up Comparison.*

Investigators can improve on the pre-/post- study design by adding a follow-up observation. This allows them to determine whether any changes observed between the pre- and post- conditions persisted or disappeared over time. While investigators may be delighted to observe a meaningful change between the pre- and post-intervention periods, if these changes do not last over time, then intervention efforts and resources may have been wasted. This **pre-/post-/follow-up design** remains in the pre-experimental category, and would be diagrammed like this:



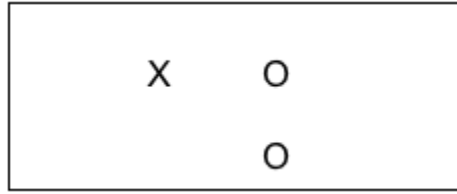
Here we see the intervention (X), with pre-intervention data, post-intervention data, and follow-up data being collected (O_1 , O_2 , and O_3). As an example, Wade-Mdivanian, Anderson-Butcher, Newman, and Ruderman (2016) explored the impact of delivering a preventive intervention about alcohol, tobacco and other drugs in the context of a positive youth development program called Youth to Youth International. The outcome variables of interest were the youth leaders' knowledge, attitudes, self-efficacy, and leadership before the program, at the program's conclusion, and six months after completing the program. The authors concluded that positive changes in knowledge and self-efficacy outcomes were observed in the pre-/post- comparison, though no significant change was observed in attitudes about alcohol, tobacco and other drugs; these differences persisted at the six-month follow-up. See the pre-/post-/follow-up design diagrammed with comparison arrows added:



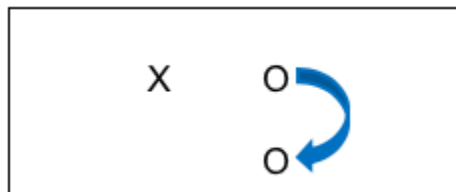
This design resolved two of the three internal validity challenges, although at an increased cost of time, effort, and possibly other resources with the added data point. But an investigator would still lack confidence that observed changes were due to the intervention itself. Let's look at some design strategies that focus on that particular challenge.

#5: Comparison Groups.

In our prior course we learned how to compare groups that differed on some characteristic, like gender for example. Comparison groups in intervention research allow us to compare groups where the difference lies in which intervention condition each received. By providing an experimental intervention to one group and not to the other group, investigator confidence increases about observed changes being related to the intervention. You may have heard this second group described as a **control group**. They are introduced into the study design to provide a benchmark for comparison with the status of the intervention group. The simplest form of a comparison group design, which is a quasi-experimental type of design, a **post-only group design** can be diagrammed as follows:



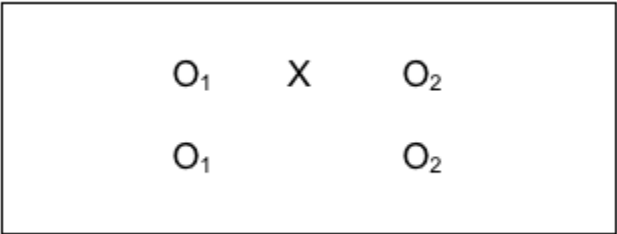
Consider the possibility in our earlier example of evaluating a social work training intervention that the team decided to expand their post only design to include collecting data from a group of social workers who are going to get their training next month. On the same day that the data were collected from the trained group, the team collected data from the untrained social workers, as well. This situation is a post-only design where the top row shows the group who received the training intervention (X) and the outcome was measured (O), and the bottom row shows the group without the training intervention (no X was applied) also being measured at the same point in time as the intervention group (O). This remains a cross-sectional study because each individual was only observed once. The following diagram shows the arrow where investigators compared the two groups on the outcome variables (knowledge and skills quiz scores, using our training example). If the training intervention is responsible for the outcome, the team would see a significant difference in the outcome data when comparing the two groups; hopefully in the direction of the trained group having better scores.



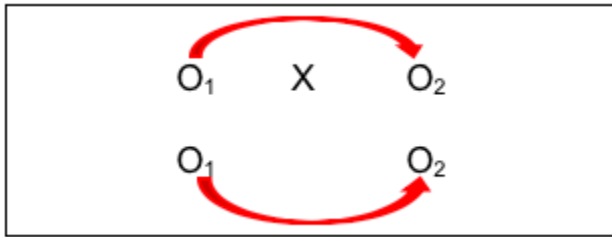
While this design has helped boost investigator confidence that the outcomes observed with the intervention group are likely due to the intervention itself, this post-only design “costs” more than a post-only single group (pre-experimental) study design. This post-only group design still suffers from a significant concern: how would an investigator know if the differences between the two groups appeared only after the intervention or could the differences always have existed, with or without the intervention? With this design, that possibility cannot be ignored. An investigator can only hope that the two groups were equivalent prior to the intervention. Two internal validity questions remain unanswered: did the outcome scores actually demonstrate a change resulting from intervention, and did any observed changes persist over time. Let’s consider some other combination strategies that might help, even though they may “cost” considerably more to implement.

#6: Comparison Group Pre-/Post- Design.

A giant leap forward in managing internal validity concerns comes from combining the strategy of having both an intervention and a comparison group (which makes it quasi-experimental) with the strategy of collecting data both before and after the intervention (which makes it longitudinal). Now investigators are able to address more of the major validity concerns. This *comparison group pre-/post- design* is diagrammed as follows:



What the investigators have done here is collect data from everyone in their study, both groups, prior to delivering the intervention to one group and not to the other group (control group). Then, they collected the same information at the post-intervention time period for everyone in their study, both groups. The power in this design can be seen in the following diagrams that include arrows for the kinds of longitudinal pre-/post- comparisons that can be assessed, measuring change.



This is a pre-/post- comparison, indicating if change occurred with the intervention group after the intervention. Investigators would hope the answer is “yes” if it is believed that the intervention should make a difference. Similarly, the investigators could compare the non-intervention group at the two observation points (the lower arrow). Hopefully, there would be no significant change without the intervention.

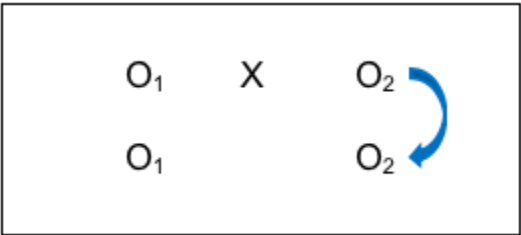
You might be wondering, “Why would there be change in the no intervention group when nothing has been done to make them change?” Actually, with the passage of time, several possible explanatory events or processes could account for change.

- The first is simple **maturation**. Particularly with young children, developmental change happens over relatively short periods of time even without intervention.
- Similarly, the passage of time might account for symptom improvement even without intervention. Consider, for

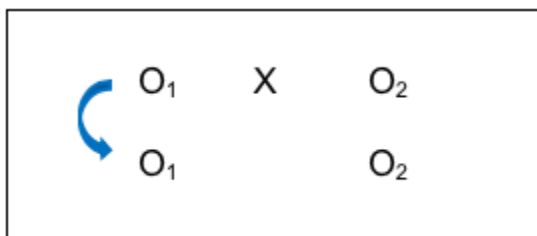
example, the old adage that, “if you treat a common cold it will take 7 to 10 days to get better; if you don’t treat it, it will take a week to a week-and-a-half.” Change without intervention is called **spontaneous or natural change**. Either way, there is change—with or without intervention.

- Third, given time, individuals in the no intervention group might seek out and receive other interventions not part of the study that also can produce change. This could be as simple as getting help and support from friends or family members; or, seeking help and advice on the internet; or, enrolling in other informal or formal treatment programs at the same time.

The benefit of combining the comparison groups and pre-/post-designs continues to emerge as we examine the next diagram showing comparisons an investigator can consider:



This group comparison of post-intervention results indicates whether there is a difference in outcomes between people who received the intervention and people who did not. Again, investigators would hope the answer is “yes,” and that the observed difference favors the intervention group. But with this design, an investigator can go even further in ruling out another possible explanation for outcome differences. Consider the power of this comparison:



By comparing the two groups BEFORE the intervention took place, an investigator can hopefully rule out the possibility that post-intervention group differences were actually a reflection of pre-existing differences; differences that existed prior to the intervention. In this case, the investigator would hope to observe no significant differences in this comparison. This “no differences” result would boost confidence in the conclusion that any observed post-intervention differences were a function of the intervention itself (since there were no pre-existing differences). Not being able to rule out this possibility was one limitation of the post-only group comparison strategy discussed earlier.

A Note About Non-Treatment, Placebo, and Treatment as Usual Comparison Groups.

What we have diagrammed above is a situation where the second group received no treatment at all. This, however, is problematic in three ways. First, the ethics of intentionally not serving individuals who are seeking help with a social work problem is concerning. Second, the scientific integrity of such studies tends to suffer because non-treatment “control” groups tend to have high rates of study drop out (attrition) as those participants seek help elsewhere. Third, the results of these studies are coming under great scrutiny as the world of behavioral science has come to realize that any

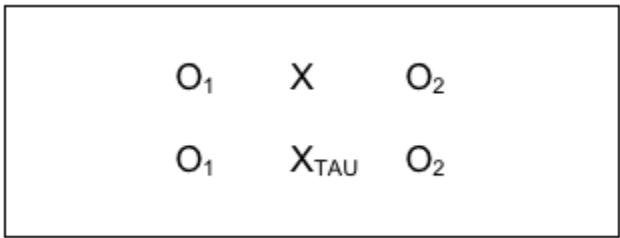
treatment is likely to be better than no treatment—thus, study results where the tested intervention is significantly positive may be grossly over-interpreted. Compared to other treatments, what appears to be a fantastic innovation may be no better.

Medical studies often include a comparison group who receives a “fake” or neutral form of a medication or other intervention. In medicine, a placebo is an inert “treatment” with a substance that has no likely known effect on the condition being studied, such as a pill made of sugar, for example. The approximate equivalent in behavioral science is an intervention where clients are provided only with basic, factual information about the condition; nothing too empowering. In theory, neither the placebo medication nor the simple educational materials are expected to promote significant change. This is a slight variation on the **non-treatment control** condition. However, over 20 years of research provided evidence of a **placebo effect** that cannot be discounted. In one systematic review and meta-analysis study (Howick et al, 2016) the size of effect associated with placebos were no different or even larger than treatment effects. This placebo effect is, most likely, associated with the psychological principles of motivation and expectancies—expecting something to work has a powerful impact on behavior and outcomes, particularly with regard to symptoms of nausea and pain (Cherry, 2018). Of course, this means that participants receiving the placebo believe they are (or could be) receiving the therapeutic intervention. Also interesting to note is that participants sometimes report negative side-effects with exposure to the placebo treatment.

In medication trials, the introduction of a placebo may allow investigators to impose a double-blind structure to the study. A **double-blind study** is one where neither the patient/client nor the practitioner/clinician knows if the person is receiving the test medication or the placebo condition. The double-blind structure is imposed as means of reducing practitioner bias in the study results from either patient or practitioner/clinician beliefs about the experimental medication. However, it is difficult to disguise

behavioral interventions from practitioners—it is not as simple as creating a real-looking pill or making distilled water look like medicine.

More popular in intervention science today is the use of a **treatment as usual condition (TAU)**. In a TAU study, the comparison group receives the same treatment that would have been provided without the study being conducted. This resolves the ethical concerns of intentionally denying care to someone seeking help simply to fulfill demands of a research design. It also helps resolve the study integrity concerns mentioned earlier regarding non-treatment studies. But this design looks a little bit different from the diagram of a comparison group design. You will notice that there is a second intervention “X” symbol in the diagram and that the new X has a subscript notation (TAU) so you can tell the two intervention groups apart, X and X_{TAU}.



#7: Single-System Design.

A commonly applied means of evaluating practice is to employ a quasi-experimental single-system design. This approach uniquely combines aspect of the case study with aspects of the pre-experimental pre-/post-design. Like the case study, the data are collected for one single case at a time—whether the element or unit of study is an individual, couple, family, or larger group, the data represent the behavior of that element over time. In that sense, the

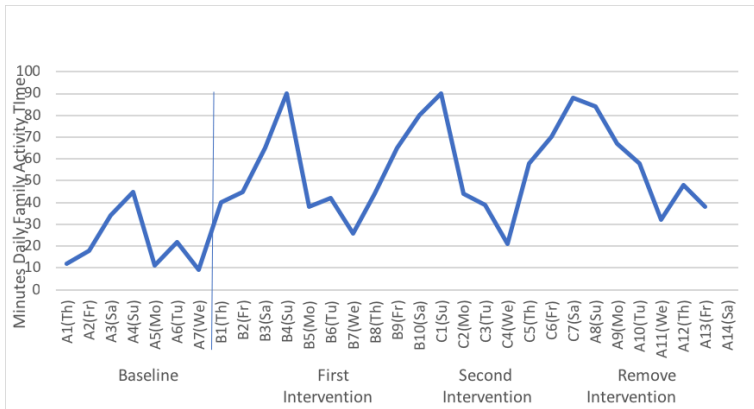
single-system design is longitudinal—repeated measurements are drawn for the same element each time. It is a quasi-experimental design in that conditions are systematically varied, and outcomes measured for each variation. This approach to evaluating practice is often referred to as a single-subject design. However, the fact that the “subject” might be a larger system is lost in that label, hence a preference for the single-“system” design label.

The single-system design is sufficiently unique in implementation that an entirely different notation system is applied. Instead of the previous Xs and Os, we will be using A's and B's (even C's and D's). The first major distinction is that instead of a pre-intervention measurement (what we called O₁) we need a pre-intervention baseline. By definition, a line is the distance between two or more points, thus in order to be a baseline measurement, at least two and preferably at least 7 pre-intervention measurement points are utilized. For example, if the target of intervention with a family is that they spend more activity time together, a practitioner might have them maintain a daily calendar with the number of minutes spent in activity time is recorded for each of 7 days. This would be used as a baseline indication of the family's behavioral norm. Thus, since there are 7 measurement points replacing the single pre-intervention observation (O₁), we designate this as A₁, A₂, A₃ and so forth to A₇ for days 1 through 7 during the baseline week—this might extend to A₃₀ for a month of baseline recording.

Next, we continue to measure the target behavior during the period of intervention. Because the intervention period is different from the baseline period, we use the letter B to indicate the change. This corresponds to the single point in our group designs where we used the letter X to designate the intervention. Typically, in a pre-/post- intervention study no data are collected during the intervention, X, which is another difference in single-system designs. Let's say that the social worker's intervention was to text the family members an assigned activity for each of 10 days, and they continued to record the number of minutes when they were engaged in family activity time each day. Because it is a series of

measurement points, we use B₁, B₂, B₃ and so forth for the duration of the intervention, which is B₁₀ in this case. The next step is to remove the daily assignment text messages, giving them just one menu on the first of the next 7 days, from which the family is expected to pick an activity and continue to record time spent in family activity time each day. This would be a different form of intervention from the “B” condition, so it becomes C₁, C₂, C₃, and so forth to the end, C₇. Finally, the social worker no longer sends any cues, and the family simply records their daily family activity time for another week. This “no intervention” condition is the same as what happened at baseline before there was any intervention. So, the notation reverts back to A’s, but this time it is A₈, A₉, A₁₀, and so forth to the end of the observation period, A₁₄. For this reason, single system design studies are often referred to as “ABA” designs (initial baseline, an intervention line, and a post-intervention line) or, in our example an “ABCA” design since there was a second intervention after the “B” condition. This single-system design notation is different from the notation using Xs and Os because data are collected multiple times during each phase, rather than at each single point during pre-intervention, intervention, and post-intervention.

The data could be presented in a table format, but a visual graph depicts the trends in a more concise, communicative manner. What the practitioner is aiming to do with the family is to look at the pattern of behavior in as objective manner as possible, under the different manipulated conditions. Here is a graphical representation from one hypothetical family. As you can see, there is natural variation in the family’s behavior that would be missed if we simply used a single weekly value for the pre- and post-intervention periods instead.

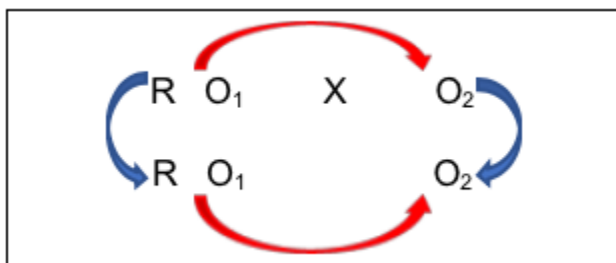


Together, the practitioner and the client family can discuss what they see in the data. For example, what makes Wednesday family activity time particularly difficult to implement and does it matter given the what happens on the other days? How does it feel as a family to have the greater activity days, and is that rewarding to them? What happens when the social worker is no longer prompting them to engage in activities and how can they sustain their gains over time without outside intervention? What new skills did they learn that support sustainability? What will be their cue that a “booster” might be necessary? What the data allow is a clear evaluation of the intervention with this family, which is the main purpose of practice evaluation using the single-system design approach.



#8: Random Control Trial (RCT) Pre-/Post- Design.

The major difference between the comparison group pre-/post-design just discussed and the **random control trial (RCT)** with pre-/post- design is that investigators do not have to rely so much on hoping that the two comparison groups were initially equivalent: they randomly assign study participants to the two groups as an attempt to ensure that this is true. There is still no guarantee of initial group equivalence; this still needs to be assessed. But the investigators are less vulnerable to the bad luck of having the two comparison groups being initially different. This is what the RCT design looks like in a diagram:



Many authors of research textbooks describe the RCT as the “gold standard” of intervention research design because it addresses so many possible internal validity concerns. It is, indeed, a powerful experimental design. However, it is important to recognize that this type of design comes at a relatively high “cost” compared to some of the others we have discussed. Because there are two comparison groups being compared, there are more study participant costs involved than the single group designs. Because there are at least two points in time when data are collected, there are more data collection and participant retention costs involved than the single time point post-only designs. And random assignment to experimental conditions or groups is not always feasible in real-world intervention situations. For example, it is difficult to apply random assignment to conditions where placement is determined by court order, making it difficult to use an RCT design for comparing a jail diversion program (experimental intervention) with incarceration (treatment as usual control group). For these practical reasons, programs often settle on group comparison designs without random assignment in their evaluation efforts.

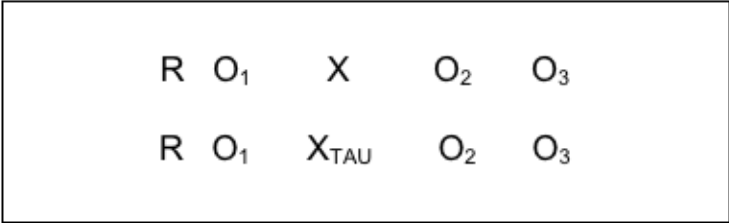


Remembering how to interpret statistics in a research report. The investigators reported that for the intervention group, the mean reduction in AUDIT-12 scores was significantly greater for the SBI intervention group than for the TAU group on a one-way Anova: $M=18.83$ and 13.52 respectively, $F(1,148)=6.34$, $p<.01$. This statement indicates the following:

1. The mean (M) for the innovative SBI intervention group $=18.83$
2. The mean (M) for the treatment as usual (TAU) group $=13.52$.
3. The computed F statistic in the analysis of variance (Anova) $=6.34$.
4. Degrees of freedom for the Anova were 1 (2 groups -1) and 148 (149 cases - 1): $df=(1, 148)$.
5. The F -distribution test statistic at those degrees of freedom is significant at a level less than the criterion $p<.05$.
6. We reject the null hypothesis of no difference between the groups, concluding that a significant difference exists.

A slight variant on this study design was used to compare a screening and brief intervention (SBI) for substance misuse problems to the usual (TAU) condition among 729 women preparing for release from jail and measured again during early reentry to community living (Begun, Rose, & LeBel, 2011). One-third of the

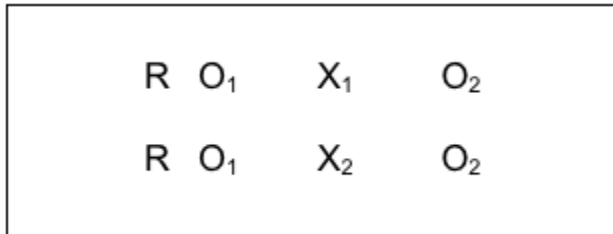
women who had positive screen results for a potential substance use disorder were randomly assigned to the treatment as usual condition (X_{TAU}), and two-thirds to the SBI experimental intervention condition (X) as diagrammed below. The investigators followed 149 women three months into post-release community reentry (follow-up observation), and found that women receiving the innovative SBI intervention had better outcomes for drinking and drug use during early reentry (three months post-release); the mean difference in scores on the AUDIT-12 screening instrument prior to jail compared to reentry after release was more than 5 points (see Back to Basics Box for more details). The *random controlled trial with pre/post/follow-up study design* looked like this:



#9: Random Control Trial (RCT) Comparing Two (or More) Interventions.

The only difference between the design just discussed and the random control trial that compares two interventions is the indicator associated with the intervention symbol, X. The X_1 and X_2 refer to two different interventions, the same way the X_1 and X_{TAU} reflected two different intervention conditions above. This kind of design is used to determine which of two interventions has greater effectiveness, especially if there is a considerable difference in cost of delivering them. The results can help with a cost-benefit comparison which program administrators and policy decision-

makers use to help make decisions about which to fund. It is possible to compare more than two interventions by simply adding additional lines for each group (X_3 , X_4 , and so on). Each group added, however, also adds considerably to the “costs” of conducting the study.



The COMBINE Project was an historic example where multiple intervention approaches were compared for their effectiveness in treating alcohol dependence (see NIAAA Overview of COMBINE at <https://pubs.niaaa.nih.gov/publications/combine/overview.htm>).

Two medications were being compared (acamprosate and naltrexone) along with medical management counseling for adherence to the medication protocol (MM), and medication with MM was compared to medication with MM plus a cognitive behavioral psychotherapy intervention (CBI). Clients were randomly assigned to 9 groups with approximately 153 clients per group:

X_1 -acamprosate placebo + naltrexone placebo + MM (no CBI)

X_2 -acamprosate + naltrexone placebo + MM (no CBI)

X_3 -naltrexone + acamprosate placebo + MM (no CBI)

X_4 -acamprosate + naltrexone + MM (no CBI)

X_5 - acamprosate placebo + naltrexone placebo + MM + CBI

X_6 - acamprosate + naltrexone placebo + MM + CBI

X_7 - naltrexone + acamprosate placebo + MM + CBI

X_8 - acamprosate + naltrexone + MM + CBI

X₉-no pills or MM, CBI only

The COMBINE Project was an extremely “costly” study to conduct given the number of study participants needed to meet the study’s design demands—one reason why it was a collaboration across multiple study sites. The results included participants in all 9 groups showing at least some improvement as measured by a reduction in drinking—intervention is almost always an advantage over no intervention. The poorest results were observed for the last group (X₉), those receiving the specialty cognitive behavioral intervention (CBI) alone, with no medication or placebo medication. Surprising was that best outcomes were observed for the group receiving CBI with both placebo medications and medication management (MM) counseling (X₅)! The other group with best outcomes was the group receiving naltrexone with MM counseling but no CBI (X₃). The investigative team concluded that pharmacotherapy combined with medication counseling can yield clinically significant alcohol treatment outcomes, and this can be delivered in primary care settings where specialty alcohol treatment is unavailable (Pettinati, Anton, & Willenbring, 2006). They were surprised, also, by the lack of observed advantage to using the medication acamprosate because it had so much evidence support from studies conducted in European countries. This map of the design does not show the multiple follow-up observations conducted in the study.

| | | | |
|---|----------------|----------------|----------------|
| R | O ₁ | X ₁ | O ₂ |
| R | O ₁ | X ₂ | O ₂ |
| R | O ₁ | X ₃ | O ₂ |
| R | O ₁ | X ₄ | O ₂ |
| R | O ₁ | X ₅ | O ₂ |
| R | O ₁ | X ₆ | O ₂ |
| R | O ₁ | X ₇ | O ₂ |
| R | O ₁ | X ₈ | O ₂ |
| R | O ₁ | X ₉ | O ₂ |

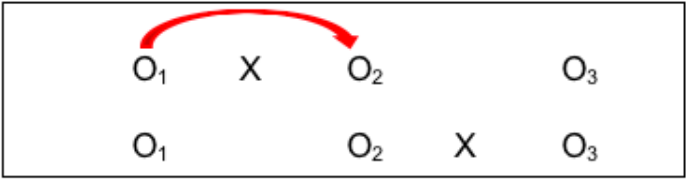
#10: Comparison Group with a Delayed Intervention Condition.

One way of overcoming the ethical dilemma of an intervention/no intervention study design is to offer the intervention condition later for the second group. Not only does this more evenly distribute risks and benefits potentially experienced across both groups, it also makes sense scientifically and practically because it allows another set of comparison conditions to be longitudinally analyzed for a relatively smaller additional “cost.” Here is what this design strategy might look like using our study design notation:

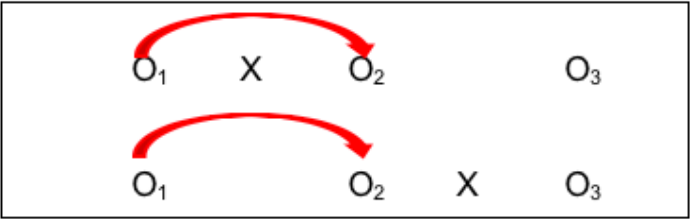
| | | | |
|----------------|---|----------------|----------------|
| O ₁ | X | O ₂ | O ₃ |
| O ₁ | | O ₂ | X |
| | | | O ₃ |

Look at all the information available through this design:

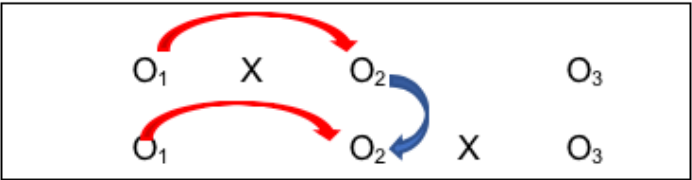
- First is a simple pre-/post- design (O_1 to O_2 with X), where investigators hope to see a difference;



- Second is the control group pre-/post- comparison (O_1 to O_2 without intervention X), where investigators hope for no difference, or at least significantly less difference than for the group with the intervention early on;

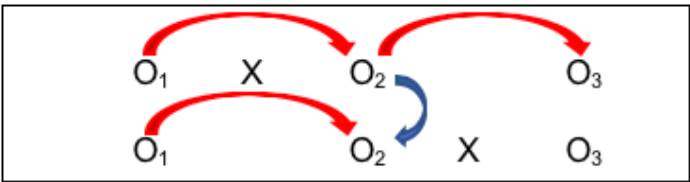


- Third is the comparison of the group with the intervention (X) to the group without the intervention, after the intervention was delivered (O_2 for each group), where the investigators hope to see a significant difference between the two groups, and that the difference favors the group who received the intervention (X);

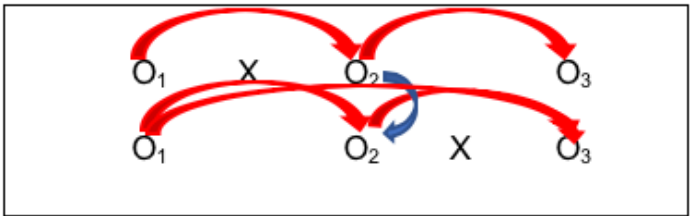


- Fourth is the post-/follow-up comparison for the group with the intervention (O_2 to O_3), where investigators hope that the changes seen on O_2 persisted at O_3 (or are even more

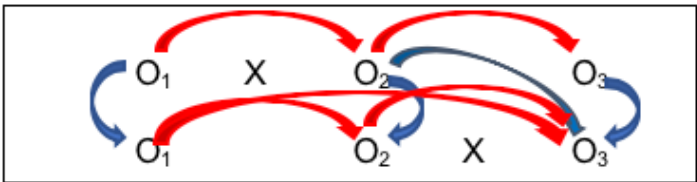
improved, not that they declined over time;



- Fifth is the ability to replicate the results of the first group receiving the intervention with the results for the second group to receive the intervention (O₂ for the first group with O₃ for the second group—actually, change from O₁ to O₂ for the first group and change from O₁ to O₃ for the second group). Ideally, the investigators would see similar results for the two groups—those who received the intervention early on and those who received it a bit later;



- Sixth is the ability to test the assumption that the two groups were similar prior to the study beginning—that the differences observed between them were related to the intervention and not pre-existing group differences (O₁ for both groups), as well as differences immediately after intervention and between follow-up and immediately post-intervention.



This study design was used to examine the effectiveness of equine-assisted therapy for persons who have dementia (Dabelko-Schoeny,

et al., 2014). The study engaged 16 participants with Alzheimer's Disease in activities with four horses (grooming, observing, interacting, leading, and photographing). Study participants were randomly assigned to two groups: one received the intervention immediately the other did not; however, this latter group received the intervention later, after the intervention was withdrawn from the first group. When not receiving the equine-assisted therapy, participants received the usual services. Thus, the investigators were able to make multiple types of comparisons and ensured that all participants had the opportunity to experience the novel intervention. The team reported that the equine-assisted therapy was significantly associated with lower rates of problematic/disruptive behavior being exhibited by the participants in their nursing homes and higher levels of "good stress" (sense of exhilaration or accomplishment) as measured by salivary cortisol levels.

Major disadvantages of this study design:

- it may be that social work interventions are best delivered right away, when the need arises, and that their impact is diminished when delivered after a significant delay;
- the design is costlier to implement than either design alone might be, because it requires more participants and that they be retained over a longer period of time.



Chapter Conclusion

As you can see, social work professionals and investigators have numerous options available to them in planning how to study the impact of interventions and evaluate their practices, programs, and policies. Each option has pros and cons, advantages and disadvantages, costs and benefits that need to be weighed in making the study design decisions. New decision points arise with planning processes related to study measurement and participant inclusion. These are explored in our next chapters.

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=171>

Module 3 Chapter 3:

Measurement

This chapter examines quantitative measurement principles as applied in intervention and evaluation research, keeping in mind that qualitative methods have the potential to:

- Inform subsequent quantitative study research questions,
- Inform the development of measurement tools, procedures, and participant recruitment/retention in subsequent quantitative studies,
- Assess intervention implementation and fidelity during a quantitative study, and
- Develop a context for interpreting the results of a quantitative study.

Thus, quantitative studies for understanding social work interventions can be enhanced when qualitative study approaches are also applied, before, during, and after.

In this chapter, you learn:

- how quantitative measurement content from our prior course (variable types, validity, reliability, and locating measurement instruments) relates to research for understanding social work interventions;
- issues related to measuring change in intervention and evaluation research;
- principles and issues in addressing cultural relevance in designing intervention and evaluation research.

Measures & Variables

You may recall from our earlier course that the tools used in data collection are used to measure variables important and relevant for the study aims, questions, and design. Searching the literature on a topic might result in many different measures being identified for gathering data about a study variable—investigators are then challenged with the need to evaluate the different options and select the one best suited to their purposes. Here are several important points considered in this decision-making process:

- how adequately the tool measures the target variable (a validity issue)
- how accurately/consistently the tool measures the target variable (a reliability issue)
- how appropriate the tool is for the population engaged in the study (age, ability, and cultural relevance, for example)
- participant burden involved with its use
- cost and whether it is available in the public domain (or permission must be purchased from the developer/copyright holder)
- whether modifications necessary to make it fit the study and what the risks to validity and reliability might be when modified (the psychometric evidence no longer applies to a modified measure)
- whether an appropriate tool exists, or one needs to be created (with unknown validity, reliability, and other psychometric properties).

Let's consider an example from a prisoner reentry study (Begun, Early, & Hodge, 2016). The research question was expressed as, "What disparities exist between the mental health and addiction treatment services that adults need during community reentry after incarceration and services they actually receive?" First, the investigators faced challenging decisions about measuring "need"

for these different service types. Three possible options considered included:

- Individual study participants self-identified their need status;
- Standardized screening tests indicated a person's needs; and,
- Individuals' past history of service use implied an ongoing, present need.

Each option had associated advantages and disadvantages. For example, while self-identification is responsive to the social work value placed on self-determination, it is also confounded by individual differences in perception. Thus, the data lack consistency in measurement values. This problem of **measurement inconsistency** means that this option for measuring the variable is not a strong measurement choice. Two different individuals facing identical circumstances might interpret their needs quite differently: one might identify a high need for service while the other similarly-situated person identifies the need as low. The result is less a measure of the need for services than it is a measure of perceptions of need. Measurement inconsistency is a relatively common problem arising in rating scales, for instance when someone is asked to rate frequency in terms of “seldom, sometimes, often.” One person's interpretation of “seldom” might be very different from another person's. Measurement inconsistency is a less relevant problem when the aim of the study involves learning about individuals' personal perceptions (internal subjective states, beliefs, opinions, mood) rather than learning about external, objective realities and events.



The use of standardized screening tools has the advantage of imposing a greater degree of measurement consistency, reducing “perceptions” as a source of individual differences and measurement error. However, standardized measurement tools may not exist for measuring study variables of interest, particularly in new, emerging areas of research. In addition, standardized measurement tools may not be appropriate for the population providing the data—they may not be appropriate for the participants’ age, ability/disability, life circumstances, or culture. For example, the standardized alcohol screening tool called the AUDIT (Alcohol Use Disorders Identification Test) has strong psychometric properties and relevance for a broad range of adults in the general population, is relatively brief (10 items taking about 2 minutes), and available in the public domain. However, the most predictive questions are the first three that ask about quantity and frequency of drinking—at least in terms of screening with young and middle-aged adults (Lee, Jung, & Choi, 2018). Many of these questions, however, are of limited value in screening persons living in restricted environments, such as jails, prisons, or residential treatment programs. The questions presume that individuals currently have access to alcohol:

- How often do you have a drink containing alcohol?
- How many drinks containing alcohol do you have on a typical day when you are drinking?
- How often do you have six or more drinks on one occasion?

To effectively screen for probable risk during community reentry after release from the institutional setting the questions might need to be modified to refer to a period of time prior to institutionalization. Otherwise, the risk of failing to identify persons with probably alcohol use disorders is high since they may be answering “never” to these items while living in a controlled environment. This modification, however, is made at the risk of unknown reliability and validity for the measure’s modified version.

The third option, using an individual’s past history as an indicator of present or future needs, relies on several assumptions, any of which may be flawed. For example, past access to services may have been influenced by racial, ethnic, gender, gender-identity, socio-economic, or diagnosis inequities in the system rather than differences in actual need. Or, for example, an individual might not have the same needs moving forward if the problem has been effectively addressed with past services. On the other hand, the need may have emerged recently and not be detected through a measure of past need.



As you can see, identifying the best way to measure variables in social work research is not always simple and straight forward. The investigators in our example (Begun, Early, & Hodge, 2016) elected to use a multi-pronged approach where all three measurement strategies were combined. This, however, was also associated with a significant degree of participant burden, generating many interview questions. The varied tools used included standardized measures, modified measures, and some items developed specifically for the study at hand. In this way, they were able to include a wide range of variables and variable types. This leads to a review of variable types and extension of what we learned in the previous course to research for understanding social work interventions. This will be helpful both in terms of thinking about the choices made in published investigations and choices you might need to make in your own work.

Variable Types & Levels Revisited and Extended

In our earlier course you learned about different types of variables; these continue to be relevant for research concerning social work interventions. Let's see how this information (still) applies.

Dependent and independent variables.

As you may recall from our earlier course, **independent variables** are manipulated in the study design, and **dependent variables** are where the impact of changing the independent variable is observed. The independent variable in intervention studies usually relate to the intervention itself, for example:

- which intervention is being delivered (e.g., innovative

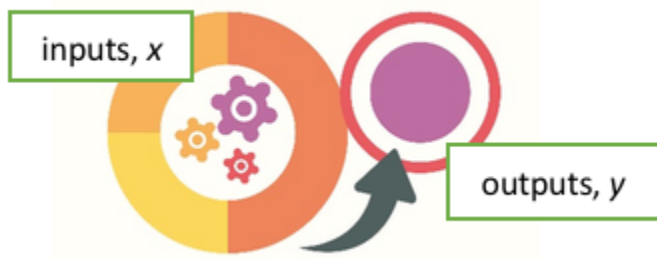
intervention vs control or treatment as usual/TAU condition;
two or more different interventions compared),

- intervention “dose” or intensity (e.g., high, medium, or low intensity/frequency),
- who delivers the intervention (e.g., MSW, BSW, or paraprofessional practitioner),
- when the intervention is delivered (e.g., at first symptoms vs when symptoms exceed a clinical threshold/diagnosis criterion point).

Dependent variables in intervention and evaluation research are usually the outcomes investigators examine in terms of an intervention effects or impact. For example, this might include variables we hope to change, such as:

- number of symptoms,
- frequency of symptoms,
- severity/intensity of symptoms,
- co-occurring problems,
- cost of delivering services,
- access or barriers to service,
- attitudes, opinions, knowledge, skills or behaviors.

Note that the definition of independent variables can be a bit confusing when the “intervention” is a natural event that no investigator intentionally manipulated—an event such as a natural disaster (hurricane, tornado, tsunami, or earthquake); a traumatic critical incident (terrorist threat, violence incident, or other human-directed event), or political protest/disruptive innovation (e.g., Black Lives Matter, #LouderTogether, #MeToo, or #NeverAgain movement). Social workers might be interested to see how a specific outcome changed in relation to a naturally occurring “intervention.”



Nominal/categorical variables.

Sometimes dependent or independent variables represent a specific set of categories. For example, in our design discussion we saw examples where the independent variable (X) had two categories:

- intervention (X) and non-intervention control,
- innovative intervention (X) and treatment as usual (X_{TAU}), or
- two different interventions (X_1 and X_2).

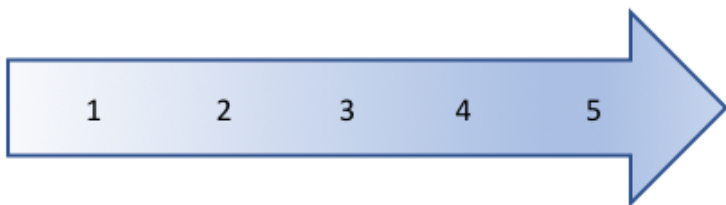
For example, the dependent variable (y) might be treatment completion versus non-completion—two categories. Or, it might be symptom change measured in three categories: symptom improvement, symptom degradation (decline), or no change in symptoms. The dependent (outcome) variable could reflect individuals meeting specific demographic characteristics—for example, national origin among new Americans, gender identity, single- or two-parent headed households, children living in their family homes or out-of-home/foster care, employment status, diagnostic category, access to healthful and affordable food, water security, exposure to violence, and others. These variables are nominal because there is no intrinsic numeric order to the categories. These variables might be independent variables in some studies (e.g., how individuals are sorted into categories); they are

dependent variables if they represent the outcome that an intervention is attempting to modify.



Ordinal variables.

The categories of some variables do have an intrinsic numeric order or sequence. An intervention research example involves the manipulation of intervention dose delivered to different groups (dose being the independent variable, x): perhaps low dose (1 visit), medium dose (2-4 visits), high dose (5 or more visits). Studies demonstrating equal effectiveness of planned brief therapies compared to lengthier or unlimited therapy included comparisons of dose in this manner (Smyrinos & Kirkby, 1993, for example). The outcome (dependent) variable might be ordinal, as well. For example, the outcome might be no symptoms present, non-clinically significant symptoms present, clinically relevant symptoms present; or, the outcome might be level of client satisfaction indicated on a 5-point scale (not at all to extremely much).



Interval/scale/continuous variables.

Sometime variables have numerical values that vary in equal-sized increments or intervals. For example, outcomes are often measured on assessment scales such as the Alcohol Use Disorder Identification Test (AUDIT) where possible scores range from 0 to 40 and higher scores mean greater probability of an alcohol use disorder.

The AUDIT example is rich in that investigators (or clinicians) also are able to convert the interval scale score into a dichotomous categorical score that could be considered ordinal. AUDIT scores of 8 or greater designate a “positive” screening result for men, and lower scores are “negative” screening results for men; AUDIT scores of 7 designate a “positive” screening result for women, and lower scores are “negative screening results for women (Bradley, Boyd-Wickizer, Powell, & Burman, 1998; Saunders, Aasland, de la Fuente, & Grant, 1993). The AUDIT scale scores also may be used to create three categories: “negative” screening, “positive” screening for hazardous drinking, and “probable alcohol dependence” (scores of 13 or greater for women, or 15 or greater for men).



While it is possible to convert a scale or continuous variable into a categorical variable, it is not possible to convert nominal or categorical data into scale or continuous variables. Consider a variable like age in years. If investigators collect data as the number of years in a person's age (or compute it as the current year minus the year of birth), they have a continuous scale variable for age. This can be converted into categories such as infant/toddler (0-2 years), preschool (3-4 years), early school age (4-10 years), preadolescent (11-12 years), adolescent (13-17 years), emerging adult (18-25 years), and so forth through adulthood and late life. However, if the investigators collected the data as ordinal categories, they cannot convert this reliably into a continuous variable since if a person is in the emerging adult category, there is no way of knowing if this means 18, 19, 20, 21, 22, 23, 24, or 25 years. We see the same issue with variables like income, duration of time (e.g., since symptoms appeared or disappeared), and scores on many measurement scales (e.g., depression, anxiety, pain, reoffending risk, conflict in interpersonal relationships, and many others). This is an important decision to make about study variables prior to collecting data

In order to refresh your skills in working with Excel and gain practice with the entering data for these different types of variables, we have an exercise in the Excel workbook to visit.

Interactive Excel Workbook Activities

Complete the following Workbook Activity:

- [SWK 3402.3-3.1 Data Entry Exercise](#)

Unit of Analysis Issues Revisited and Extended

The concept of **unit of analysis** was previously introduced in our first course—this it both a measurement and an analysis issue. When a study concerns the benefits of an individual-level intervention, it is relatively easy to identify the unit of analysis: data are collected about the participating individuals and analyzed at the individual level. The picture becomes a little more complicated when studying more complex systems. For example, imagine a social work BSW program wanting to investigate the impact of adding content designed to improve graduates' performance on the state licensure exam. The program might compare graduates' passing rates during the “pre” and “post” intervention (curriculum change) years. Data would be collected about each individual student's exam outcome, but each student appears only once in the study: when they took the licensure exam. On the surface, this looks like a cross-sectional study. However, the unit of analysis is not change for individual students—none of the students experienced both curriculum conditions. The unit of analysis is change for whole cohorts of

students, not change at the level of individual students. In other words, while students are the source of the pre-/post- data, they are not the unit of analysis. The unit of analysis is the program in this longitudinal comparison study (before and after the curriculum change).

Another social work example of the unit of analysis issue appears with policy evaluation. The data points might represent individuals, families, neighborhoods, or other local entities. But, the change related to the policy intervention is evaluated at the level of a county, region, state, or nation. That higher level is the unit of analysis, not the individuals who provided data which investigators aggregated at the more macro level. Imagine evaluating a new policy to reduce a community's rate at which individuals with serious mental disorders become incarcerated—the intervention involves a police/court/mental health partnership diverting these individuals to outpatient services for managing their mental disorders. Even though each person arrested provides the data used to measure the overall rate of incarceration (before and after the policy is implemented), the unit of analysis is change in the community-wide rate, not changes for individual persons.



We encounter this unit of analysis question in many systems, such as research about school-based interventions. Even if we collect data about individual students, the unit of analysis might be:

- teacher/classroom (aggregating across 25 or more students to learn about an intervention at a school),
- school (aggregating across classrooms to learn about an intervention in a school district),
- school district (to learn about an intervention with a state school system), or
- school systems (to learn about an intervention at the national level).

These may seem like subtle distinctions but understanding unit of analysis issues is important in terms of study design, measurement, and data analysis strategies.

The special case of measuring intervention fidelity.

The importance of adhering to an intervention protocol has been previously discussed. You learned that it is important to monitor how well an intervention's delivery adheres to the tested intervention protocol, and that this is assisted by staff training, detailed manuals, and routine process monitoring. Social work interventions at macro levels, such as policy changes, are no different in terms of the importance of monitoring fidelity or adherence in their implementation. The logic model for a policy-level intervention should have included the steps, processes, and output indicators associated with the policy. Monitoring these steps and processes is a form of fidelity assessment. For example, a federal policy to help address homelessness and families at risk of becoming homeless is known as the Section 8 Housing Choice Voucher Program. The policy was designed to provide resources to low-income families to secure housing in their local community's private market. The policy has numerous rules and guidelines concerning who is eligible to receive the vouchers, who can

administer voucher programs, and different types of vouchers. Assessing fidelity to the policy would require tracking how communities perform along each of these different dimensions. In addition, it would be wise to assess barriers to their implementation, factors such as natural disasters and economic trends that affect affordable housing availability. As survivors of the 2018 Camp Fire that decimated most of Paradise, California can attest, just because funds are available does not mean that housing is available.



It is important for intervention and evaluation investigators to consider how variables associated with fidelity to intervention at any level (micro, meso, macro, or global) will be measured. A search of the literature may or may not result in identifying a strong set of tools for measuring fidelity. This is an area of intervention research and measurement science that is not yet as well developed as others. Thus, investigators and evaluators may need to be creative in developing their strategies for tracking and measuring intervention fidelity.

These steps should assist in developing a fidelity tracking/measurement plan.

- Identify the critical components of the intervention that need to be assessed for fidelity. For example, this might mean specific content/information/topics being covered, screening or assessment being conducted, feedback being delivered, resources being shared, referrals being made, or other critical features.
- Identify key indicators for each critical component having been delivered. This might be in the form of client or practitioner checklists, review of case notes, supervisory review of a recorded session, or other strategies.
- Specify a sampling plan for tracking fidelity. For example, fidelity might be assessed for every contact with every participant, a random sample of contacts for each participant, or routinely for the first, middle, and final contact with a random sample of each practitioner's clients.
- Specify a measurement plan for tracking each key indicator.

Simple accounting measures.

An intervention protocol may call for a specific number, amount of time, frequency/pace, or duration over time of client contacts. For example, a batterer treatment program might require participants to attend a minimum of 24 2-hour group sessions and 8 1-hour individual counseling sessions during a 28-week period to meet court-ordered requirements. This type of information is relatively easy to track through administrative, scheduling, attendance, staff service, or billing data. These objective data, however, do not provide information concerning what transpired during those group or individual counseling sessions. More complex measures of fidelity are needed for this purpose.



Complex fidelity measures.

In many instances, the critical elements of an intervention involve how a session progressed and the qualities of the interactions, rapport, and transactions between the provider and client. In other words, determining not only THAT something transpired, but HOW it transpired. Consider, for example, interventions delivered through motivational interviewing (MI):

“Motivational interviewing (MI) is an empirically supported intervention for substance abuse and other behavioral problems. However, for this intervention to work, it must be provided with fidelity and skill” (Madson & Campbell, 2006, p. 67).

A systematic review of MI and motivational enhancement literature resulted in the identification of five measures for assessing practitioner fidelity to a psychotherapy or MI delivered intervention (Madson & Campbell, 2006). For example, the Motivational Interviewing Skill Code (MISC) requires a skilled rater to code

videotaped MI sessions between a practitioner and client. The MISC rater codes the following:

- the practitioner's behavior globally throughout the MI session, elements deemed critical to implementing MI (acceptance, egalitarianism, empathy, genuineness, warmth, and the "spirit" of MI)
- client behavior globally across the session (affect, cooperation, disclosure, and engagement)
- interactions globally between the practitioner and client across the session (level of collaboration and benefit of the interaction)
- each therapist utterance (affirming, confronting, open or closed questions, reflection types, or reframing)
- each client utterance (asking questions, neutral response, resisting change, or change talk)
- amount of time the practitioner and client engaged in talk during the session.

As you can imagine, a 45-minute session might require hours of time for coding using this fidelity measure. While this measure might help assess the practitioner's adherence (fidelity) to the MI intervention protocol, it is difficult and resource intensive (expensive) to use. Furthermore, it may be difficult to locate a rater qualified to evaluate MI sessions with strong reliability in this deconstructed manner. And, as a research measure used to assess fidelity, it is important to learn more about this measure's validity and reliability (psychometric) properties (Madson & Campbell, 2006). Let's review what this means for measurement in general.



Measurement Validity & Reliability Revisited and Extended

In our prior course (Module 3, chapter 5), we examined issues of **measurement validity** and **measurement reliability** in quantitative research. In our previous course, we discussed measurement validity and reliability as indicators of an instrument's measurement accuracy. Measurement validity and reliability are important to all quantitative measurement efforts and continue to be relevant in intervention and evaluation research. We have additional issues to address, as well: **measurement sensitivity**, **measurement specificity**, and the use of clinical screening and assessment tools intervention and evaluation research.

Measurement validity.

Measurement validity is an index of the extent to which a measurement tool adequately measures the concept, construct, or variable that it purports to measure (Grinnell & Unrau, 2014; Perron & Gillespie, 2015). If we believe that we are measuring a construct, validity is concerned with whether we are actually measuring that construct, factor, or variable. It concerns whether our measurement is on target, but also whether our measurement is robust enough to cover all or most of the dimensions of that construct. With validity, we need to determine whether we are measuring an underlying ability/capability/capacity or simply measuring a performance. For example, in a quiz on this course material, do you feel that your true ability to work with the material is reflected in your performance? There exist many reasons why a person might not perform at top capacity on any measure—situational, context, motivation, and other reasons.



An important consideration in measurement validity relates to cultural relevance and inclusiveness. In our earlier course we discussed the problem of using human intelligence tests with a gorilla (Koko); one criticism of traditional intelligence and mental

illness diagnostic measures concerns analogous cultural bias built into the instruments.

“Many tests have been objected to, as they produced poor results for the ethnic or racial minorities (students), as compared to the racial majorities. The problem lies not with the test-taker, but with the test itself”
(<https://psychologenie.com/understanding-cultural-bias-with-examples>).

An example presented in a test designed to demonstrate this point is the question: “What number comes next in the sequence, one, two, three, _____?” Those of us from majority culture would respond “four.” However, the “correct” answer in this test is “mong” (translated as “many”) because this is an “Original Australian Test of Intelligence” and the counting system in this culture can be translated as “one, two, three, many” (the word *mong* meaning numbers from 4 to 9 or 10, <http://www.wilderdom.com/personality/intelligenceOriginalAustralianAnswers.html>). While this example may seem extreme, analogous cultural biases appear in many measures used in quantitative research. It is important that this phenomenon of cultural bias be assessed with the tools investigators wish to use, at least in terms of the use with the populations investigators wish to study.



For example, investigators wished to use the Beck Hopelessness Scale in research with youth living in Kenya (Kagotho, Bowen, Ssewamala, Vaughn, & Kirkbride, 2018). Originally, the Beck Hopelessness Scale was developed in western culture, as part of the Beck Depression index (Beck, Weissman, Lester, & Trexler, 1974). The first challenge was to translate the instrument into the youths' national language (Kiswahili). To ensure adequacy of the translation, the translation was then translated back to English to ensure that the original and re-translated versions remained consistent. Next, the investigators collected data with 3,965 Kenyan students (aged 9-18 years) from 5 regions of the country. (When offered their choice, all the students opted to complete the English version.) Investigators analyzed the instrument's psychometric characteristics of the 20-item true/false question instrument using these data. They concluded that the Beck Hopelessness Scale remains a valid measure in use with this non-western population. Furthermore, the students' responses appeared not to systematically vary as a function of age group, meaning that it is

reasonably valid for both the younger children (aged 9-12 years) and the older youth aged 13-18 years).



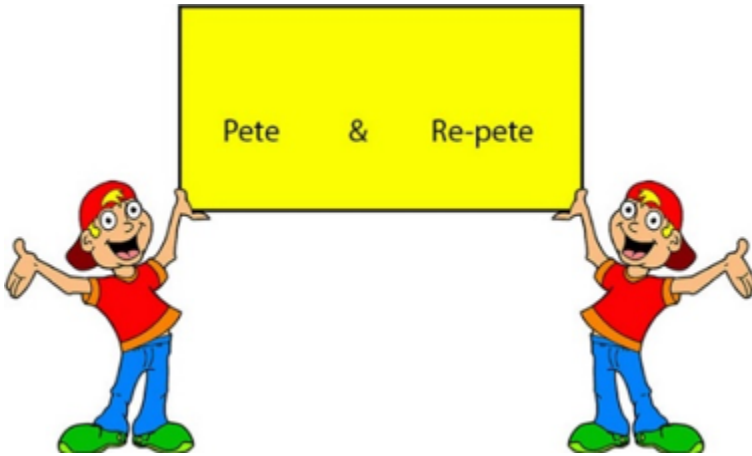
While these two examples (aboriginal Australia and sub-Saharan Kenya) represent differences between western and other cultures, differences exist within a country, state, city, or local population, as well—anywhere that human diversity exists. Consider, for example, the differences in candidates' yard signs you see in single neighborhoods at the time of an election. “Cultural” differences are relevant in measurement regarding many dimensions of human diversity beyond race, ethnicity, language, and national origin—for example, sex, sexuality, and gender identity; ability and disability; age; income and socio-economic status; religion and spirituality; and, the intersections among these dimensions.

Measurement reliability.

Measurement reliability is an index of the extent of instrument's stability and consistency in performing its measurements. Three aspects of measurement reliability are important to consider in research: test-retest, internal, and inter-rater reliability.

Test-retest reliability:

A measurement instrument's test-retest reliability is especially important in longitudinal intervention or evaluation research. The idea here is that the measure is consistent over time, when time is the only variable that is known to change. Using the instrument or measure at one time point should provide the same result as using it again at a different time point with the same individuals. Consider a yardstick—if you measure the height from the floor to your hip you expect to obtain the same result today, tomorrow, and the next day as long as all other conditions remain unchanged (same type of flooring, same footwear, and same point on your body). In social and behavioral research we aim for the same degree of consistency in measurement.



Let's examine an example that a social worker might use in studying an intervention designed to prevent or address intimate partner violence in dating relationships among adolescents (Wolfe, et al., 2001). The authors embarked on this effort to develop an age-appropriate measure, the Conflict in Adolescent Dating Relationships Inventory (CADI), because the instruments typically used with adults in relationships were not particularly relevant or valid in work with adolescent relationships. As part of a four-study series to assess the measure, they conducted a two-week test-retest reliability study with 70 students in 9th through 11th grades. This means that participants completed the measure at one time point and again two weeks later. The reliability over time was indicated in a correlation coefficient ($r=.68$ for abuse, $r=.75$ for restricted abuse). This is reasonably strong test-retest reliability (a positive correlation, far from 0 and closer to 1). Intervention and evaluation researchers rely on strong test-retest reliability in measures because it allows them to conclude that differences in scores between two time points can be attributed to the intervention rather than to inconsistency in the measurements.

Internal consistency:

Internal consistency is about getting similar responses to similar items on a multi-item measure, and is sometimes called internal reliability. This information conveys to investigators how well items that should be providing similar results do provide similar results. For example, if you wanted to measure client satisfaction with the services received, you would probably ask about satisfaction on several different dimensions of care, and use multiple items about each of those dimensions, rather than ask one single satisfaction question. Internal consistency would be present if the multiple satisfaction responses about one dimension were similar, even if satisfaction about different dimensions differed. In other words, how well items in a single group fit together based on how people responded to the measure.



Looking once again at the Conflict in Adolescent Dating Relationships Inventory (CADI), the investigators analyzed internal reliability of the measure's items on five subscales or groupings of items that should, in theory, have elicited similar responses from the study participants (Wolfe, et al., 2001). The statistic they reported is called Cronbach's alpha (using the same α symbol we have seen previously to designate alpha) which is low when close to 0 and high

when close to 1. The group of items related to verbal or emotional abuse were acceptably related ($\alpha=.82$), as were the items related to physical abuse ($\alpha=.83$); items for threatening behavior were not as internally consistent ($\alpha=.66$), relational aggression ($\alpha=.52$), or sexual abuse ($\alpha=.51$). An intervention investigator might decide to use this instrument to study verbal/emotional and physical abuse, but perhaps not the other forms of relationship aggression.

Inter-rater reliability:

The topic of inter-rater reliability has been discussed in our previous readings as inter-rater, inter-coder, and inter-observer reliability. While internal consistency is about the degree of consistency in the participants' responses, inter-rater reliability is about the degree of consistency in how responses are coded or rated by investigators observing participants' behavior. This construct is important in all forms of measurement where the data are formed from observations rather than collected from the target individuals themselves. For example, parents and teachers or fathers and mothers or grandparents and parents might provide different judgments of a child's behavior. The more similar their judgments, the greater their inter-rater or inter-observer reliability, and the more confidence investigators have in the ratings provided using that measure.



A variation on this theme was introduced in the study concerning the Conflict in Adolescent Dating Relationships Inventory (CADI): since this is a measure describing a relationship between two persons, the investigators also looked at the degree of partner agreement in rating their relationship (Wolfe, et al., 2001). They report that, in general, agreement between the male and female dating partners in describing their relationship was “fairly good” (p. 285) and that no significant differences were observed in the level of abuse partners reported in their relationships ($p > .05$ on a paired- t test, so they did not reject the null hypothesis of no difference).

Measurement sensitivity & specificity.

In assessing the measurement tools investigators might utilize, it is important to consider instrument precision. There are two sides to this precision issue: measurement sensitivity and measurement specificity. This discussion takes us back to prior discussions about

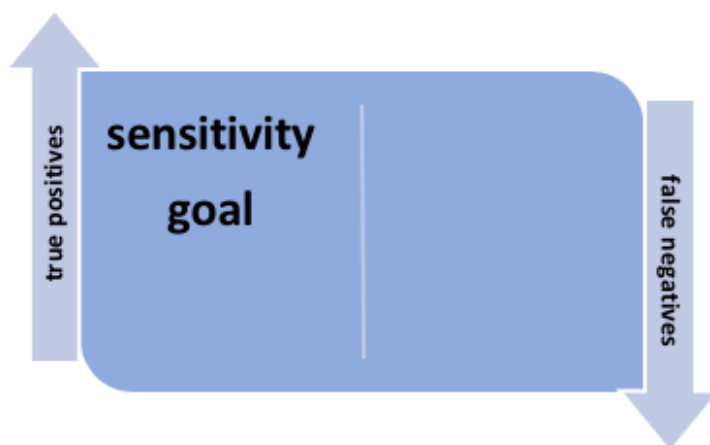
probability: the probability of properly classifying individuals or individual cases using a clinical tool. This situation concerns clinical tools that identify individuals or cases dichotomously: positive or negative for the condition of interest. For example, a tool might be designed to determine if individuals meet criteria for depression (positives) or fail to meet the criteria (negatives).

Sensitivity.

Measurement sensitivity concerns achieving the highest possible rate of accuracy in detecting the problem of interest. In other words, the percentage of individuals identified as meeting the criteria, classified as “positive,” would be as close as possible to 100% of those who actually have the condition or problem. In other words, the goal with sensitivity is to maximize identification of true positives and minimize the number of false negatives (those who are classified as not meeting criteria but who actually have the condition of interest). At-home pregnancy tests, for example, ideally report “positive” results and do not report “negative” results for women who are pregnant—true positives are identified and false negatives are minimal.

For example, investigators recently compared the sensitivity of two depression measures when used in mental health screening of adult Mexican American women (Valencia-Garcia, Bi, & Ayón, 2017). They found that one measure (the Kessler Psychological Distress Scale, or K-10) had a significantly higher degree of sensitivity than the other (depression questions on the Composite International Diagnostic Interview, Short Form, or MDD CIDI-SF) when the clinical “gold standard” of the Patient Health Questionnaire (PHQ9) was used as the comparison classification tool: 81.1% of those screened as positive for depression on the “gold standard” PHQ9 also were classified as positive on the K-10 compared to only 56.8% with the MDD CIDI-SF instrument. In other words, sensitivity of the

PHQ9 was deemed better than the MDD CIDI—true “positives” were better identified and “false negatives” were fewer.



Specificity.

On the other hand, it is equally important to make sure that the measurement instrument excludes as many individuals as possible that do not meet the criteria. In other words, the goal with specificity is to minimize the “false positive” (individuals without the condition being incorrectly classified as positive) and maximize the “true negative” rates (individuals without the condition being properly classified as negative). At-home pregnancy tests, for example, would hopefully show “negative” results and not show “positive” results for women who are not pregnant.

Let’s turn back to the example of the study comparing depression measures used with Mexican American women (Valencia-Garcia, Bi, & Ayón, 2017). The investigators found that the measure identified as being more sensitive (correctly identifying women with depression) also had significantly worse specificity: the K-10’s specificity was 78.6% and specificity of the MDD CIDI-SF was 88.7%. In other

words, the K-10 did not do as well as the other measure in identifying women who did not have depression based on the PHQ9 “gold standard.” The authors’ conclusion: the two measures appear to be complementary and perhaps both should be used when screening for depression in this population.



Measuring Change

A significant way in which intervention and evaluation research differ from the kinds of research we explored in our prior course revolves around what, exactly, is being measured. In any type of research, investigators might be interested in measuring outcomes related to a behavior, characteristic, or other dependent variable (y). In intervention and evaluation research, however, the focus is on measuring change in dependent (outcome) variables. This has important implications for the measurement plan.

Sensitivity to change.

First, investigators need to ensure that the measurement tools are sufficiently sensitive to detect even small changes. For example, imagine that a brief school-based program for students experiencing clinically diagnosable depression is being evaluated. Assuming the evaluators selected a depression measure that is categorical in nature: an individual is either scored as clinically depressed or not clinically depressed. At the brief program's conclusion, it may be that very few (if any) of the student participants seemed to have improved on the depression measure—moved from the clinically depressed to not-depressed category. However, if the evaluation team selected a scaled depression measure with 50 points ranging from “no depression” (0) to “severe depression” (50), it is quite possible that the investigators would observe progress in terms of a significant downward trend in the values of the depression scores for the program participants. One measure leads to the conclusion that the program was a failure and waste of resources, whereas the other measure leads to the conclusion that significant improvement is possible with the brief program, even if many students remained above the threshold for a clinical assessment of depression. The scaled measure is more nuanced and sensitive to change and may be a better choice than categorical measures to answer research questions associated with evaluation or intervention research.



Clinically or statistically significant change.

Second, investigators need to understand the difference between **statistically significant** change and **clinically significant** change. Going back to our example of the school-based depression program, the evaluation team might observe a statistically significant decline in depression scores among program participants: perhaps the mean score dropped from 44 to 42. While this change might have been statistically significant (leading investigators to reject the null hypothesis of no difference), was the change meaningful in a real-world or clinical way? Would a social work practitioner and the students, their parents, siblings, and friends/peers notice an average 2-point improvement? Hypothetically, it might require a 10-point difference on this depression measure before clinical significance or meaningfulness is achieved. Clinical significance gets at the issue of what is a relevant amount of change, beyond what is statistically significant. In terms of measurement, it is important to know how much

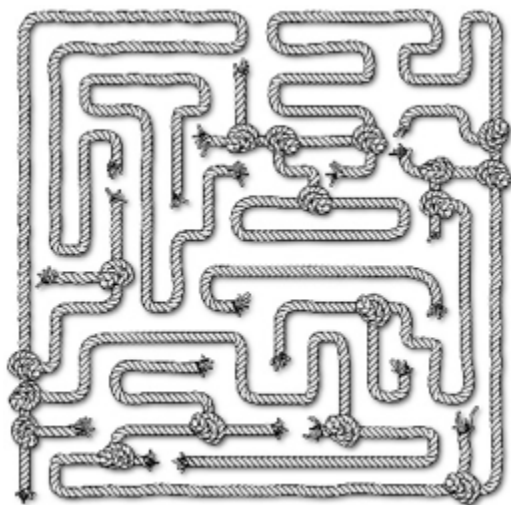
difference or change would be clinically meaningful with the instruments used.



Validity for measuring change.

Third, investigators need address the possibility that changes in scores on their measures are validly measuring change and not some other difference phenomenon. For example, imagine that investigators want to know whether an intervention helps improve problem solving abilities among aging persons without dementia. They might choose a measurement tool that is based on how fast a set of puzzles can be solved (a timed-trial). If they administer the very same puzzles before and after the intervention, they may see significant improvement in the participants' solution times the second time around. However, the improvement may not be a (sole) result of the intervention: it may be that participants figured out and learned the puzzles the first time and were able to remember how to solve them faster the second time. This problem is a test

recall effect (rather than measuring a real change in participants' capability), and the investigators might draw the wrong conclusions about the impact of their intervention based on these results.



Locating Data Sources

Before exploring sources for locating measurement tools, let's revisit the topic of sources of data. In research for understanding social work interventions, investigators have a number of options for locating available data (Kapp & Anderson, 2010). Working with available data eliminates or reduces the need for time and resource intensive data collection and can reduce participant burden for providing new data. **Secondary data** may not help answer an investigator's evaluation research questions since the questions are about a specific program or agency and the secondary data were likely collected elsewhere and may originally have been collected to answer a different research question. The secondary data may help

inform the local evaluation effort and measurement strategy, but local data is desired for answering local research questions.

On the other hand, **administrative data** is a form of available data commonly collected by the agencies, programs, and institutions being evaluated locally. While pre-existing, available data may seem like a gift to investigators, it is important to assess their usefulness. Investigators might consider three types of available and administrative data (Kapp & Anderson, 2010):

Systematically gathered administrative data.

Administrative processes generate a variety of forms of data that may be helpful in evaluating a program. For example, data concerning the number, frequency, duration, and timing/intervals of client contacts with the program and service delivery team members may be present in routine billing documentation. Certain demographic information might be available in client records, as well. For example, clients' ages can easily be calculated from their birth dates. At the organization level, records are maintained concerning staff delivering the interventions—their qualifications, percent effort directed to the intervention being evaluated, cost of delivering the services, and more.

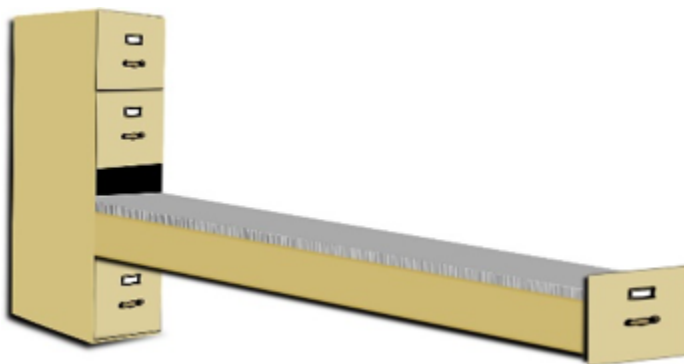


Treatment documentation.

Available data might also include information located in the records or case files for each client served. Individual records may include information in the form of screening and clinical assessment reports, case notes and treatment/service plan recorded by practitioners delivering services, case history details recorded in the client record over the course of intervention, critical incident reports, and other documentation about the intervention's process, fidelity, and outcomes. These sorts of data (often qualitative in nature) can be more difficult and time consuming to extract than most administrative data, however these data are often very rich and more accurate than retrospective data since they are recorded at the time of events happening (Kapp & Anderson, 2010). This type of information is typically located in client records because social work practitioners are trained to link intervention plans to assessment results and are encouraged to use evidence-supported assessment tools in this process (Jordan & Franklin, 2003).

Consider the possibility that investigators wish to understand client termination from an intervention program. It is difficult to obtain information from individuals who are no longer participating in an intervention, so client records may be a source of information about what preceded departure from the intervention: was it because their goals had been successfully achieved or were other factors suggestive of a pattern in client dropout? Recently, a social work student designed a study around this question, using client case records as the source of her data to determine if there existed patterns in dropout among young men in residential treatment for sexual offending behavior (Brenner, unpublished honors thesis project). She reviewed all 94 case records for past clients in the program, coding the information on completion status (successful, unsuccessful), level of family's involvement in the treatment process (low, moderate, high), scores on the intake Juvenile Sex offender Assessment Protocol (J-SOAP), adoption or foster care history (yes/

no), and victim type (sibling, friend, other). She found that low scores on the J-SOAP and high levels of parental involvement significantly predicted successful program completion.



Clinical or administrative judgments.

This category, as described by Kapp and Anderson (2010), is vague compared to the other types but has important implications for how evaluation results are interpreted. This type of information is closely tied to issues of intervention process, context, and fidelity. Evaluation investigators might record practitioner, staff, and administrator observations about the service delivery process. These observations might be presented or discussed in staff meetings or supervision sessions and may be reflected in formal or informal communications (e.g., internal agency memos, procedural guidelines, email conversations, and other artifacts). These types of information are particularly helpful in evaluating the implementation surrounding a new intervention or the modification of an existing intervention.



Working with secondary and administrative data has the previously noted drawbacks: investigators cannot be certain about data quality as its collection may not have been routinely monitored, and investigators are limited to what was being collected rather than being able to precisely tailor the data collection tools to the variables of interest. Thus, they may need to rely on **proxy variables** in their research efforts. For these reasons, intervention researchers and evaluators may wish to rely on **primary data** collection strategies instead or in addition. This leads us to a discussion of where one might turn to locate measurement tools for intervention and evaluation studies.

In order to refresh your memory of working with Excel, we have an activity in the Excel Workbook related to how variables appearing in an administrative data set might be transformed into variables suited to conducting an evaluation analysis.

Interactive Excel Workbook Activities

Complete the following Workbook Activity:

- [SWK 3402.3-3.2 Exercise in Transforming Variables](#)

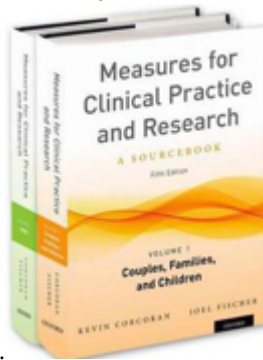
Locating Measurement Tools

In our prior course, you read about a few places where you might search for measurement tools that might be useful for intervention or evaluation research studies—particularly **screening and assessment tools** (Module 3, Chapter 5). First and foremost, you could turn to the literature. Surveying the articles published by other investigators on the topic, intervention, population, or variables of interest is a great place to start. For example, in our prior course you learned some electronic search terms that help investigators and practitioners locate pre-existing measures on a specific topic: tests, measures, measurements, assessment(s), scale(s), and screening.

In addition, you were introduced in our prior course to a variety of tools for locating and selecting general measurement tools. For example, the *Mental Measurements Yearbook* (Carlson, Geisinger, & Jonson, 2017). You also were introduced to several that are specific to measuring variables that might be of interest in efforts to understand social work interventions. These and others include:

- *Handbook of clinical rating scales and assessment in psychiatry and mental health* (Baer & Blais, 2010)
- *Handbook of family measurement techniques*, v. 1-3 (Touliatos, Perlmutter, Straus, & Holden, 2001)
- *Handbook of psychiatric measures*, 2nd ed. (Rush, First, & Blacker, 2008)
- *Handbook of research design and social measurement*, 6th ed. (Miller & Salkind, 2002)

- *Measures for clinical practice and research*, 5th ed, v. 1&2 (Corcoran & Fischer, 2013)
- *Positive psychological assessment: A handbook of models and measures* (Lopez & Snyder, 2003)
- *Rating scales in mental health* (Sajatovic & Ramirez, 2012)
- *Outcome measures for child welfare services: Theory and applications* (Magura & Moses, 1986)
- *Measuring violence-related attitudes, beliefs, and behaviors among youths: A compendium of assessment tools* (Dahlberg,



Toal, & Behrens, 1998).

Cautionary reminder.

In the earlier course you learned two important cautions related to the use of clinical screening and assessment tools as research measures.

- The need for procedures to provide their test results to each study participant.
- The need to determine how to work with the scores obtained using clinical measures—whether actual scores can be used as a scale variable or if scores can only determine the related to whether or not the person meets criteria for the condition being studied (dichotomous yes/no data).

Norms.

An important third consideration concerns the availability of standardized norms for the measure, particularly when applied with the population of interest. Normed scores provide a context for interpreting clinical screening and assessment scores by providing information about how others have scored on that measure (Jordan & Franklin, 2003), and are an important part of providing social work clients with personalized normed feedback about the behaviors being screened or assessed. For example, the following chart was used to provide personalized feedback to participants' responses concerning the number of drinks containing alcohol they consumed in an average week (Miller & Mattson, 2004). It offers men, women, practitioners, and researchers comparative values on this variable. A man who consumes 20 drinks per week is consuming more alcohol than 91% of other men in the United States; a woman who consumes 20 drinks per week is consumer more alcohol than 99% of other women. This information might be a big surprise to men and women who drink the same or less than their drinking buddies; it can help them reevaluate their own drinking behavior as being non-normative compared to the general population.

Table 3-1. Alcohol consumption norms for U.S. adults (based on 1995 National Alcohol Survey, adapted from Miller & Mattson, 2004)

| drinks per week | men | women | | drinks per week | men | women |
|-----------------|-----|-------|--|-----------------|-----|-------|
| 0 | 28% | 43% | | 14 | 87% | 98% |
| 1 | 54% | 77% | | 15 | 88% | 98% |
| 2 | 61% | 83% | | 16-17 | 89% | 98% |
| 3 | 68% | 88% | | 18-19 | 90% | 99% |
| 4 | 71% | 90% | | 20-21 | 91% | 99% |
| 5 | 73% | 92% | | 22-23 | 92% | 99% |
| 6 | 76% | 93% | | 24-26 | 93% | 99% |
| 7 | 77% | 94% | | 27-30 | 94% | 99% |
| 8 | 79% | 95% | | 31-36 | 95% | 99% |
| 9 | 80% | 96% | | 37-42 | 96% | 99% |
| 10 | 82% | 97% | | 43-49 | 97% | 99% |
| 11 | 84% | 97% | | 50-59 | 98% | 99% |
| 12 | 85% | 98% | | 60+ | 99% | 99% |
| 13 | 86% | 98% | | | | |

Normed scores also help determine clinical cutpoints for categorizing individuals as meeting or not meeting problem criteria. In other words, where to draw the line for categorizing their scores into the dichotomous “yes” or “no” groups. For example, the AUDIT (Alcohol Use Disorders Identification Test, Babor, Higgins-Biddle, Saunders, & Monteiro, 2010) has two different cut points based on population norms for screening individuals for a possible alcohol use disorder. Originally, the normed cutpoint for the AUDIT was a score of 8 or greater. More recently, authors suggested that the normed cutpoint for women should be a score of 7 or greater instead, and that the normed cutpoint of 8 remains suitable for men.



Chapter Conclusion

In this chapter you reviewed several important topics related to measurement and extended what you know about these topics as they apply to intervention and evaluation research. These included the relationship between measures and variables, types and levels of variables, unit of analysis issues, measurement validity and

reliability, and how to approach locating intervention and evaluation research measures. The expanded topics included instrument sensitivity and specificity, measuring intervention fidelity, measuring change (sensitivity to change and clinically significant change issues), considering the use of administrative data, and the importance of norms when clinical screening or assessment tools are used. In addition, you dusted off your skills in working with Excel on two data-related exercises: one concerning data entry for different variable types, the other about transforming variables that might appear in administrative data for use in an evaluation effort. The remaining topic in this module concerning design issues concerns study participants, the topic of our next chapter.

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=175>

Module 3 Chapter 4:

Participant Recruitment, Retention, and Sampling

Until this point, most of our discussions have treated intervention and evaluation research as being very similar. One major way in which they differ relates to participants and the pool or population from which they are selected. Recall that intervention research aims to draw conclusions and develop generalizations about a population based on what is learned with a representative sample. Thus, intervention research is strongest when the participants are systematically drawn from the population of interest. The aim of evaluation research is different: the knowledge gained is to be used to inform the practice and/or program being evaluated, not generalized to the broader population. As a result, evaluation research typically engages participants receiving those services. While the principles of systematic and random sampling might apply in both scenarios, the pool or population of potential participants is different, and the generalizability of results derived from the sample of participants differs, as well. The principles learned in our prior course about sampling and participant recruitment to understand social work problems and diverse populations applies to social work intervention and evaluation research for understanding interventions. Because much of evaluation and intervention research is longitudinal in nature, participant retention, as well as participant recruitment, is of major concern.

In this chapter you :

- review features of sample size and filling a study design, and learn how they apply to effect sizes and research for

- understanding social work interventions;
- review features of participant recruitment and retention, and learn how they apply to research for understanding social work interventions;
- learn about random assignment of participants to study design conditions in intervention and evaluation research.

Sample Size Reviewed & Expanded

Sample size is not a significant issue if interventions are being evaluated from a qualitative approach where the aim is depth of data rather than generalizability from a sample to a population. Sample size in qualitative studies is generally kept relatively small as means of keeping manageable the volume of data needing to be analyzed.

Sample size does matter in quantitative approaches where investigators will generalize from the sample to a population. In our prior course you learned how sample size matters in terms of the sample's ability to represent the population. Remember the green M&Ms example where the small samples were quite varied compared to each other and to the true population, but the larger (combined) samples were less different? Sample size issues remain important in intervention research where generalizations are to be made to the population based on the sample. This might be an issue, as well, in evaluation research where there are many participants involved in the intervention being evaluated and the investigators choose to work with data from a sample rather than participants representing the entire population served. In either case, intervention or evaluation research, investigators need to determine what constitutes an adequately sized sample. Two issues need to be addressed: numbers needed to fulfill the requirements of a study design and sample size needed to detect meaningful effects.

Filling a quantitative study design:

You may recall from our prior course how a study design relates to the number of study participants that need to be recruited (and retained). The study design might include two or more independent variables (the ones being manipulated or compared). To ensure sufficient numbers of participants for analyzing these variables, investigators need to be sure that participants of the designated types are recruited and retained so that their outcome (dependent variable) data can be analyzed. Here is an example of the numbers of each type needed to fulfill a 2 X 3 design. This example has neighborhoods as the unit of analysis; individual participants are embedded within those neighborhood units. This example is relevant to research for understanding social work interventions at a meso or macro level.

Imagine a study concerning the impact of a community empowerment intervention designed to help members of local communities improve health outcomes by reducing exposure to air and water environmental toxins and contaminants inside and outside of their homes. Investigators are concerned that the intervention might differently impact very low-income, low-income, and moderate-income neighborhoods. They have chosen to conduct a random assignment study where ½ of the neighborhoods receive the intervention immediately and the other ½ receive it one year later (delayed intervention with the no intervention period serving as the control). They have determined that for the purposes of their analysis plan, they need a minimum of 12 neighborhoods in each condition. The sampling design would look like this:

| | | Neighborhood Income Status | | |
|--------------|-----------|----------------------------|------------|-----------------|
| | | very low income | low income | moderate income |
| Intervention | immediate | 12 | 12 | 12 |
| Group | Delayed | 12 | 12 | 12 |

Filling the study design cells for this 2 X 3 design requires a minimum of 72 neighborhoods (6 cells times 12 units each=72 units total). These would be recruited as: 24 very low income, 24 low income, and 24 moderate income neighborhoods. Within each neighborhood, they hope to engage 15-20 households, meaning that they will engage with between 1080 and 1440 households ($15 \times 72=1080$, $20 \times 72=1440$).

Sample size related to effect size.

Previously in this chapter you read about differences that are clinically meaningful. Intervention researchers are often asked to consider an analogous problem: what is the size of the effect detected in relation to the intervention? While an observed difference might be statistically significant, it is important to know whether the size of that difference is meaningful. **Effect size** information helps interpret statistical findings related to interventions—their power to effect meaningful amounts of change in the desired outcomes. The size or magnitude of the effect detected is determined statistically, and sample size is one part of the formula for computing effect sizes. As a result, the size of a study's sample has an impact on the size of effect that can be detected.

Here, the logic can sometimes become a bit confusing. This diagram helps explain the relationship between effect size and sample size without getting into the detailed statistics involved.

| | | Effect Size | |
|-------------|-------|---|---|
| | | small | large |
| Sample Size | small | <ul style="list-style-type: none"> intervention may have had meaningful impact that could not be detected due to small sample size (Type II error if discarding the intervention as a result) intervention may have no meaningful impact (correct to discard the intervention) which is true would need to be confirmed with larger sample | <ul style="list-style-type: none"> intervention had a meaningful impact—it was detected even with a small sample size (correct to preserve the intervention as a result) |
| | large | <ul style="list-style-type: none"> intervention had little or no meaningful impact—effect not detected even with a large sample size (correct to discard the intervention as a result, preserving the intervention would be a Type I error) | <ul style="list-style-type: none"> intervention had a meaningful impact (correct to preserve the intervention as a result) larger sample allows detection of even small effects |

In other words, if an investigator wishes to be as sure as possible to detect an effect if it exists, a larger sample size will help; having a small sample leaves the question unanswered if no effect is detected (see the small/small peach colored box)—the study will have to be repeated to determine if there really is no effect (see the large/large pink colored box) or there actually is an effect of the intervention—large or small.

In order to refresh your skills in working with Excel and gain practice with the topic of sample size related to effect size, we have an exercise in the Excel workbook to visit.

Interactive Excel Workbook Activities

Complete the following Workbook Activity:

- [SWK 3402.3-4.1 Exercise on Sample and Effect Sizes](#)

Participant diversity.

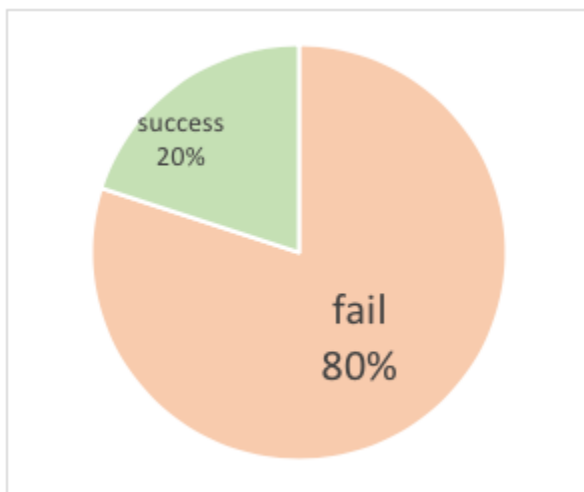
In our prior course we also examined issues related to participant diversity and heterogeneity in study samples. Intervention and evaluation research working with samples need to consider the extent to which those samples are representative of the diversity and heterogeneity present in the population to which the intervention research will be generalized or the population of those served by the program being evaluated. Ideally, the strategies for **random selection** that you learned in our prior course (probability selection) aid in the effort to achieve representativeness. Convenience sampling and snow-ball sampling strategies, as you may recall, tend to tap into **homophily** and create overly homogeneous samples.

Diversity among participants in qualitative studies is not intended to be representative of the diversity occurring in the population. Instead, heterogeneity among participants is intended to provide breadth in the perspectives shared, as a complement to depth in the data collected.

Participant Recruitment & Retention Reviewed & Expanded

In our prior course you learned about the distinction between **participant recruitment** (individuals entering into a study) and

participant retention (individuals remaining engaged with a longitudinal study over time). Intervention and evaluation research investigators need to have a strong recruitment and retention plan in place to ensure that the study can be successfully completed. Without the right numbers and types of study participants, even the best designed studies are doomed to failure: over 80% of clinical trials in medicine fail due to under-enrollment of qualified participants (Thomson CenterWatch, 2006)! How suitably this figure represents what happens in behavioral and social work intervention research is unknown (Begun, Berger, & Otto-Salaj, 2018). In fact, several authors have recommended having study personnel assigned specifically to the tasks of implementing a detailed participant recruitment and retention plan over the lifetime of the intervention or evaluation study (Begun, Berger, & Otto-Salaj, 2018; Jimenez & Czaja, 2016).



Here is a brief review of some of those principles and elaboration of several points relevant to intervention and evaluation research.

Participant recruitment.

Investigators are notorious for over-estimating the numbers of participants that are available, eligible, and willing to participate in their studies, particularly intervention studies (Thoma et al., 2010). This principle has been nicknamed Lasagna's Law after the scientist Louis Lasaj who first described this phenomenon:

“Investigators all too often commit the error of (grossly) overestimating the pool of potential study participants who meet a study's inclusion criteria and the number they can successfully recruit into the study” (Begun, Berger, & Otto-Salaj, p. 10).



You may recall learning in our prior course about a 3-step process related to participant recruitment (adapted from Begun, Berger, & Otto-Salaj, 2018): generating contacts, screening, and consent.

Generating contacts.

The first step involved generating initial contacts, soliciting interest from potential participants. Important considerations included the

media applied (e.g., newsletter announcements, radio and television advertising, mail, email, social media, flyers, posters, and others) and the nature of the message inviting participation. Remember that recruitment messages are invitations to become a participant and need to respond to:

- why someone might wish to engage in the study
- need for details to make an informed choice about volunteering
- how they can become involved
- cultural relevance of the invitation message.

One strong motivation for individuals to participate in intervention research is the potential for receiving a new form of intervention, one to which they might not otherwise have access. This might be particularly motivating for someone who has been dissatisfied with other intervention options. The experimental option might seem desirable because it is different from something that has not worked well for them in the past, it seems more relevant, it is more practical/feasible than other options, or they were not good candidates for other options.

On the other hand, one barrier to participation is the experimental, unknown nature of the intervention—the need to conduct the study suggests that the outcomes are somewhat uncertain, and this may include unknown side effects. This barrier might easily outweigh the influence of a different motivation: the altruistic desire to contribute to something that might help others. Altruism is sometimes a motivation to engage in research but cannot be relied on alone to motivate participation in studies with long-term or intensive commitments of time and effort. It may be sufficient to motivate participation in an evaluation for interventions that are part of their service or treatment plan, however.



Screening.

Screening is part of the intervention research recruitment process. Intervention and evaluation research typically require participants to meet specific study criteria—**inclusion criteria** and **exclusion criteria**. These criteria might relate to their condition or problem, their past or present involvement in services or treatment, or specific demographic criteria (e.g., age, ethnicity, income level, gender identity, sexual orientation, or others). For example, inclusion criteria for a study might specify including only persons meeting the DSM-5 diagnostic criteria for a substance use disorder. Exclusion criteria for that same study might specify excluding all persons with additional DSM-5 diagnoses such as schizophrenia or dementia unrelated to substance use and withdrawal.

Study investigators need to establish clear and consistent screening protocols for determining who meets inclusion/exclusion criteria for participation. Screening might include answers to simple questions, such as “Are you over the age of 18?” or “Are you currently pregnant?” Screening might also include administering one or more standardized screening instruments, such as the Alcohol Use Disorder Identification Test (AUDIT), the Patient Health Questionnaire screening for depression, the mini mental state examination (MMSE) screening for possible dementia,

or the Hurt, Insult, Threaten, Scream (HITS) screening for intimate partner violence.

Regardless of the tool, screening information is not data since screening occurs prior to participant consent. The sole purpose of screening is to determine whether a potential volunteer is eligible to become a study participant. Ethically, screening protocols also need to include a strategy for referring persons who made an effort to participate in the intervention (expressed a need for intervention) but could not meet the inclusion criteria. In other words, if the door into the study is closed to them, alternatives need to be provided.



Consent.

Informed consent is the third phase of the recruitment process. You learned in our prior work what is required for informed consent. For intervention research with a generalizability aim the consent process should be reviewed by an Institutional Review Board (IRB). Evaluation research, on the other hand, that is to be used primarily to inform the practitioner, program, agency, or institution might not require IRB review. However, the agency should secure consent to

participate in the evaluation, particularly if any activities involved fall outside of routine practice and record-keeping.

Important to keep in mind throughout the three phases of recruitment is that all three phases and what transpires during intervention relate to participant retention over the course of a longitudinal study. While it is critical during recruitment to consider, from the participants' point-of-view, why they might wish to become involved in an intervention or evaluation study, in longitudinal studies it is equally important to consider why they might wish to continue to be involved over time. This topic warrants further attention.



Participant retention.

A great deal of resources and effort devoted to participant recruitment and delivering interventions is wasted each time a study participant drops out before the end of a study (called study attrition, this is the opposite of retention). Furthermore, the integrity of study conclusions can be jeopardized when study attrition occurs. An interesting meta-analysis was conducted to assess the potential impact on longitudinal studies of our nation's high rates of incarceration, especially in light of extreme racial disparities in incarceration rates (Wang, Aminawung, Wildeman,

Ross, & Krumholz, 2014). The investigators combined the samples from 14 studies into a complete sample of 72,000 study participants. Based on U.S. incarceration rates, they determined that longitudinal studies stand to lose up to 65% of black men from their samples. Under these conditions, study results and generalizability conclusions are potentially seriously impaired, especially since participant attrition is not occurring in a random fashion equivalently across all groups.



Relationship & Retention.

As previously noted, a major factor influencing participants' willingness to remain engaged with an intervention or evaluation study over time is the nature of their experiences with the study. An important consideration for intervention and evaluation researchers is how they might reduce or eliminate barriers and inconveniences associated with participation in their studies (Jimenez & Czaja, 2016). These might include transportation, time, schedule, child care, and other practical concerns. Another factor influencing potential participants' decisions to commit to engaging

with a study concerns stigma—the extent to which they are comfortable becoming identified as a member of the group being served. Consider, for example, the potential stigma associated with being diagnosed with a mental illness, identified as a victim of sexual assault, or categorized as “poor,” or labeled with a criminal record. Strategies for minimizing or eliminating the stigma associated with participation in deficit-defined study and emphasizing the strengths base would go a long way toward encouraging participation. For example, recruiting persons concerned about their own substance use patterns is very different from recruiting “addicts” (see Begun, 2016).

Before launching a new program or extending an existing program to a new population, social workers might solicit qualitative responses from potential participants, to determine how planned elements are likely to be experienced by future participants. This may be performed as a preliminary focus group session where the group provides feedback and insight concerning elements of the planned intervention. Or, it may be conducted as a series of interviews or open-ended surveys with representatives of the population expected to be engaged in the intervention. Similarly, investigators sometimes conduct these preliminary studies with potential participants concerning the planned research activities, not only the planned intervention elements.

Two examples of focus groups assisting in the planning or evaluation of interventions come from Milwaukee County. HEART to HEART was a community-based intervention research project designed to reduce the risk of HIV exposure among women at risk of exposure by virtue of their involvement in risky sexual and/or substance use behavior. Women were to be randomly assigned to a preventive intervention protocol (combined brief HIV and alcohol misuse counseling) or a “control” condition (educational information provided about risk behaviors). Focus group members helped plan the name of the program, its identity and branding, many of the program elements, and the research procedures to ensure that it was culturally responsive, appropriate, and welcoming. Features

such as conducting the work in a non-stigmatizing environment (a general wellness setting rather than a treatment center), creating a welcoming environment, gender and ethnic relevant materials, and providing healthful snacks were considered strong contributors to the women's ongoing participation in the longitudinal study (Begun, Berger, & Otto-Salaj, 2018).

In a different intervention research project, a focus group was conducted with partners of men engaged in a batterer treatment program. The purpose of the focus group was to develop procedures for safely collecting evaluation data from and providing research incentive payments to the women at risk of intimate partner violence. The planned research concerned the women's perceptions of their partners' readiness to change the violent behavior, and investigators were concerned that some partners might respond abusively to a woman's involvement in such a study. The women helped develop protocols for the research team to follow in communicating safely with future study participants and for materials future participants could use in safely managing their study participation (Begun, Berger, & Otto-Salaj, 2018).



Random Assignment Issues

First, it is essential to remember that random selection into a sample is very different from the process of randomization or random assignment to experimental conditions.

Random selection refers to the way investigators generate a study sample that is reflective and representative of the larger population to which the study results are going to be generalized (external validity)...**Random assignment**, often colloquially called **randomization**, has a different goal and is used at a different point in the intervention research process. Once we have begun to randomly select our participants, our study design might call for us to assign these recruited individuals to experience different intervention conditions” (Begun, Berger, & Otto-Salaj, 2018, p. 17-18).

Several study designs examined in Chapter 3 of this module involved random assignment of participants to one or another experimental condition. The purpose of random assignment is to improve the ability to attribute any observed group differences in the outcome data to the groups rather than to pre-existing differences among group members (internal validity). For example, if we were comparing individuals who received a novel intervention with those who received a treatment as usual (TAU) condition we would be in trouble if there happened to be more women in the novel treatment group and more men in the TAU group. We would not know if differences observed at the end were attributable to the intervention or if they were a function of gender instead.

Consider, for example, the random controlled trial (RCT) design from an intervention study to prevent childhood bullying (Jenson et al, 2010). A total of 28 elementary schools participated in this study, with 14 having been randomly assigned to the experimental condition (the new intervention) and 14 to the no-treatment control group. This allowed investigators to compare outcomes of the two

treatment conditions with considerable confidence that the observed differences were attributable to the intervention; however, they were somewhat unlucky in their randomization effort since a greater percentage of children in the experimental condition were Latino/Latina than in the control condition. This ethnicity factor needs to be taken into consideration in conclusions about the observed significant reduction in bully victimization among students in the experimental schools compared to the control group.



Random assignment success & failure.

Randomly assigning participants to different experimental or intervention conditions requires investigators to introduce chance to the process. Randomness means a lack of systematic assignment. So, if you were to alternately assign participants to one condition or the other based alternating how each enrolled for the study

a certain degree of chance is invoked: persons 1, 3, 5, 7 and so forth=control group, persons 2, 4, 6, 8 and so forth=experimental group. This system is only good if there is nothing systematic about how they were accepted into the study—nothing alphabetical or gendered or otherwise nonrandom. Systems of chance include lottery, roll of the dice, playing card draws, or use of a random numbers table—the same kinds of systems you read about in our prior course when we discussed how individuals might be randomly selected for participation in the sample. What could possibly go wrong?



Unfortunately, relying on chance does mean that random assignment (randomization) may fail to result in a balanced distribution of participants based on their characteristics even if the size of different assigned groups is even. This unfortunate luck was evident in the Jenson et al (2010) study previously mentioned where the distribution of Latino/Latina students was disproportionate in the two intervention condition groups. However, those investigators were only unlucky on this one dimension—there was reasonable comparability on a host of other variables.

Another way that random assignment sometimes goes wrong is through failure to stick to the rules of the randomization plan. Perhaps a practitioner really wants a particular client to experience the novel intervention (or the client threatens to participate only if assigned to that group). Suppose the practitioner somehow manipulates the individual's assignment with the intent to replace

that person with another individual so the numbers assigned to each group even out. Unfortunately, the result is reduced integrity of the overall study design—those individuals' assignments not being random means that systematic assignment has crept into the study, jeopardizing study conclusions. “Randomization ensures that each patient has an equal chance of receiving any of the treatments under study” and generates comparable groups “which are alike in all the important aspects” with the notable exception of which intervention the groups receive (Suresh, 2011, p. 8). In reality, investigators, practitioners, and participants may be tempted to “cheat” chance to achieve a hoped-for assignment.



Assessing randomization results.

Investigators need to determine the degree to which their random assignment or randomization efforts were successful in creating equivalent groups. To do this they often turn to the kinds of statistical analyses you learned about in our prior course: chi-square, independent samples t-test, or analysis of variance (Anova), depending on the nature of the variables involved. The major difference here, compared to the analyses we previously practiced,

is what an investigator hopes the result of the analysis will be. Here is the logic explained:

1. the null hypothesis (H_0) is no difference exists between the groups.
2. if the groups are equivalent, the investigator would find no difference.
3. the investigator hopes not to find a difference—this does not guarantee that the groups are the same, only that no difference was observed.

To refresh your memory of how to work with these three types of analyses and to make them relevant to the question of how well randomization worked, we have three exercises in our Excel workbook.

Interactive Excel Workbook Activities

Complete the following Workbook Activities:

- [SWK 3402.3-4.2 Exercise Testing Randomization with Chi-Square Analysis](#)
- [SWK 3402.3-4.3 Randomization Check: Independent Samples t-Test](#)
- [Workbook SWK 3402.3-4.4 Randomization Check: Anova](#)

Chapter Summary

In this chapter you reviewed several concepts related to samples, sampling, and participant recruitment explored in our prior course. The sample size topic was expanded to address how sample size relates to effect size in intervention and evaluation research. Issues related to participant recruitment were reviewed, particularly as they relate to the need for engaging a diverse and representative sample of study participants and how these concerns relate to a study's external validity. This topic was expanded into a 3-phase model of recruitment processes: generating contacts, screening volunteers for eligibility, and consenting participants. You then read about issues concerning participant retention over time in longitudinal intervention and evaluation studies, especially the importance of participants' experiences and relationships with the study. This included a discussion of participants' experiences with random assignment to study conditions, depending on the study design, and how randomization might or might not work. You learned a bit about how to assess the adequacy of the randomization effort in our Excel exercises and to think about how randomization successes and failures might affect a studies integrity and internal validity.

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=176>

Module 3 Summary

Throughout this module emphasis was placed on the kinds of study design and implementation features that need to be addressed in developing or critiquing an intervention or evaluation research study. Topics related to study approach, study design, measurement, and participants were reviewed from what you learned in our prior course and expanded in terms of their relevance to research for understanding social work interventions. Planning for and assessing intervention fidelity was a major new theme in this module. Sources of data, data collection strategies, and ten strategies for quantitative designs in intervention and evaluation research were other major new areas of content. Mastering the contents of this module prepares you for the next module, how data from intervention and evaluation research might be analyzed.

Module 3 Key Terms and Definitions

administrative data: information routinely collected by agencies, organizations, or institutions that might be appropriate sources for analysis in evaluation research.

case study: detailed, qualitative description of a single case (or very small number of cases), including initial assessment, description of the intervention(s) applied, and observed outcomes.

clinically significant: as opposed to statistical significance, relates to the practical importance of a finding or observed difference in data.

community-based participatory action research (CBPR): an approach to research formed out of a collaborative partnership between community-based members and individuals with research expertise to meet information needs of the various partners.

control group: a comparison condition in experimental study designs where participants do not receive the novel or experimental conditions delivered to the experimental group.

cost-related evaluation: research designed to answer questions concerning an intervention's (or program's) efficiency, particularly in relation to the extent of impact it has on the target problem.

cross-sectional study: data collection at one time point with a single unit of study (individuals, couples, families, or other elements); does not require participant retention over time.

DALYs (disability-adjusted life years): units designed to indicate “disease burden,” calculated to represent the number of years lost to illness, disability, or premature death (morbidity and mortality), often used as an outcome indicator in evaluating medical interventions.

dependent variables: the variable(s) presumed to vary as a function

of changes in the independent variable; sometimes called the outcome variable(s).

double-blind study: an intervention study approach where both the study participants and those delivering the intervention remain unaware about the group to which any participant has been assigned until the study's conclusion (the "unblinding" phase).

effect size: a statistical means of quantifying the magnitude or size of a difference between groups or time points (in a longitudinal study) being compared.

exclusion criteria: standards applied in screening potential study participants where good fit with the sample criteria leads to an "ineligible to participate" decision.

experimental designs: a research study approach in which at least one variable is intentionally varied or manipulated (independent variables) and the influence of other variables is controlled to maximize the investigator's ability to identify the effects of the manipulated variable on one or more outcome variables (dependent variables).

external validity: the extent to which conclusions based on observations about a sample can appropriately be generalized to a population or to other situations.

follow-up: collecting data to answer questions about durability of intervention effects over time following completion of intervention (as compared to immediate post-intervention data collection)

formative evaluation: evaluation designed to address intervention planning questions, such as feasibility or needs assessment (as opposed to process or summative evaluation).

homophily: the tendency for similar individuals to aggregate or associate together, separately from individuals who are different.

inclusion criteria: standards applied in screening potential study participants where good fit with the sample criteria leads to an "eligible for participation" decision.

independent variables: the variable(s) intentionally varied or

manipulated in an experimental design to determine their effect on an outcome (dependent variable).

internal validity: the extent to which potentially confounding factors are controlled in an experimental study, enhancing confidence in the main study results concerning the impact of the studied variables.

intervention fidelity: the degree to which implementation of an intervention replicates the original, studied intervention protocol with integrity.

longitudinal study: data collection at two (or more) time points with the same units of study (individuals, couples, families, or other element); requires participant retention over time.

manualized intervention: one strategy to enhance fidelity and integrity in intervention implementation, involving the development and dissemination of detailed logic model and implementation guidelines.

maturation: when change occurs as a result of naturally occurring developmental processes rather than resulting from application of an intervention intended to produce change.

measurement inconsistency: the degree to which a measurement tool fails to consistently measure a construct and measurement error is introduced from individual differences in interpretation of the measurement tool rather than actual differences in events or experiences presumably being measured.

measurement reliability: indicates the degree of accuracy or precision in measuring the construct or variable of interest.

measurement sensitivity: the rate of accuracy a measurement tool has in detecting the problem of interest; the tool accurately identifies individuals meeting the criteria (“positives”) and does not miss those who should be classified as “positive” (few false negatives).

measurement specificity: the rate of accuracy a measurement tool has in not erroneously identifying individuals as “positives” when they should not be (false positive rate low) and correctly identifying individuals as “negatives” when they should be.

measurement validity: the extent to which a measurement tool or instrument adequately measures the concept, construct, or variable of interest, comprised of several types of validity (construct, concurrent, predictive, and others).

naturalistic observation: collecting data about behavior occurring in its natural environment and context, where the observer is non-participant (compared to laboratory, contrived, or controlled circumstances and to participatory action research).

non-treatment control: a study design where the comparison group receives no form of intervention in contrast to the experimental group.

outcome or impact evaluation: evaluation designed to answer questions about the effects of an intervention (see summative evaluation).

participant recruitment: the process of engaging participants in a study.

participant retention: process of keeping participants engaged in a study.

participatory action research (PAR): an approach to intervention or evaluation research where investigators are involved in both observational aspects of research and as agents of change (action oriented); understanding comes from changing the situation and observing the impact of the change efforts.

participatory observation: a form of naturalistic or semi-naturalistic observation whereby the investigator is or becomes an integrated member of the group being studied.

participatory research: a set of approaches to research where participants direct the activities for change and for investigation in collaboration with investigators.

placebo effect: an effect produced by exposure to a neutral “treatment” where the effects (positive/beneficial or negative/harmful) cannot reasonably be attributed to its characteristics, but instead to the experience of having some intervention delivered rather than none.

post-only design: data collection at one point in time after the intervention.

pre-experimental designs: study designs that lack control groups, suffering reduced internal validity as a result.

pre-/post-design: data collection at two comparison time points are before the intervention and immediately after the intervention.

pre-/post-/follow-up design: data collection at three comparison time points are before the intervention, immediately after the intervention, and at a significant point following the intervention.

primary data: information collected for the specific purposes to which it is used in a research study, tailored to the study's selected aims, design, and variables.

process evaluation: evaluation research designed to answer questions about how an intervention was implemented (as differentiated from intervention outcomes, see summative evaluation).

proxy variables: a variable that serves in place of another variable that could not be or was not directly measured itself; proxy variables should have a close association (correlation) with the variables they represent.

quasi-experimental designs: research designs that include comparison (control) groups but lack random assignment to those groups.

random assignment: elements in an experimental study are assigned to study conditions in such a manner (randomly) that potential sources of group membership bias are minimized or eliminated; also called randomization.

random control trial (RCT): an experimental study design where participants are randomly assigned to different conditions, such as the novel experimental or control groups (may be non-treatment, placebo, TAU or a different intervention).

randomization: see random assignment.

random selection: elements (participants) in a study are selected from the population in a manner (randomly) that maximizes the

sample's representativeness of (generalizability to) the study's target population and minimizes bias in the final selected sample.

screening and assessment tools: instruments used in practice to either identify persons who are at risk or possibly experience the problem of interest (screening) or to determine the extent/severity of the problem identified (assessment).

secondary data: research data originally gathered to meet one study's aims, design, and demands, but can be re-analyzed as a separate study to meet a different study's aims without having to collect new primary data.

selection bias: the risk of non-representative results due to a non-representative sample.

spontaneous or natural change: changes that occur without or outside of intentional intervention efforts.

statistically significant: determination that an observed relationship between variables exists beyond what might be expected by chance.

summative evaluation: evaluation research designed to answer questions about the effects, impact, or outcomes of an intervention (as differentiated from intervention implementation, see process evaluation).

treatment as usual condition (TAU): an experimental design where the control group receives whatever is the usual and customary form of intervention compared to the experimental group receiving the novel intervention being tested.

Module 3 References

- Babor, T.F., Higgins-Biddle, J.C., Saunders, J.B., & Monteiro, M.G. (2010). *The Alcohol Use Disorders Identification Test (AUDIT): Guidelines for use in primary care*, 2nd ed. Geneva: The World Health Organization (WHO), Department of Mental Health and Substance Dependence.
- Baer, L., & Blais, M.A. (2010). *Handbook of clinical rating scales and assessment in psychiatry and mental health*. NY: Humana Press.
- Beck, A.T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The hopelessness scale. *Journal of Consulting and Clinical Psychology*, 42(6), 861-865.
- Begun, A.L. (2016). Consider the language we use: Well worth the effort. *Journal of Social Work Practice in the Addictions*, 16(3), 332-336.
- Begun, A.L., Berger, L.K., & Otto-Salaj, L.L. (2018). *Participant recruitment and retention in intervention and evaluation research*. NY: Oxford University Press, Pocket Guides to Social Work Research.
- Begun, A.L., Early, T.J., & Hodge, A.I. (2016). Mental health and substance abuse service engagement by men and women during community reentry following incarceration. *Administration & Policy in Mental Health & Mental Health Services Research*, 43(2), 207-219.
- Begun, A.L., Rose, S.J., & LeBel, T. (2011). Intervening with women in jail around alcohol and substance abuse during preparation for community reentry. *Alcoholism Treatment Quarterly*, 29(4), 453-478.
- Berkowitz, J. M., Huhman, M., Heitzler, C. D., Potter, L.D., Nolin, M.J., & Banspach, S.W. (2008). Overview of formative, process, and outcome evaluation methods used in the VERBTM campaign. *American Journal of Preventive Medicine*, 34(6S), S222-S229.
- Bradley, K.A., Boyd-Wickizer, J., Powell, S.H., & Burman, M.L. (1998).

- Alcohol screening questionnaires in women: A critical review. *JAMA*, 280(2), 166-171.
- Brenner, M. (unpublished honors thesis report). *Variables for predicting successful completion of a treatment program for juvenile males who have committed sex offenses*. Columbus, OH: The Ohio State University. Retrieved from https://kb.osu.edu/bitstream/handle/1811/84919/1/MorganBrenner_HonorsThesis.pdf
- Carlson, J.F., Geisinger, K.F., & Jonson, J.L. (Eds.). (2017). *The twentieth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.
- Cherry, K. (2018). How the placebo effect works in psychology. Verywellmind. Retrieved from <https://www.verywellmind.com/what-is-the-placebo-effect-2795466>
- Corcoran, K., & Fischer, J. (2013). *Measures for clinical practice and research*, 5th ed., v. 1-2. NY: Oxford University Press.
- Dabelko-Schoeny, H., & King, S. (2010). In their own words: Participants' perceptions of the impact of adult day services. *Journal of Gerontological Social Work*, 53(2), 176-192.
- Dabelko-Schoeny, H., Phillips, G., Darrough, E., DeAnna, S., Jarden, M., Johnson, D., & Lorch, G. (2014). Equine-assisted intervention for people with dementia. *Anthrozoös*, 27(1), 141-155.
- Dahlberg, L.L., Toal, S.B., & Behrens, C.B. (1998). *Measuring violence-related attitudes, beliefs, and behaviors among youths: A compendium of assessment tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Control and Prevention, Division of Violence Prevention.
- Engel, R.J., & Schutt, R.K. (2013). *The practice of research in social work, third edition*. Los Angeles: Sage.
- Fisher, A.K., Simmons, C., & Allen, S.C. (2016). An experiential approach to clinical supervision training: A mixed methods evaluation of effectiveness. *Journal of Teaching in Social Work*, 36(5), 460-476.
- Grinnell, R.M., & Unrau, Y.A. (2014). *Social work research and*

- evaluation: Foundations of evidence-based practice*, 10th ed. NY: Oxford University Press.
- Harris, M.J. (2010). *Evaluating public and community health programs*. San Francisco: Jossey-Bass.
- Holkup, P.A., Tripp-Reiner, T., Salois, E.M., & Weinert, C. (2004). Community-based participatory research: An approach to intervention research with a Native American community. *Advances in Nursing Science*, 27(3), 162-175.
- Howick, J., Friedemann, C., Tsakok, M., Watson, R., Tsakok, T., ...Heneghan, C. (2016). Correction: Are treatments more effective than placebos? A systematic review and meta-analysis. [Correction of article originally published 2013]. PLoS ONE, 11(1).
- Jackson, A.C. (2013). Current and future approaches to the therapy of human rabies. *Antiviral Research*, 99(1), 61-67.
- Jimenez, D.E., & Czaja, S.J. (2016). Recruitment and retention: Two of the most important, yet challenging, tasks in behavioral intervention research. In L.N. Gitlin & S.J. Czaja, (Eds.), *Behavioral intervention research: Designing, evaluating, and implementing*. NY: Springer.
- Jordan, C., & Franklin, C. (2003). *Clinical assessment for social workers: Quantitative and qualitative methods*, second edition. Chicago: Lyceum.
- Kagotho, N., Bowen, N.K., Ssewamala, F.M., Vaughn, M.G., & Kirkbride, G. (2018). Cross-national validity of the Beck Hopelessness Scale for children and adolescents: Findings from the YouthSave-Impact Study Kenyan. *International Journal of Culture and Mental Health*, published online (January). <https://doi.org/10.1080/17542863.2017.1417457>
- Kapp, S.A., & Anderson, G.R. (2010). *Agency-based program evaluation: Lessons from practice*. Los Angeles: Sage.
- King, J.A., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *How to assess program implementation*. Thousand Oaks, CA: Sage.
- Lee, J.H., Jung, K.Y., & Choi, Y.H. (2018). Screening test for at-risk drinking: Development of new abbreviated version of alcohol use

- disorder identification test for young and middle-aged adults. *Emergency Medicine International*, 2018, 1-6.
- Lopez, S.J., & Snyder, C.R. (2003). *Positive psychological assessment: A handbook of models and measures: A handbook of models and measures*. Washington, DC: American Psychological Association.
- Madson, M.B., & Campbell, T.C. (2006). Measures of fidelity in motivational enhancement: A systematic review. *Journal of Substance Abuse Treatment*, 31, 67-73.
- Magura, S., & Moses, B.S. (1986). *Outcome measures for child welfare services: Theory and applications*. Washington, DC: Child Welfare League of America.
- Miller, W.R., & Mattson, M.E. (2004). *Combined behavioral intervention manual: A clinical research guide for therapists treating people with alcohol abuse and dependence*. COMBINE Monograph Services, Volume 1. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism.
- Miller, D.C., & Salkind, N.J. (2002). *Handbook of research design & social measurement*, 6th ed. Thousand Oaks, CA: Sage.
- Perron, B.E., & Gillespie, D.F. (2015). *Key concepts in measurement*. NY: Oxford University Press. Pocket Guides to Social Work Research.
- Pettinati, H.M., Anton, R.F., & Willenbring, M.L. (2006). The COMBINE Study. *Psychiatry*, 3(10), 36-39.
- Pyles, L. (2015). Participation and other ethical considerations in participatory action research in post-earthquake rural Haiti. *International Social Work*, 58(5), 628-645.
- Quinn, C.R., Volpe, E.M., Nichols-Hadeed, C., Pierre, P., Douglas, V., Stewart, A.E., & Cerulli, C. (2017). Investigating barriers and facilitators to develop a randomized control trial using community-based participatory research with vulnerable youth at risk of homelessness and violence, (p. 1-20). *Sage Research Methods Cases, Part 2*. London: Sage.
- Rush, A.J., First, M.B., & Blacker, D. (2008). *Handbook of psychiatric*

- measures, 2nd ed. Washington, DC: American Psychiatric Publications.
- Sajatovoc, M., & Ramirez, L.F. (2012). *Rating scales in mental health*, 3rd ed. Baltimore, MD: Johns Hopkins University Press.
- Saunders, J.B., Aasland, O.G., de la Fuente, J.R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption. *Addiction*, 88(6), 791-804.
- Smyrinos, K.X., & Kirkby, R.J. (1993). Long-term comparison of brief versus unlimited psychodynamic treatments with children and their parent. *Journal of Consulting and Clinical Psychology*, 61(6), 1020-1027.
- Suresh, K.P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences*, 4(1), 8-11.
- Touliatos, J., Perlmutter, B.F., Straus, M.A., & Holden, G.W. (2001). *Handbook of family measurement techniques*, vols. 1, 2 & 3. Thousand Oaks: Sage.
- Trochim, W.M.K., & Donnelly, J.P. (2007). *The research methods knowledge base*, third edition. Mason, OH: Atomic Dog/Thomson Custom Solutions Publishing.
- Valencia-Garcia, D., Bi, X., & Ayón, C. (2017). Sensitivity and specificity in three measures of depression among Mexican American women. *Journal of Immigrant and Minority Health*, 19(3), 562-571.
- Wade-Mdivanian, R., Anderson-Butcher, D., Newman, T.J., & Ruderman, D.E. (2016). Exploring the long-term impact of a positive youth development-based alcohol, tobacco and other drug prevention program. *Journal of Alcohol & Drug Education*, 60(3), 67-90.
- Wang, E.A., Aminawung, J.A., Wildeman, C., Ross, J.S., & Krumholz, H.M. (2014). High incarceration rates among black men enrolled in clinical studies may compromise ability to identify disparities. *Health Affairs*, 33(5), 848-855.
- Wolfe, D.A., Scott, K., Reitze-Jaffe, D., Wekerle, C., Grasley, C., &

Straatman, A.L. (2001). Development and validation of the Conflict in Adolescent Dating Relationships Inventory. *Psychological Assessment*, 13(2), 277-293.

MODULE 4

Module 4 Introduction

This module extends what was learned about descriptive and comparative statistics in our previous course to understanding how various statistical approaches help answer questions about social work interventions. Using Excel software, we review earlier lessons about group comparisons as applied to comparing intervention groups (chi-square, t-tests, Anova, and correlations). We introduce for the first-time repeated measure analysis for comparing pre- and post-intervention longitudinal data (paired t-test), and address how single-system research data might be analyzed. We briefly re-introduce nonparametric statistical principles and briefly introduce the idea behind logistic regression to test outcomes in intervention or evaluation research.

READING OBJECTIVES

After engaging with these reading materials and learning resources, you should be able to:

- Describe appropriate data analytic approaches for descriptive (mean, median, standard deviation) and group comparison questions based on the nature of the research questions and type of variables (single sample t-test, independent samples t-test, one-way analysis of variance, chi-square);
- Explain why repeated measures analysis (paired-t test) would be used on longitudinal data and describe the approach;
- Identify basic non-parametric and logistic regression principles relevant in analysis of intervention data;
- Recognize and explain approaches to analyzing single system design data for understanding social work interventions;
- Define key terms related to analyzing data for understanding

social work interventions.

Module 4 Chapter 1: Review and Extension of Descriptive and Group Comparison Analyses

In our previous course you learned to distinguish between univariate and bivariate analyses, and to both conduct and interpret a variety of each type of analysis. Those activities were directed toward evidence for understanding diverse populations, social work problems, and social phenomena. Much of what was learned applies to our current course concerned with intervention and evaluation research for understanding social work interventions. In this chapter, these concepts will be reviewed in the context of intervention and evaluation research.

In this chapter you:

- review principles of univariate analysis (mean, median, and standard deviation), this time as applied to data for understanding social work interventions;
- review principles of bivariate analysis (single sample t-test, independent samples t-test, analysis of variance, chi-square, and non-parametric analytic approaches), this time as applied to data for understanding social work interventions;
- learn basic principles about logistic regression which can answer questions about dichotomous intervention outcomes.

Review of Univariate Statistics Extended to Intervention Research

In our previous course (Module 4, Chapter 2), you learned about analyses that describe data one, single variable at a time—univariate analysis. These descriptive statistics and analyses continue to be relevant in research for understanding social work interventions:

- Frequency & proportion data (categorical variables)
- Central tendency analysis (numeric/continuous variables)
 - mean
 - median
 - mode
- Distribution analysis (numeric/continuous variables)
 - range
 - variance
 - standard deviation
 - normal curve
 - skew & kurtosis

For example:

- In evaluating an intervention to help families who are or at risk of homelessness, you might collect and analyze data concerning the proportion of families who were able to achieve stable housing status for a year or more (a dichotomous, categorical variable).
- You might wish to compare the mean number of disciplinary actions that occur in middle school classrooms before and after teachers are trained to employ trauma-informed teaching and classroom management practices.
- In evaluating the impact of an intervention on native-speakers of English with its impact on individuals for whom English is a second or third language, you might examine the variance in

outcomes for these two groups—little variability in the second group might indicate that the intervention was not being adequately “received” or “processed” by these individuals.

- You may need to assess the distribution of values on an outcome variable in planning your statistical analyses in an intervention study—whether or not the values are relatively normally distributed may determine which statistical tests are most appropriate.

Review of Bivariate Analysis Extended to Intervention Research

Recall that bivariate analysis is about assessing the nature and strength of relationships that exist between two variables. In our previous course you learned several statistical tests for answering different types of research questions using different types of variables. You learned about the:

- single sample t-test
- independent samples t-test
- analysis of variance (Anova)
- chi-square test
- correlation test

In evaluating social work interventions, any of these tests might be applied, depending on the nature of the study questions, study design, and study variables involved. Here we will review all but the correlation test because that it seldom used as a means of evaluating outcomes in intervention research.

Single sample *t*-test.

In a very simple post-only study design an investigator might wish to compare a group's mean outcome score to a pre-determined, established standard or norm value. This would be used, for example, to evaluate the impact of an intervention to improve the health of infants born to mothers who smoke cigarettes by reducing or eliminating smoking during pregnancy (and, ideally, after birth, as well). The outcome measure could be the newborn infants' 5-minute Apgar scores, a measure routinely recorded at birth for all babies born in the United States (and many other countries). An Apgar score of 7, 8, or 9 means an infant is healthy; lower Apgar scores are associated with poorer health outcomes. Apgar scores can only be measured within minutes following birth, so a post-only design might make sense. The null hypothesis would be:

H_0 : the post-intervention mean Apgar score (measured at 5-minutes post-birth) is 7 or greater for babies born to mothers who received the intervention.

You can refer to your prior course Excel workbook exercises as a reminder of how to implement the steps involved in a **single sample *t*-test**. The selection of 7 as the comparison mean value is a bit arbitrary. Ideally, the investigators have historical data from before the intervention was made available concerning the mean 5-minute Apgar score for mothers who smoke. For now, the investigation uses the “healthy baby” value as the comparison mean.

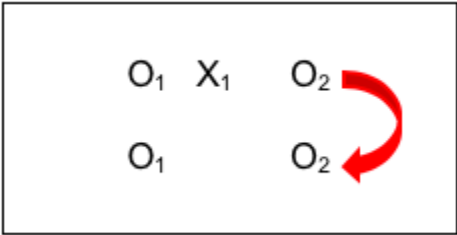


Independent samples t -test.

In Module 3 we considered study designs for answering research questions that involve comparing two intervention conditions. This might be a new intervention being compared to no intervention, a placebo, or a treatment-as-usual (TAU) “control” condition. Once investigators have collected outcome data for study participants who are exposed to these different intervention conditions, it is time to analyze those data to see if the hypothesized differences between the groups are actually observed. This is identical to the use of **independent samples t -tests** examined in our prior course where they were used to answer questions comparing groups to answer questions about diverse populations, social work problems, or social phenomena. Recall that certain assumptions need to be fulfilled in order to use this test—assumptions related to normal distribution and variance equality, for example. Otherwise, non-parametric analyses would be preferable. For the sake of clarity, let’s work with a specific example using independent samples t -tests to evaluate an intervention.

A Canadian team of investigators (Toneatto, Pillai, & Courtice, 2014) wanted to know if an intervention that enhanced Cognitive Behavioral Therapy (CBT) for problem gambling with Mindfulness

Training was better than no intervention (the control group participants were on a wait-list for the program and did, eventually, receive the intervention). Both groups in this pilot study (N=18 participants total, 9 in each group) were measured before and after the intervention. The study design looked like this (note there was no random assignment to the intervention or wait-list control condition):



The Variables.

In this example, the observation (O₂) data reflect the outcome variable. In this case, the outcome variables were expected to be dependent on the intervention condition manipulated by the investigators: Mindfulness Enhanced Cognitive Behavioral Therapy (M-CBT) versus Wait List Control (no intervention). Thus, the outcome variables are the dependent variables in this example. The outcome variables were a set of measures of gambling urges and symptoms; we can demonstrate the important points with just two of these, gambling diagnosis symptoms and gambling urges. Both of these variables are numeric in nature, with the scale on each showing that a higher score means a more serious gambling problem.

The Null Hypothesis.

The research question being asked is if there existed a significant difference between the outcomes for the group receiving the experimental intervention (X_1) and the group on the waiting list (non-treatment control)—was the difference between these two groups meaningfully different from zero (no difference)? In statistics logic, the investigators were testing the null hypothesis that the difference between the two groups was zero.

H_0 : there exists no statistically significant difference in symptoms or craving outcomes for the two treatment groups.

If the results of analysis led the investigators to reject the null hypothesis, it means they were reasonably confident that there existed a meaningful difference between the two groups (the difference was NOT zero). If the analysis led the investigators to fail to reject the null hypothesis, it means that no difference was detected (but they cannot conclude that no difference exists).

The Statistical Analysis Approach.

Since exactly two groups were compared, the independent groups t-test was appropriate. (Remember, in our prior course you learned that if three or more groups were being compared, an analysis of variance would be required; if only one group was involved, a single-sample t-test would be appropriate.) The underlying assumptions for this type of analysis are exactly the same as what you learned in in our prior course:

- Type of (Outcome) Variable: The scale of measurement for the dependent variable is continuous (interval).
- Normal Distribution: The dependent variable is normally distributed in the population, or a sufficiently large sample size was drawn to allow approximation of the normal distribution.

Note: the “rule of thumb” is that neither group should be smaller than 6, and ideally has more.

- Independent Observations: Individuals within each sample and between the two groups are independent of each other—random selection indicates that the chances of any one “unit” being sampled are independent of the chances for any other being sampled.
- Homogeneity of Variance: Variance is the same for the two groups, as indicated by equal standard deviations in the two samples.

The independent groups t-test analysis (sometimes called a student’s t-test) involved dividing a difference score by a variance estimate. This is accomplished by the following steps:

1. calculating the mean score for each group (2 groups’ means are called M_{group1} and M_{group2});
2. computing a “difference from the mean” score for each individual participant in each of the 2 groups—this is called the deviation score for each participant—then squaring that deviation score for each participant in each of the 2 groups;
3. computing the sum of the squared deviation scores for each group;
4. calculating the estimate for variance as the standard deviation squared for each group (S_{group1}^2 and S_{group2}^2) by taking the sum of the squared deviation scores for each group calculated in step 3 and dividing by the number of cases in that group minus 1 ($n_{\text{group1}} - 1$ and $n_{\text{group2}} - 1$);
5. compute the t-value as the difference between the two groups’ means (this is $M_{\text{group1}} - M_{\text{group2}}$) divided by the square root of a variance estimate computed as $[(S_{\text{group1}}^2 \text{ divided by } n_{\text{group1}}) + (S_{\text{group2}}^2 \text{ divided by } n_{\text{group2}})]$;
6. compare the computed t-value with the criterion value identified using the degrees of freedom for the total $N-2$ (where N is the total number of participants in the two groups)

combined) and the $\alpha < .05$ criterion.

Here is how it worked out in our example about the gambling intervention.

The Analysis Results and Interpretation.

The investigators observed a statistically significant difference in both outcome variables when they compared the enhanced intervention group to the non-treatment (wait listed) control group. Furthermore, the difference was in the expected direction: the intervention group’s mean scores were lower, indicating less severe gambling symptoms, than the control group’s mean scores.

Table 1. Outcome measures by group (adapted from Toneatto et al, 2014).

| | M-CBT (n=9) | Wait List (n=9) |
|-----------------------------|-------------|-----------------|
| DSM symptoms ^a | 3.44 | 6.71 |
| Gambling urges ^b | 7.67 | 15.29 |

^a $t(14)=2.72, p<.05$
^b $t(14), 2.97, p<.05$

For both study outcome measures (symptoms and urges), the differences between the two treatment group means were statistically significant (the $p<.05$). Thus, the investigators rejected the null hypothesis of no difference (that the difference was zero) and concluded that mindfulness enhancement of cognitive behavioral therapy for problem gambling warrants further, more systematic research attention. One caution about over-interpreting these results is that intervention of any type is usually preferable

to no treatment; future research should compare the mindfulness enhancement to the treatment-as-usual condition or the cognitive behavioral therapy alone.

Remember, also, that a strong intervention or evaluation research study also includes information concerning effect sizes—not only whether the differences were statistically significant but were they of meaningful magnitude, were the differences clinically significant?

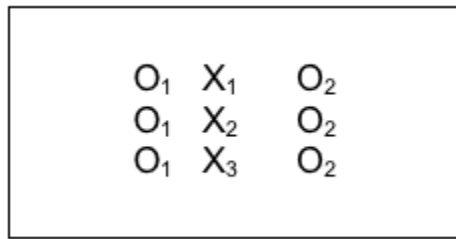


Analysis of variance (Anova).

The independent samples t-test is a fine analysis when two groups are compared. But, as you learned in our prior course, **analysis of variance (Anova)** is preferred when more than two groups are being compared. Again, recall that certain assumptions need to be fulfilled in order to use this test—assumptions related to normal distribution and variance equality, for example. Otherwise, non-parametric analyses would be preferable. Consider the example presented by Walsh and Lord (2004) in their study of client satisfaction with social work services. In one analysis, the authors compared satisfaction scores for three groups of parents referred for social work services in a pediatric hospital setting. The three groups were:

- parents referred for practical assistance(X_1)
- parents referred for counselling (X_2)
- parents referred for both practical assistance and counselling (X_3)

A diagram of this scenario might look like this:



The Variables.

In this example, the observation (O_2) data reflected the outcome (dependent) variable—client satisfaction. Client satisfaction was numeric in nature with higher scores indicating greater satisfaction. The investigators wondered if this was dependent on the parents' referral—a 3-category categorical variable.

The Null Hypothesis.

The research question asked was if there existed a significant difference between the outcomes (client satisfaction) for parents referred only for practical assistance ($n=4$), only for counseling ($n=9$), and for both practical assistance and counseling ($n=6$). In statistics logic, the investigators were testing the null hypothesis that the difference between the three groups was zero.

H₀: There exists no statistically significant difference in client satisfaction between these three groups of parents.

If the results of analysis led the investigators to reject the null hypothesis, it means they could be reasonably confident that there existed a meaningful difference between at least two groups, but they would not know which two groups without conducting further analyses. For example the difference(s) could be between:

- X₁ and X₂
- X₁ and X₃
- X₂ and X₃
- X₁ and X₂ and X₃.

The Anova test would be called an **omnibus test of significance**, meaning that it covers the whole group of possible comparisons, and the investigators would follow up with *post hoc* analyses to determine where the differences actually lie.

The Statistical Analysis Approach.

Since more than two groups were compared, the analysis of variance (Anova) test should have been appropriate—but there is a problem with the sample size (see the assumptions list below). The underlying assumptions for this type of analysis are exactly the same as what you learned in in our prior course:

- Type of (Outcome) Variable: The scale of measurement for the dependent variable is numeric or continuous (interval).
- Normal Distribution: The dependent variable is normally distributed in the population, or a sufficiently large sample size was drawn to allow approximation of the normal distribution.
Note: the “rule of thumb” is that no group should be smaller than 6, and ideally has more. In our example, this assumption was violated—the groups were as small as 4, 6, and 9.

- Independent Observations: Individuals within each sample and between the two groups are independent of each other—random selection indicates that the chances of any one “unit” being sampled are independent of the chances for any other being sampled.
- Homogeneity of Variance: Variance is the same for the two groups, as indicated by equal standard deviations in the two samples.

The Analysis Results and Interpretation. The investigators observed no statistically significant difference in client satisfaction scores across the three groups of parents. What they stated was:

“The mean client satisfaction score for parents referred for practical assistance was 22.75 (SD = 6.65), while the mean for parents referred for counselling was 27.67 (SD = 3.32), and the mean for parents referred for both practical assistance and counselling was 28.33 (SD = 2.34). An ANOVA revealed that there was a trend for parents referred for counselling to express more satisfaction with the service provided, and the association between reason for referral and satisfaction approached significance ($F(17) = 2.79, p = .09$)” (Walsh & Lord, 2004, p. 48).

Breaking this information down into its relevant pieces, we start with the information about the three group means for client satisfaction: 22.75, 27.67, and 28.33. Just looking at these values there seems to be a trend where the practical assistance only group is less satisfied than the two groups that received counseling. We also see that the F-statistic computed for the three groups’ differences was 2.79 and that this was not great enough to meet our $p < .05$ criterion for rejecting the null hypothesis. We also see that there were 17 degrees of freedom in this analysis—and we can figure out that there were 2 between groups degrees of freedom (one less than the number of groups).

There is some disagreement in the field as to whether results

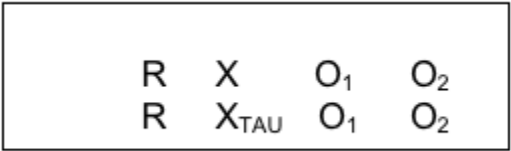
that “approach significance” should be reported in the literature. On one hand, it is potentially useful information in reviewing a body of literature. On the other hand, it is too easily over-interpreted as a meaningful finding and there is also the possibility that a Type II error is being made—if the sample were sufficiently large to meet the test assumptions, a significant effect might have been observed, but because the sample was so small the null hypothesis could not be rejected. Looking closely at the reported results also suggests that the assumption of equal variances might also have been violated in this sample. Hence, the earlier comment that the Anova might have been the appropriate analysis plan—this is a case where **nonparametric analysis** might have been preferable if the sample size could not be expanded.



Chi-square analysis.

We have reviewed what is done when the outcome variable in an intervention study is numeric (t-test or Anova). We should also consider the best choice for a situation where both the independent and outcome/dependent variable are categorical in nature: **chi-square analysis** is suited for this job. Consider a study conducted in Iran evaluating a brief, home-based, motivational intervention

delivered by social workers designed to encourage men who use methamphetamine to enter substance misuse treatment. The study was designed as a randomized control trial where the control group received treatment-as-usual consulting services and data were collected one week and 3 months post-intervention. The study design looked like this:



The Variables.

One outcome variable was whether the men entered a treatment program (dichotomous, yes or no). In this example, both the independent (treatment condition) and dependent (treatment entry) variables were categorical in nature; in fact, both were dichotomous categorical variables. Another way of describing this study in terms of the sample is as a 2 X 2 design (2 groups on each variable). For this reason, chi-square analysis was performed on the data.

The Null Hypothesis.

The research question asked was if there existed a significant difference between the outcomes (proportion entering treatment) for men who received the innovative intervention compared to men who received only treatment as usual. In statistics logic, the investigators were testing the null hypothesis that the difference between the two groups of men was zero.

H₀: There exists no statistically significant difference in proportion of men entering treatment between those who do and do not receive the innovative intervention.

If the results of analysis led the investigators to reject the null hypothesis, it means they could be reasonably confident that there existed a meaningful difference in treatment entry related to whether the men received the innovative intervention or the treatment-as-usual condition. If the results led the investigators to fail to reject the null hypothesis, they could only say that they did not detect a significant difference.

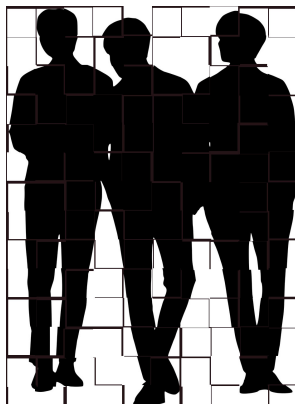
The Statistical Analysis Approach.

Since both variables were categorical in nature, chi-square analysis was appropriate.

The Analysis Results and Interpretation.

The investigators reported that a total of 56 men participated in the study, with equal groups of 28 randomly assigned to each of the intervention conditions. The authors described the results of their chi-square analysis for entering a treatment program within one week after the intervention as the innovative intervention group having significantly higher treatment program participation than the treatment-as-usual group (75% vs 10.7%), where $\chi^2(56) = 21.073$, $p < 0.001$ (Danaee-far, Maarefvand, & Rafiey, 2016, p. 1866). Furthermore, they also reported that the rate of retention in treatment at three months follow-up was greater for the innovative intervention group compared to the treatment-as-usual comparison group (60.7% vs. 14.3%) where $\chi^2(56) = 12.876$, $p < 0.001$ (Danaee-far, Maarefvand, & Rafiey, 2016, p. 1866). The authors concluded that the brief social work intervention contributes to

treatment participation and retention compared to the usual consulting services for men who use methamphetamines.



Logistic regression.

You now know how investigators might analyze intervention or evaluation data where both the independent and dependent variables are categorical in nature (chi-square). You have ways of analyzing data where the independent variable is categorical and the dependent variable is numeric/continuous/interval (t-test and Anova). You even know what to do when both variables are numeric (continuous) in nature: that is when correlation analysis is often helpful.

What is worth mentioning is the remaining possibility: the dependent variable is categorical and the independent variable is numeric/continuous/interval. In this case, a different approach to analysis is adopted, on the spectrum of what are called regression analyses. These operate from a different set of assumptions and have capabilities that differ somewhat from the parametric analyses we have studied in our two-course sequence. They are based on linear algebra rather than differences in means and variance.

Without going into a great deal of detail, the test that might be used in this case—as long as the dependent variable is a dichotomous (2-category) categorical variable—would be what is called **logistic regression**. An example might be the outcome (dependent) variable is treatment completion vs drop-out and the independent variable is a numeric variable like age or a score on symptom severity or distance from home to the intervention location. A real-world example comes from the Safe At Home data where treatment completion (or dropout) was related to the number of weeks between the intimate partner violence incident and intake to batterer treatment. While we are not studying this regression approach in detail, it is important to recognize that this analysis approach might be applied to the special case where independent variables are numeric and dependent variables are both categorical and dichotomous in nature.

Stop and Think



following activity.

Take a moment to complete the



*An interactive or media element has been excluded
from this version of the text. You can view it online*

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=336>

Chapter Summary

In this chapter you revisited topics related to univariate analysis, this time in relation to intervention and evaluation research questions and study designs: frequency and proportion data for categorical variables, as well as central tendency and distribution analyses for numeric/continuous variables. You also revisited topics related to bivariate analysis: single sample and independent samples t-tests, analysis of variance, and chi-square analysis. You were briefly introduced to logistic regression and reminded that non-parametric approaches might be preferable to parametric approaches for analyzing small-sample data sets. One important data analysis approach remains to be explored: what to do with longitudinal data, such as the pre-/post- comparison situation. This is the topic of the next chapter in this module.

Module 4 Chapter 2: Paired t-Test

At this point you have developed skills in the analysis of group comparison data. However, because intervention research designs are often longitudinal in nature, it is time to learn about how comparing pre-/post-intervention data collected from the same participants at two different time points requires a different approach to statistical analysis.

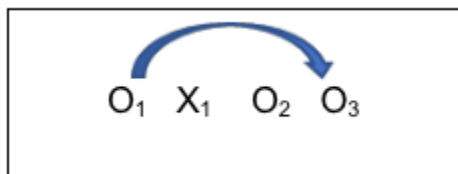
In this chapter you learn:

- the importance of repeated measures analysis approaches for longitudinal data;
- steps involved in conducting and interpreting paired-t test analysis (including an Excel exercise).

Pre-/Post-Intervention Comparisons Are Different (Longitudinal Data)

We considered analyses for an intervention research question in Chapter 2 concerning differences in gambling symptoms with and without a novel intervention being delivered— X_1 and wait list no intervention control conditions (Toneatto et al., 2014). The investigators' design permitted them also to longitudinally examine change in the participants' gambling symptoms before and after the experimental intervention (X_1). Once the investigators collected the post-intervention outcome data for participants (O_2), they also collected 3-month post-intervention follow-up data (O_3). Let's look at how they analyzed the data intervention group's baseline (O_1) and follow-up (O_3) data to see if the observed differences (O_1 to O_2)

seemed to hold up over time. Using our study design notation, it looks like this:



Investigators in this situation cannot simply apply the independent samples t-test or the Anova analysis approach as if they had two independent groups to compare because of that longitudinal design. And, this is why...

We already know something about how to statistically compare two groups—but here the investigators did not have two independent groups to compare. Instead, they had the same elements (study participants) being measured twice—before and 3-months after the intervention. In this case, an important assumption in the independent samples t-test and the analysis of variance (Anova) is being violated in the longitudinal data: it is right there in the name of the t-test, “independent samples.” With longitudinal data we no longer have independent groups—by the very nature of having been collected from the same individuals the later round of data is not independent of the first-round data. Participants providing that later round of data are completely dependent on having been the same participants providing the first round of data. This non-independence means that a portion of the variance in the data is due to the individual differences that existed at the first point in time, not due to intervention-caused differences across the two points in time.

Statisticians have developed an elegant solution to this problem in what is called the **paired t-test**. We use the paired t-test approach in analyzing data where the data involve repeated measurement

with the same participants, such as comparing pre- and post-intervention or pre- and follow-up longitudinal data. Paired t-test (and other repeated measure analysis) solutions account for the potential impact of the lack of independence resulting from repeatedly measuring the same individual elements. It is still critically important that the individual pairs of scores be independent of all other pairs of scores (the participants be independent of each other), but it is acceptable that each individual's longitudinal scores are non-independent at the two points in time—they cannot be since they are produced by the same person. The solution to this problem when comparing two points in time (pre- and post- or pre- and follow-up, for example) is to use a paired t-test instead of an independent samples t-test. In the case of more than two time points being compared, there is an analogous repeated measures analysis of variance, *rmanova*, to replace the one-way analysis of variance we previously learned—for example, a single analysis that included all three time points, O_1 , O_2 , and O_3 all together (which is preferable as a means of avoiding a Type II error since it is one less analysis to risk making the wrong decision).

Re-Visiting the Gambling Study Example.

Let's continue to work the example from the problem gambling study (Toneatto et al., 2014) to examine change in the participants' gambling symptoms before and at follow-up 3 months after intervention.

The Variables. In this example, the observation (O_1 and O_3) data reflect the problem gambling symptoms variable (DSM symptoms); the example also works for the variable called gambling urges. In this case, the follow-up means were expected to be different from the pre-intervention (baseline) means in the intervention group. As a reminder, both gambling-related variables are numeric in nature, with the scale on each showing that a higher score means a more

serious gambling problem. The two points in time represent a dichotomous categorical variable.

The Null Hypothesis (H_0). The primary research question being asked in this analysis is if there exists a significant difference between the pre- and follow-up gambling variable means for the group receiving the experimental intervention (X_1)—is the difference between pre- and follow-up meaningfully different from zero (no difference)? In statistics logic, the investigators were testing the null hypothesis that the difference between the two points in time was zero.

H_0 : No difference exists between the pre- and follow-up intervention means.

If the results of analysis lead the investigators to reject the null hypothesis, it means they can be reasonably confident that there is a meaningful difference between the pre- and follow-up means (the difference is NOT zero). If the analysis leads the investigators to fail to reject the null hypothesis, it means that no difference was detected (but they cannot conclude that no difference exists). The investigators are hoping to reject this null hypothesis for the intervention group with both the DSM symptoms and the gambling urge outcome variables.

The Statistical Analysis Approach.

Since the data being compared at exactly two points in time is paired (non-independent), the paired t-test analysis is appropriate. With the exception of this aspect of (non)independence in paired, longitudinal data, the underlying assumptions for this type of analysis are exactly the same as what you learned in our prior course and previously reviewed for the independent samples t-test:

- *Type of (Outcome) Variable*: The scale of measurement for the dependent variable is continuous (interval).

- Normal Distribution: The dependent variable is normally distributed in the population, or a sufficiently large sample size was drawn to allow approximation of the normal distribution. Note: the “rule of thumb” is that neither group should be smaller than 6, and ideally has more.
- Independent Observations: Individuals within each sample and between the two groups are independent of each other—random selection indicates that the chances of any one “unit” being sampled are independent of the chances for any other being sampled.
- Homogeneity of Variance: Variance is the same for the two groups, as indicated by equal standard deviations in the two samples.

The paired *t*-test analysis, like the independent samples *t*-test, involves dividing a difference score by an estimate of variance. The major difference lies in how the difference scores and the variance estimate (square of standard deviation) are handled:

1. calculate the difference score for each individual participant ($\text{score}_{\text{time 2}} - \text{score}_{\text{time 1}}$) and making certain that you preserve the “sign” of each difference as being a positive or negative value;
2. compute the “difference” mean by adding together the individual participants’ difference scores and dividing by the number of participants (this is why it is important to preserve the “sign” for each difference score because you may need to add in some negative numbers along with some positive numbers)—this value will be used as the numerator in your calculation of the *t*-value;
3. for the denominator in your calculation of the *t*-value, begin by calculating the square of each difference score from step 1, then add these squared differences together (this will be a sum of squares)—it would be written as $\sum(\text{difference}^2)$;
4. then calculate the sum of the difference scores, square this

value, and divide by the sample size, N —it would be written as $(\sum \text{difference})^2 / N$ —notice the difference where the symbol for squaring is located inside the parenthesis and outside the parenthesis;

5. multiply the sample size minus 1 by the sample size—it would be written as $(N-1) * (N)$;
6. compute the t -value denominator as the square root of [(step 3 minus step 4) divided by step 5];
7. compute the t -value as the numerator from step 2 divided by the denominator from step 6;
8. compare the computed t -value with the criterion value using the degrees of freedom for the total $N-1$ (where N is the total number of participants) and the $\alpha < .05$ criterion.

The Analysis Results and Interpretation.

The investigators in the problem gambling study (Toneatto et al., 2014) observed a statistically significant difference in pre-/follow-up outcome variables for the experimental intervention group (X_1). Furthermore, the difference was in the desired direction: the follow-up mean scores were lower 3 months following the intervention than at baseline (pre-intervention), indicating less severe gambling symptoms (see Table 1).

Table 1. Outcome measures at 3-months post-intervention compared to pre-intervention baseline (adapted from Toneatto et al, 2014).

| | M-CBT (n=9) | |
|----------------|-------------|-------------------|
| | pre | post |
| DSM symptoms | 7.07 | 3.00 ^a |
| Gambling urges | 19.36 | 5.36 ^b |

^a $t(13)=4.41, p<.001$
^b $t(13)=9.18, p<.001$

For both outcome measures reported here, the differences between the group means were statistically significant. Thus, the investigators rejected the null hypothesis of no difference (that the difference was zero) and concluded that intervention effects persisted at 3-month follow-up. They believe that mindfulness enhancement of cognitive behavioral therapy for problem gambling warrants further, more systematic research attention. One caution about over-interpreting these results is that intervention of any type is usually preferable to no treatment; future research should compare the mindfulness enhancement to the treatment-as-usual condition or the cognitive behavioral therapy alone.



Interactive Excel Workbook Activities

Complete the following Workbook Activity:

- [SWK 3402.4-2.1 Paired t-test Analysis Exercise](#)

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online here:

<https://ohiostate.pressbooks.pub/swk3402/?p=339>

Chapter Summary

In this chapter you learned a new statistical approach to bivariate analysis of quantitative data. The paired-t test is one example of repeated measures analysis that is appropriate when the data being compared are not independent because they are from the same individuals at two points in time. You learned why this is important (the estimate of variance needs to be adjusted for this non-independence) and you learned how to engage in this type of analysis. You now know how investigators and evaluators might work with data from most types of intervention and evaluation study designs. What remains is understanding how to work with data generated from single-system design studies. That is the topic of the final chapter in this module.

Module 4 Chapter 3: Analysis of Single System Design Data

As you learned in the previous module, one approach to evaluating social work practice involves collecting data using a single system design. Many of the approaches we have studied are about what happens with different groups of participants. The single system design is sometimes called “intrasubject” research: the data are related to what happens within an individual system (Nugent, 2010). Single system designs generally involve collecting baseline data repeatedly for a period prior to implementing an intervention (the “A” phase) and collecting data during the intervention period (the “B” phase). A common variation on this design include collecting data during a period when the intervention is removed (another “A” phase), hence the ABA designation. Many other variations exist including adding and comparing alternative interventions (B₁, B₂, B₃ for example) and collecting data during each of those phases. The issue in this current chapter is what to do with the collected data, how to analyze it for meaning.

In this chapter you learn:

- visual graphic approaches to analyzing single system design data;
- assessing changes between single system design phases in level, trend, variability, overlap, means, persistence, and latency of change;
- two statistical analysis approaches for complementing visual approaches (non-overlap of all pairs analysis and two-standard deviation band analysis).

Graphing Single System Design Data

The first step in analyzing single-system design data is the visual analysis. According to Engel and Schutt (2013), visual examination of the graphed data is “the most common method” of analysis (p. 200). The purpose is to determine whether the target variable (outcome) changed between the baseline and intervention phases. Let’s begin by examining the elements of the single system design graph.

First, the phase of a single study design can be considered as the **independent variable** (the one that is manipulated experimentally) and the measured outcome as the **dependent variable**. Social work interventions often have multiple goals or objectives—perhaps addressing multiple behaviors, or a behavior and an attitude, or knowledge on several topics. For the purposes of simplifying the understanding single system design data, we will focus on data measuring only one outcome variable at a time.

The single system design graph has “time” as the horizontal (**x**) **axis** and the frequency or number count of events on the vertical (**y**) **axis**. Clearly identified in the graph are transition points in the phases of the single system design—when baseline ends and intervention begins, for example.

- **x axis:** Time might be measured in minutes, hours, days, or weeks, depending on the natural frequency of the behavior targeted for change. For example, the frequency of facial and spontaneous vocalization tics might be measured in minutes or hours for a child experiencing Tourette’s syndrome, while the frequency of a couple’s arguments might be measured in days or weeks or the frequency of temptation to drink (alcohol) might be measured in hours or days.
- **y axis:** The scale used for the “count” data running up the vertical axis is important to consider. A major factor involves the degree of variability in the values counted. If the lowest value is 0 and the highest value is in the hundreds, then it will

be very difficult to discern differences of 5 or 15—they will seem very close together in the scale. On the other hand, if the variability runs between 0 and 20, a difference of 5 or 15 will show up very well.

- **transition points:** One consistent approach is drawing a vertical line through the point in time when a transition occurs and labeling it. It is also possible to use color changes if the graph can be presented in a color format (some printers or copiers will lose this feature).

Here is an example of an ABAB single system design graph to consider (adapted from Prochaska, www.sjsu.edu/people/fred.prochaska/courses/ScWk240/s1/ScWk-240-Single-Case-Designs-Slides-Week-8.pdf).

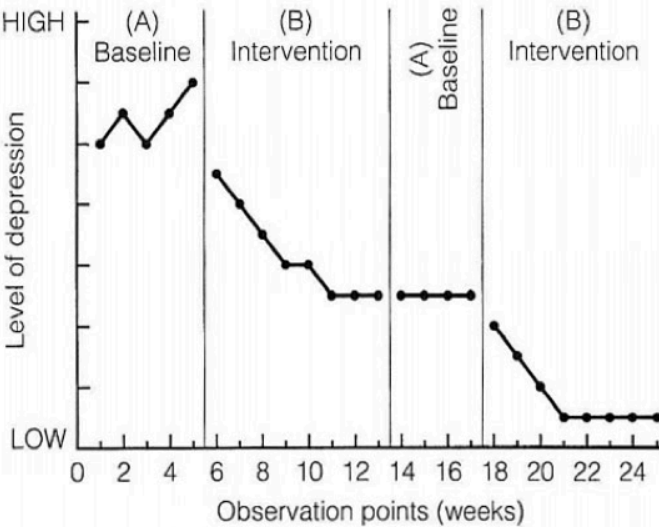


Figure 14-8 Graph of Hypothetical Outcome of ABAB Design Supporting Intervention Efficacy

The client and social worker together can see that both the level and variability of depression were high during the five weeks of baseline and there may have been an upward trend occurring at the point when intervention began (week 6). There was a noticeable downward trend in level of depression during most of the intervention period, with a plateau occurring around week 11. Without additional intervention during weeks 14 to 17, the client's level of depression did not worsen again, nor did it continue to improve—it remained at the previous plateau level. With additional intervention, the level of depression again showed a downward trend until it leveled off at a low level during the final weeks of intervention.

Note that it would be unusual to leave a person experiencing depression without intervention for 5 weeks of baseline data collection; however, this example was created as a hypothetical case. Also note that sometimes target behaviors show a degree of “natural” improvement during baseline as a benefit of the screening and assessment process. For example, in a study of women experiencing alcohol dependency, investigators observed significant improvement in drinking behavior prior to the beginning of intervention, during assessment of their drinking problems:

“Changes in drinking frequency occurred at all four points in the pretreatment assessment process, resulting in 44% of the participants abstinent before the first session of treatment. A decrease in drinking quantity across the assessment period also was found” (Epstein et al., 2005).

The authors concluded several points from these observations:

- the assessment process itself might be therapeutic for some participants—significant changes began with seeking treatment, before treatment started;
- interpreting intervention outcome results should be informed by this fact—interventions should be evaluated in comparison to initial intake/screening levels rather than assessment

process data to detect the full impact of the entire experience of assessment and intervention combined;

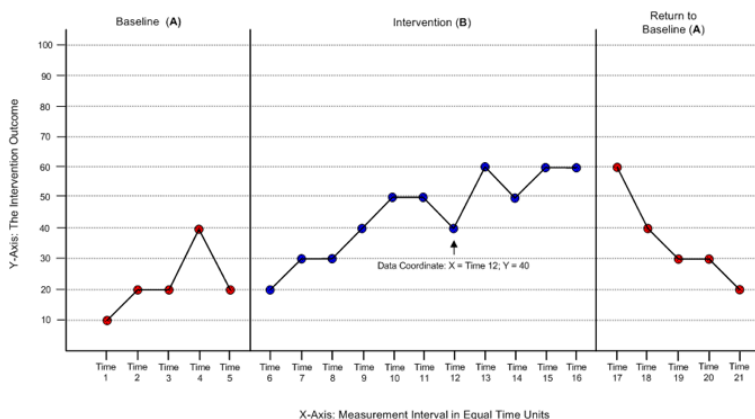
- intervention outcome results should be informed by the observation that women who showed assessment phase improvement also showed the greatest outcomes during and following treatment (12-month follow-up).

Assessing Change

According to Nugent's (2010) guide for analyzing single system design data, visual methods allow evaluators to detect several important types of changes or contrasts in the data for different study phases. Let's explore what these different kinds of information suggested by Nugent (2010) offer in understanding and evaluating social work interventions.

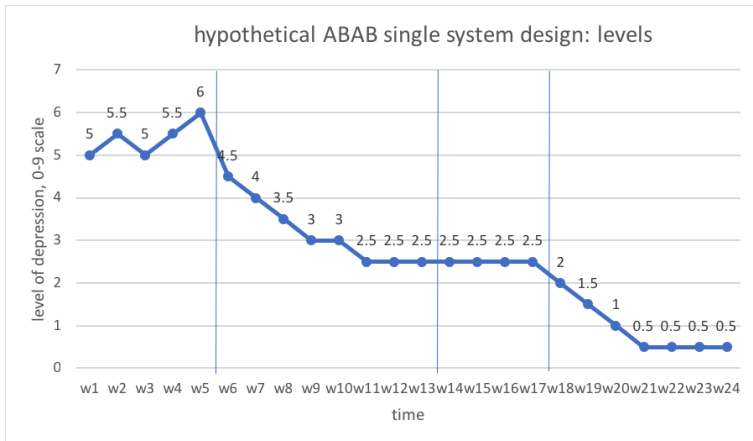
Level

Plotting the "levels" data is simply a matter of identifying the point in a graph where the "y" value for each "x" time point is located. This graph shows where the "point" for week 12 at a value of 40 would go on an ABA (baseline, intervention, remove intervention) single system design graph. Note that it is critically important that the units of time be of equal intervals. It would not work to have the baseline phase measured in days and the intervention phase in weeks, for example.



Let's work with the concrete example concerning a client's depression levels in an ABAB single system design. Here, in week 1 the level was 5, in week 2 the level was 5.5, and so forth on a 10-point scale (0 to 9 values). This chart shows the data from the previous hypothetical depression levels example where you can see the levels reported for each of the 4 phases:

- A₁-weeks 1-5 (first baseline phase), levels range 5.0-6.0
- B₁-weeks 6-13 (first intervention phase), levels range 2.5-4.5
- A₂-weeks 14-17 (removal of intervention phase), levels range 2.5-2.5
- B₂-weeks 18-24 (second intervention phase), levels range 0.5-2.0



Interactive Excel Workbook Activities

Complete the following Workbook Activity:

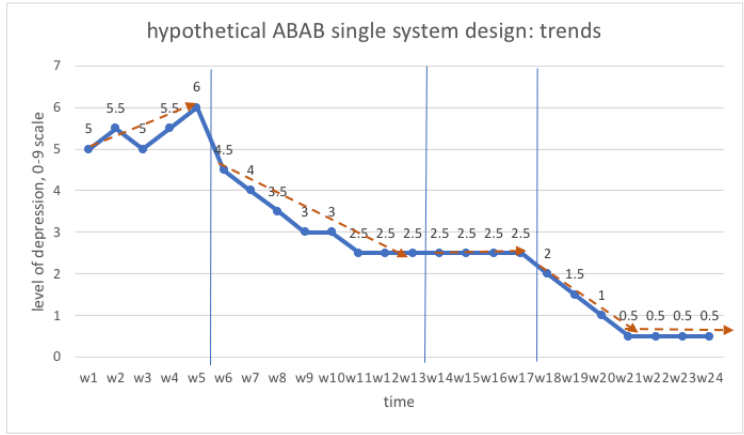
- [SWK 3402.4-3.1 Graphing Single Systems Design Data](#)

Trend

Trends are less about amount or level of change and more about the rate of change. Rate of change is indicated by the slope of the line between points in a graph, slope being the angle of the line. A “flat” line which runs horizontally, parallel to the x axis, indicates

no change. The greater the angle, the steeper the slope, indicating greater degrees of change between points in time. Two methods for determining how the line should be placed on the graph are the statistical computation of the slope for a line connecting the data points and the “Nugent” method (Engel & Schutt, 2013). The statistical method called ordinary least squares computes the “best fitting” line that has the shortest total of distances from each data point to the line. A simpler method resulting in similar conclusions is to simply draw a line between the first and last data points in each phase—this is the “Nugent” method (described by Engel & Schutt, 2013, referring to the author of Nugent, 2010).

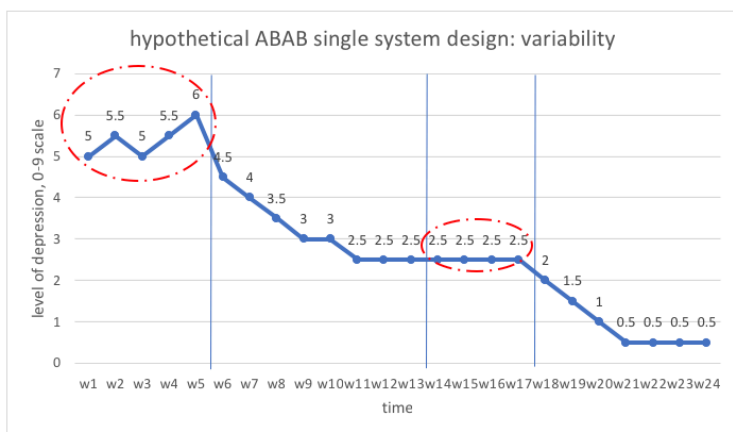
Looking at data from our previous hypothetical depression levels data, we can see a couple of trends represented as orange arrow lines created using the Nugent method:



As you can see, the baseline trend was in an undesired direction, the trend during the two intervention phases was in a desired direction, and two “flat” constant periods indicated no change.

Variability

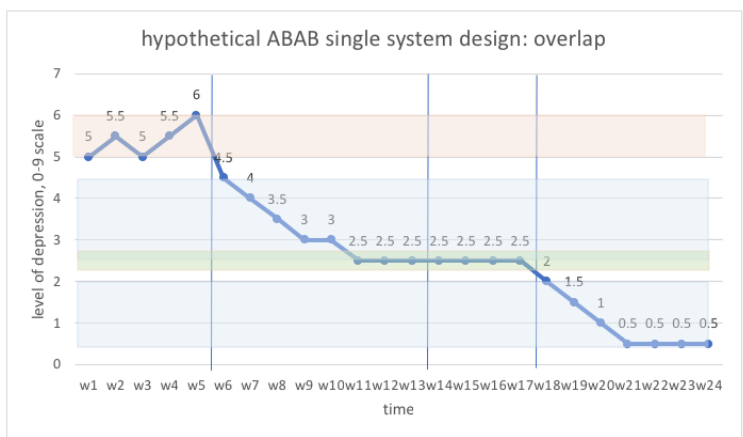
Fluctuations in the dependent variable may not be consistent across phases of a study. For example, the baseline period might be characterized by wildly fluctuating values and the intervention might help tame the variability, creating greater stability in the behavior of concern. See, for example the differences in the initial baseline (A_1) phase compared to the post-intervention phase without intervention (A_2). Variability looks different from a unidirectional trend in that it involves up-and-down fluctuations.



Overlap

The concept of overlap has to do with the range of levels observed in different phases. This is depicted by drawing lines through the top and bottom range values observed in each phase, then seeing if the ranges overlap or are separated. The greater the width of overlap area, the less evident the degree of change; smaller areas of overlap meant change is more evident (Nugent, 2010). Here is what overlap

mapping looks like in our hypothetical levels of depression example. There is no overlap between the baseline range and any of the other phases; some overlap between the first intervention phase and the removal of intervention phase; and, no overlap between the second intervention phase and any of the other phases.



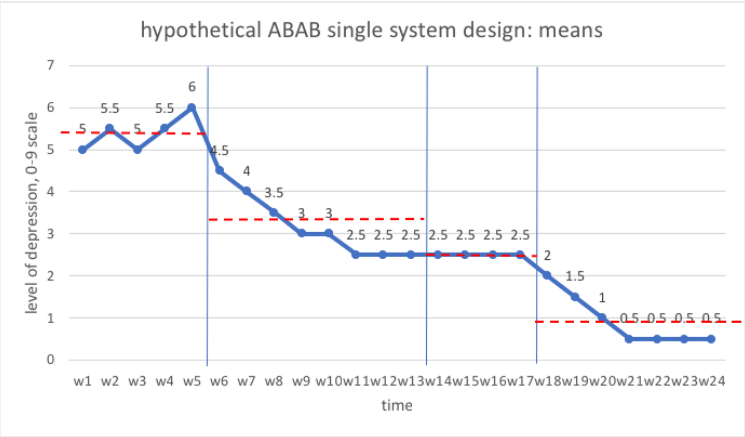
Means & medians

One strategy for detecting differences in data from different phases of the single system design study is to compute the mean and/or median of the values in each phase, then draw a line through the phase at each of those points. From our prior coursework, you know how to compute the mean (average) and the median (the 50th percentile, where ½ are above and ½ are below that value). Excel software can compute both of these values in very short order.

In our depression levels example, this is what the values would be:

| | Phase A ₁ | Phase B ₁ | Phase A ₂ | Phase B ₂ |
|--------|-------------------------|-------------------------|-------------------------|-------------------------|
| mean | 5.4 | 3.2 | 2.5 | 0.9 |
| median | 5.5 | 3.5 | 2.5 | 1.25 |
| N | 5 | 8 | 4 | 7 |

Here is an example of what the graph might look like with the mean lines drawn in, using the data from our hypothetical depression levels ABAB design example. As you can see, there is progressive improvement in levels of depression across the 4 phases of the study design. Later we will look at how to analyze these differences for statistical significance.



One thing to keep in mind, based on what you know about how “extreme” values influence the mean, is that the median might be a better tool to rely on when there exists a great deal of variability in the range of values (Engel & Schutt, 2013). Also Keep in mind that clinical significance is of great importance in analyzing results of the

intervention: a difference from 5.4 to 0.5 could be of great clinical significance in an individual's experience of depression, depending on what the 0-9 scale implies! Engel and Schutt (2013) suggest three means of determining clinical significance in these situations:

- *Establish criteria:* prior to engaging in the evaluation effort, determine what might be relevant criteria for success. “If the intervention reaches that point, then the change is meaningful” (Engel & Schutt, 2013, p. 202). The criteria could be defined in terms of a specified degree, amount, or level of change. These values should be informed by the literature concerning the measure used.
- *Use established cut point scores:* if the outcome is measured using clinical tools with established norms and/or cut point scores, determine “whether the intervention has reduced the problem to a level below a clinical cut-off score” (Engel & Schutt, 2013, p. 202). Thus, in determining clinical significance, the outcome (dependent) variable becomes a dichotomous categorical variable—above or below the clinical cut point.
- *Weigh costs and benefits:* in this case, the goal is to determine whether the efforts to produce change are resulting in sufficient change to “be worth the cost and effort to produce the improvement” observed (Engel & Schutt, 2013, p. 204).

Latency/immediacy of change

One other dimension that Nugent (2010) advises considering has to do with how long it takes before change is evident. In some cases, **latency** is short—the beginning of change attributable to intervention are almost immediately observed. In other instances, latency might be longer—it takes a while before the impact is beginning to be observed in the data. For example, changes in knowledge about a topic might happen quickly with intervention

but changes in attitudes, values, and beliefs may take longer to appear. This is particularly true of complex behaviors. In our example, changes with intervention seemed to take a relatively short time since we observed change beginning in the week of intervention implementation each time—latency was short.

Statistical Analysis of Single System Design Data

Graphs are quite useful in helping social work professionals and their clients interpret single system design data. However, precision interpretations are supported by statistical analysis of these data. Specifically, when there appear to be changes between phases, how do we know if the observed changes are meaningful? A variety of statistical approaches are described in the literature. Two approaches are presented in this chapter, based on three criteria to consider in selecting analytic approaches adapted from six criteria described by Manalov et al (2016). The statistical approach selected for analysis should:

1. be simple to compute and interpret without a high degree of statistical expertise and present a reduced likelihood of misinterpretation;
2. complement (rather than duplicate) visual analysis, especially when trend and variability patterns complicate visual analysis, providing different information;
3. be free from assumptions of data independence (longitudinally) or homoscedasticity (equal variance).

Thus, the two approaches introduced in this chapter are NAP analysis (non-overlap of all pairs) and the two-standard deviation band approach.

Approach #1: Non-overlap of all pairs (NAP) analysis.

This approach is essentially a non-parametric form of analysis based on probability: the Mann-Whitney U-test. **Nonoverlap of all pair (NAP) analysis** is about pairs of observations for two different phases in a single system study being in the desired direction. The “all pair” aspect in the title of the approach means that every observation in the time 2 data is paired to each of the observations in the time 1 data. This means that the number of pairs compared is the number of observations in time 1 multiplied by the number of observations in time 2. The probability default in this analysis is that no difference exists between two phases of single system data being compared. If “improvement” means that the observed frequency or amount decreases (such as number of arguments, need for time out, cigarettes smoked, or anxiety scores), then improvement means the “time 1” data will be greater than the “time 2” data more often. The null hypothesis is that there is no difference in the number of times that the paired values reflect improvement compared to the number that reflect no improvement.

To make easier to understand, imagine that we wish to compare the baseline (A₁) to the intervention phase (B) in our depression levels example. In this case we had 5 baseline observations and 8 intervention phase observations:

| A ₁ phase | B phase | A ₂ phase |
|----------------------|---------|----------------------|
| 5 | 4.5 | 2 |
| 5.5 | 4 | 1.5 |
| 5 | 3.5 | 1 |
| 5.5 | 3 | 0.5 |
| 6 | 3 | 0.5 |
| | 2.5 | 0.5 |
| | 2.5 | 0.5 |
| | 2.5 | 0.5 |

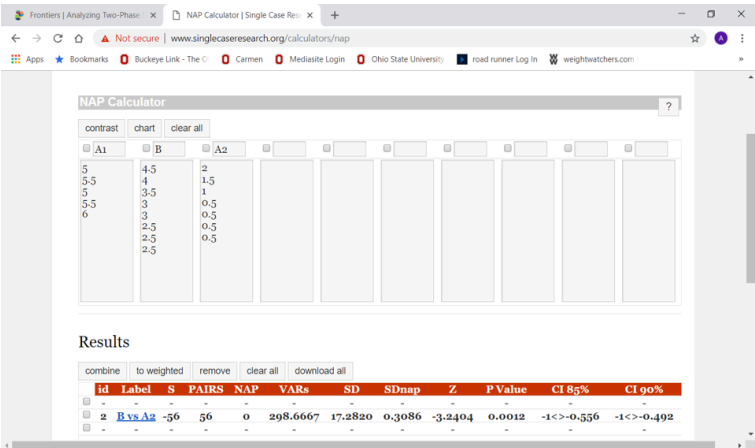
Thus, we have a total of 40 pairs of data ($5 \times 8=40$). This can be hand-calculated for a small data set using a somewhat complex formula to compute the probability. Or, the data can be entered in 2 columns in a calculator program at www.singlecaseresearch.org/calculators/nap (nap is an abbreviation for nonoverlap of all pairs; the calculator is attributed to Vannest, Parker, Gonen, & Adiguzel, 2016). In this example, it would look like this after clicking on “contrast” for the comparison of A vs B phases:

The screenshot shows the NAP Calculator interface. Under the 'contrast' tab, two columns are selected: A₁ and B. The A₁ column contains values 5, 5.5, 5, 5.5, 6. The B column contains values 4.5, 4, 3.5, 3, 3, 2.5, 2.5, 2.5. The results table is displayed below.

| id | Label | S | PAIRS | NAP | VARs | SD | SDnap | Z | P Value | CI 85% | CI 90% |
|----|---------------------|-----|-------|-----|----------|---------|--------|---------|---------|------------|------------|
| 1 | A ₁ vs B | -40 | 40 | 0 | 186.6667 | 13.6626 | 0.3416 | -2.9277 | 0.0034 | -1<>-0.508 | -1<>-0.438 |

The p -value for the Z test-statistic (-2.93) for our 40 pairs was 0.0034; our $p < .05$ allows us to reject the null hypothesis of no difference. In other words, there appears to be a statistically significant improvement between the intervention (B) phase and the baseline (A₁) phase.

Similarly, if we wish to compare the intervention (B) phase to the post-intervention (A₂) phase, we have a total of 56 possible pairs (8 x 7=56). Entering the data into the NAP calculator program we obtain the following result:



Once again, we observe that $p < .05$ (Z statistic for our 56 pairs is -3.24, $p = 0.0012$). Therefore, we can again reject the null hypothesis of no difference and conclude that the changes continued in a downward direction during the post-intervention period.

Approach #2: Two-standard deviation band method

This second approach is also valued the simplicity of computations

involved (Orme & Cox, 2001). It is called the **two-standard deviation band method** because it involves computing the mean and standard deviation from the baseline phase data, drawing the “band” representing two standard deviations (2 SDs) around the baseline mean onto the graph, and determining where the intervention phase data points fall in comparison to that “band.” The band includes values that would be 2 standard deviations (SDs) above and 2 standard deviations below the mean for the phase. Another name for this “banded” graph is a Shewhart chart (Orme & Cox, 2001).

A rule of thumb offered by Gottman and Leiblum (1974) is that if at least two consecutive intervention phase data points fall outside the band, a meaningful change was observed. The logic for this rule of thumb is that the probability of this happening by chance is less than the criterion of $p < .05$ (Nourbakhsh & Ottenbacher, 1994). Here is what it would look like for the hypothetical depression levels example with which we have been working throughout this chapter.

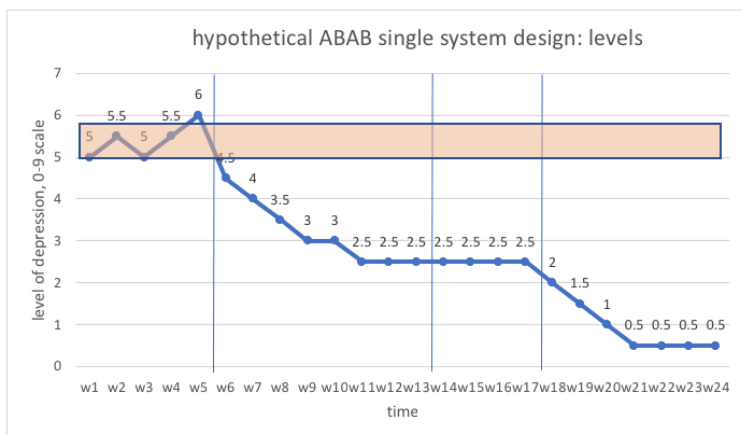
First, compute the mean and standard deviation for the baseline (A_1) phase data. Then, compute the values for the mean \pm SD (mean plus SD and mean minus SD). Using Excel, we find the following:

$$M=5.4$$

$$SD=0.42$$

$$2 \text{ SD band}=(4.98, 5.82)$$

Drawn onto the original levels graph, the two-standard deviation band graph would appear as:



As you can see, more than two (in fact, all) of the intervention phase data points fall outside of the two-standard deviation band width. The conclusion would be that a significant degree of change has occurred between these two phases, baseline (A_1) and intervention (B). To see this in action, work the exercise presented in your Excel Workbook.

Interactive Excel Workbook Activities

Complete the following Workbook Activity:

- [SWK 3402.4-3.2 Analyzing Single Systems Design Data](#)

Stop and Think



Take a moment to complete the following activity.



An interactive or media element has been excluded from this version of the text. You can view it online

here:

<https://ohiostate.pressbooks.pub/swk3402/?p=341>

Chapter Summary

This chapter concerned approaches to analyzing data collected through single system design efforts to evaluate intervention outcomes. First, you learned about the visual and graphing approaches. You learned how social work professionals and clients can use levels, trends, variability, overlap, means, medians, persistence, and latency information in interpreting their outcomes data. In addition, you learned two statistical analytic approaches that might be used to determine the significance of observed changes in single system design data. The non-overlap of all pairs

(NAP) analysis utilizes a form of nonparametric analytic logic to determine if a null hypothesis of no meaningful change should be rejected—the test statistic was a Z score (the calculation and distribution of which we have not previously studied), and the criterion of $p < .05$ remained useful to informing the decision. Another easily computed statistic you learned about was the two-standard deviation band analysis which entails calculating the mean and standard deviation for baseline data, then computing the range of values encompassing two standard deviations above and below that mean. The decision concerning meaningful levels of change is based on whether there are at least two consecutive data points in the intervention phase that fall outside of the calculated two standard deviations band. This can be visualized on a graph, but does not need to be—it can be determined based on reviewing a table of values once the band width has been computed. This chapter concludes our module on analyzing data for understanding social work interventions.

Module 4 Summary

This module emphasized approaches to analyzing the kinds of quantitative data often collected in intervention and evaluation research to answer questions about change and outcomes related to intervention. The chapter began with a review of some univariate and bivariate analysis topics learned in our prior course, applying these topics to intervention and evaluation research efforts. In addition to the descriptive analyses, you witnessed how chi-square, t-tests, and analysis of variance (Anova) might be utilized in intervention and evaluation research to answer group comparison questions. You learned a new test appropriate for use with repeated measures (longitudinal) data: the paired t-test. The importance of using appropriate repeated measures analysis strategies for analyzing data collected longitudinally was emphasized, as assumptions about independence of “groups” data are violated in longitudinal designs. You were reminded of the possible importance of nonparametric analytic approaches (when parametric data assumptions are violated) and were briefly introduced to logistic regression for the specific situation when the dependent variable is dichotomous and the independent variable is numeric. The final chapter examined approaches to analyzing data collected through single-system designs. This brings us to the next, concluding module for our two-course sequence.

Module 4 Key Terms and Definitions

dependent variable: the “outcome” variable in intervention research, sometimes referred to as the “y” variable, hypothesized to be influenced by the independent variable.

independent variable: the “input” variable in intervention research, sometimes referred to as the “x” variable, hypothesized to influence the outcomes (dependent variable).

latency: the stretch of time before something occurs, such as an event or onset of change.

nonoverlap of all pair (NAP) analysis: a form of analysis based on all possible pairwise combinations of data points, comparing two phases of a single system design, resulting in a Z-statistic with a p-value that informs the decision to reject or fail to reject null hypothesis of no meaningful change/difference.

nonparametric analysis: statistical analyses not based on the same set of assumptions about the data that parametric analyses require.

omnibus test of significance: a statistical test that measures overall significance of a set of explanatory (independent) variables without distinguishing which one or ones contribute to the solution—additional post hoc analysis are needed to make the distinctions by individual variables.

paired t-test: a parametric statistical analysis where the dependent variable data were collected longitudinally at two points in time (or otherwise nonindependent on the independent variable), using the t-distribution and calculation of a t-statistic based on shared variance.

two-standard deviation band method: an approach to statistical analysis of single system design data comparing the baseline phase band of values two standard deviations above and below the

mean to the intervention phase where two or more consecutive data points falling outside the band indicate meaningful change.

x axis: horizontal dimension on a line graph (and some other forms of bivariate graphs)

y axis: vertical dimension on a line graph (and some other forms of bivariate graphs)

Module 4 References

- Danaee-far, M., Maarefvand, M., & Rafiey, H. (2016). Effectiveness of a brief home-based social work motivational intervention for male methamphetamine users in Tehran: A randomized clinical trial. *Substance Use & Misuse*, 51(14), 1863-1869.
- Engel, R.J., & Schutt, R.K. (2013). *The practice of research in social work, third edition*. Los Angeles: Sage.
- Epstein, E.E., Drapkin, M.L., Yusko, D.A., Cook, S.M., & McCrady, B.S. (2005). Is alcohol assessment therapeutic? Pretreatment change in drinking among alcohol-dependent women. *Journal of Studies on Alcohol*, 66(3), 369-378.
- Gottman, J. M., & Leiblum, S. R. (1974). *How to do psychotherapy and how to evaluate it*. New York: Holt, Rinehart & Winston
- Manolov, R., Lasada, J.L., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2016). Analyzing two-phase single-case data with non-overlap and mean difference indices: Illustration, software tools, and alternatives. *Frontiers in Psychology*, 7(32), 1-16.
- Nourbakhsh, M.R., & Ottenbacher, K.J. (1994). The statistical analysis of single-subject data: A comparative examination. *Physical Therapy*, 74(8), 80-88.
- Nugent, W.R. (2010). *Analyzing single system design data*. NY: Oxford University Press. Pocket guides to social work research.
- Toneatto, T., Pillai, S., & Courtice, E.L. (2014). Mindfulness-enhanced cognitive behavior therapy for problem gambling: A controlled pilot study. *International Journal of Mental Health & Addiction*, 12, 197-205.
- Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, R. (2016). Single case research: Web based calculators for SCR analysis, version 2.0. College Station, TX: Texas A&M University.
- Walsh, T., & Lord, B. (2004). Client satisfaction and empowerment through social work intervention. *Social Work in Health Care*, 38(4), 37-56.

MODULE 5

Module 5 Introduction

As we conclude our two-course sequence about research and statistics for social work, we work on skills related to how social work professionals present evidence about interventions to various audiences and constituents. For example, we may use evidence to help clients/client systems make informed decisions about the array of intervention options available to them. Or, we may use evidence to inform program staff, program/policy decision makers, and potential funders about “best practice” interventions. These presentation skills build on the data reporting topics explored throughout the two courses. We also revisit how to graphically display findings as related to intervention and evaluation research, as well as continuing to practice elements from the American Psychological Association (APA) style guidelines for professional writing. This module concludes the two-course sequence with summaries of topics presented in the two courses and the Excel workbook exercises, as well as some thoughts about the future.

READING OBJECTIVES

After engaging with these reading materials and learning resources, you should be able to:

- Formulate a strategy for coordinating information and evidence about interventions generated from a search of literature;
- Generate a structured outline of a report about intervention evidence;
- Describe what it means to identify as an evidence-informed social worker;
- Recognize major topics covered in this course;

- Define a plan for continuing education and how this relates to the professional Code of Ethics.

Module 5 Chapter 1: Presenting Evidence for Understanding Social Work Interventions

Throughout this two-course sequence we have explored evidence and its relationship to strong social work practice. You have learned about how to locate, analyze, compare, and generate evidence to inform and evaluate practice using a host of critical thinking skills. These processes are important aspects of thoughtful, reflective practice. This point is emphasized in the observation that in arenas where social workers often practice, a large, severe gulf exists between evidence and practice:

*“with the majority of services delivered in usual care settings having little or no relation to practice supported by research”
(Chorpita, Bernstein, & Daleiden, 2012, p. 470).*

Our clients deserve better than this; they deserve access to the best possible practices and services. Helping clients select from among options is informed by evidence, as well as the social worker’s experience and the client’s values and preferences. These three pillars of evidence-based practice place responsibility on social work professionals to present evidence in a manner that facilitates client understanding.

In this chapter you:

- Learn a format for coordinating evidence from the literature about intervention options;
- Review a format for presenting evidence from practice or program evaluation;

- Recall what it means to identify as an evidence-informed social worker;
- Review major topics learned throughout this and the previous course;
- Consider the future in terms of professional development.

Coordinating Intervention Evidence from Literature

Presented here is a 7-part format for coordinating intervention evidence located in the literature. This format is not necessarily appropriate for presenting complex information to clients, program leaders, policy decision makers, and potential funders. The actual presentation needs to be tailored to the audience (as learned in Module 5 of our first course).

- **Specify the intervention question:** Using COPES or a similar approach, clearly identify the practice question about which intervention evidence is sought.
- **Locate, review, and summarize available evidence:** Applying skills learned throughout these two courses, identify literature/evidence sources relevant to the specified intervention question. For each source, collect and record the following information.
- **Type of evidence:** Identify the intervention(s) the evidence concerns, the intervention questions the evidence addresses, and the approach(es) used to develop the evidence.
- **Generalizability:** Specify the population to which the evidence applies (i.e., the “sample,” “subjects,” or “participants” involved, and where limits to generalizability might exist.
- **Intervention elements:** Identify the theory or logic model underlying the intervention, critical elements of the intervention, and requirements for implementing the

intervention with fidelity (the who, how, when, where, what aspects).

- **Strength of evidence:** Specify whether the evidence is represented in single studies, studies with comparable or competing outcomes, systematic review, meta-analysis, and/or scoping review. Make a determination about strength of the evidence based on the design and analysis methods used to develop the evidence.
- **Conclusions:** Identify conclusions that are appropriately drawn from the evidence (and which conclusions are not supported by evidence).

This information can be organized in a table like this, adding rows as needed for additional sources evidence and for additional intervention options.

Intervention
Option 1:
(specify)

| Source | Evidence | Generalizability | Elements | Strength | Conclusions |
|--------|----------|------------------|----------|----------|-------------|
|--------|----------|------------------|----------|----------|-------------|

Intervention
Option 2:
(specify)

| Source | Evidence | Generalizability | Elements | Strength | Conclusions |
|--------|----------|------------------|----------|----------|-------------|
|--------|----------|------------------|----------|----------|-------------|

Intervention
Option 3:
(specify)

| Source (citation) | Evidence | Generalizability | Elements | Strength | Conclusions |
|----------------------|----------|------------------|----------|----------|-------------|
|----------------------|----------|------------------|----------|----------|-------------|

Analysis Report for Audiences

The evidence table generated through the previously described activities may or may not be appropriate for presenting to the intended audiences. The following 5-part outline describes what might be included in a report tailored to specific audiences (clients, colleagues, program administrators, policy decision makers, or funders).

This type of five-step summary helps the social worker organize a complex body of information. In this way, social workers support informed choices.

- *Part 1.* Clearly state the intervention question being addressed.
- *Part 2.* Summarize the evidence reviewed (as in the previous section).
- *Part 3.* Present a relevance analysis. This analysis is about assessing the available evidence in terms of how well it applies to the specific client, client system, or population identified in the first step. This goes back to the generalizability issue and comparability of the research participants or samples to the clients for whom intervention questions are being asked. This might include assessing demographic and situational characteristics, such as:

“presenting problem(s), age, gender, ethnicity, or clinical service setting” (Chorpita, Bernstein, & Daleiden, 2012, p. 472).

- *Part 4.* Summarize implementation details/elements, costs, benefits, and feasibility of each analyzed option. Remember that cost/benefit analysis is not simply about financial costs and savings; important aspects concerning quality of life, time and effort expended/saved, goodness-of-fit dimensions (including “cultural” relevance with culture broadly defined) are also important aspects. Feasibility involves professional

competencies and training for providing the intervention with sufficient fidelity, as well as other required resources (time, space, tools). Feasibility also addresses fit with professional ethics, regulations, policies, and billing/funding criteria that might be involved.

- Part 5. Identify outstanding, remaining, uncertain, or unanswerable questions about the intervention evidence gathered.

The format for such a report should be clearly structured, following a logical outline. The 5-part list above could serve the purpose of structuring the outline.

Stop and Think



Take a moment to complete the following activity.

Visit the website for [Practicewise.com](https://www.practicewise.com), an interactive site synthesizing a vast amount of evidence concerning mental health treatment options for children, adolescents, and their families. Many millions of dollars in funding supported the development of the PracticeWise contents and tools—the result far exceeds what any one practitioner could be expected to generate. Practicewise is a fee-for-use service for practitioners in different disciplines who

provide mental health services to children, adolescents, and their families. Since we are not PracticeWise members, we cannot use the services but there are important lessons to be learned from reviewing the PracticeWise service offerings.

1. View the overview video in the PWEBS Database when you choose the “Our Services” menu. This 4-minute recording talks about how the literature is summarized for practitioner use in treatment planning and decision-making with clients. What does this recording tell you about the kinds of information you want to elicit from the literature and present to your audiences?
2. Select the PracticeWise Practitioner Guidelines next for the demonstration of practice guides and process guides. Try entering a search command like “Anxiety” and see what comes up as Practice Guide options. In the Process Guide menu, try entering the search command for “Diversity” and see what comes up. (You will not be able to actually open the .pdf files, unfortunately, since you are not a PracticeWise member. They include detailed, step-by-step guidelines for practitioners to follow.)
3. Returning to the main menu, select MATCH to see what it is about. What does the decision-tree/flowchart approach suggest to you about organizing treatment options?

Presenting Evidence from Practice Evaluation

In addition to learning how to identify and critically review existing evidence to inform social work practice, you also developed a set of skills related to evaluating practice. In this section we look at different ideas for presenting the evidence that you have generated through your evaluation efforts.

Client Reports

The format of your feedback to clients depends, to a great extent, on the clients themselves—what they already believe and understand about the practice question, their expressed preferences, aspects of their specific circumstances, and the circumstances of the feedback situation. The information needs to be tailored to clients' cognitive abilities for processing the information (e.g., age, cognitive impairment, emotional state, and more).

A generic framework for presenting/discussing evaluation results with clients might include the following:

- Specification about the variable(s) measured in the evaluation effort.
- Specification about the strengths and limitations of the measurement tool(s) used in the evaluation effort (reliability and validity, as well as sensitivity to change measurement).
- Presentation of the evidence/data and how the results might be interpreted.
- Specification about the strengths and limitations of the evaluation design—how this might influence the conclusions drawn from the evaluation effort.
- Conclusions/recommendations developed together with the client(s) based on the evaluation results, your practice expertise, and their preferences.



Formal Reports

In your evaluation work, you may be called on to present results to professional audiences—colleagues, agency or program administrators, community leaders, policy decision makers, or funders. In our first course you learned about the structure of written and presented research reports:

- Abstract or Initial Summary
- Introduction
- Methodology
- Results or Findings
- Discussion or Recommendations

In presenting evidence from your own intervention or evaluation research efforts for understanding intervention, these elements

remain an excellent outline. What you have learned throughout this and the prior course have provided you with the necessary knowledge and skills to create such a report or presentation. Review the topics presented in Module 5 of our first course related to making strong presentations, including how to create graphs, charts, figures, tables, and infographics.

As a reminder, the Social Work Code of Ethics emphasizes that it you need to ensure that individual clients/participants are not identifiable in any data files or summary reports that you share with others.

5.02.n Social workers who report evaluation and research results should protect participants' confidentiality by omitting identifying information unless proper consent has been obtained authorizing disclosure.

Not only are we concerned about the obvious identifiers (name, address, phone numbers), we are concerned about the ways that individuals' demographic data could be assembled to make an individual identifiable. For example, the combination of information about ethnicity, age, and gender might make an individual stand out and become identifiable to an audience, particularly in a small population or sample.



Module 5 Course Conclusions

Let's review the major topics covered in this second in the 2-course sequence:

- *Research and Statistics for Understanding Diverse Populations and Social Work Problems, and*
- *Research and Statistics for Understanding Social Work Interventions.*

Overall: Embracing Evidence-informed Social Work.

An article by Ferguson (2017) described reflective practice as an ideal and core concept across professions, represented by practitioners thinking about what they are doing while they are doing it, and subsequently thinking about what occurred and how their practice linked to knowledge. The concept of engaging in evidence-informed practice presented throughout this two-course sequence is based on the premise that research and statistics evidence are used to inform and evaluate practice.

Corey Powell (2011) observed that inquisitive souls exist in our world, individuals with “a ready curiosity, a desire to gather empirical evidence, a willingness to get dirty in order to put a theory to the test.” These individuals have the right instincts for science, he says, and are identifiable through their reactions to a “Wet Paint” sign: they have the impulse to check if the paint really is wet. Ideally, our coursework has stimulated or reinforced your curiosity and enthusiasm for critically thinking about evidence. Remaining curious, asking questions, and locating and evaluating evidence are important qualities for social work professionals.



Module 1. Introduction and Context of Social Work Intervention Research.

The first module in our second course began by drawing connections between the topics of the first course, *Research and Statistics for Understanding Diverse Populations and Social Work Problems*, and our second course, *Research and Statistics for Understanding Social Work Interventions*. The key point was that the skills and knowledge learned in our first course all apply to the skills and knowledge developed in our second course. What we understand about the people we serve, the social problems we address, and the social phenomena involved are all critical underpinnings to the interventions we apply. Next, Module 1 explained what is meant by “social work intervention” at multiple levels in complex, dynamic biopsychosocial systems. The relationships between evidence and intervention were

demonstrated: evidence informs intervention and evidence is used to evaluate (and improve) interventions. You learned to distinguish between evidence-informed practice, evidence-based practices (EBPs), and the evidence-based practice (EBP) process.

Next, Module 1 reviewed the types of questions asked in evaluating practice. Flowing from the research questions, you learned about different formats or approaches to evaluating practice: needs assessment, outcome evaluation (for practice, programs, and policy), process evaluation, and cost-effectiveness evaluation. Finally, Module 1 identified and explored ethical concerns that commonly arise in intervention research—not only do general professional ethics and research ethics concerns apply, there are several that arise in the specific context of intervention research. These related to issues of consent to participate, relationships between research and intervention activities, designing studies with control groups, and what happens to participants when the research phase ends.

Module 2. Identifying and Using Evidence to Inform All Levels of Social Work Practice.

This second course module specified 6 steps involved when social workers engage in the evidence-based practice (EBP) process: specify the practice question, identify best evidence, critically appraise the evidence, integrate evidence with practice expertise and client preferences/circumstances, take appropriate action, monitor and evaluate outcomes and process. You learned specific skills related to each of these steps, beginning with use of the COPES (and PICO) framework for formulating practice questions. You learned about the important role of logic models and the role of theory to intervention planning, and what evidence about mediators mean in applying theory to intervention planning (or evaluation). An entire chapter extended what you learned in the first course

about locating evidence in the literature and emphasized the value of systematic, scoping, and meta-analytic reviews for practitioners. The next chapter provided direction in terms of what to look for when reviewing research reports about interventions, including information about assessing appropriateness to the practice decision at hand, and concluded with a transition to implementing, monitoring, and evaluating the intervention plan. The process was demonstrated through an example addressing high school dropout/retention rates.



Module 3. Social Work Intervention Research Designs and Methods.

In both courses, the third module was about research approaches and study design. In this course, you learned that intervention and evaluation research is conducted through qualitative, quantitative, and mixed methods approaches depending on the nature of the research or evaluation questions: formative, process, intervention

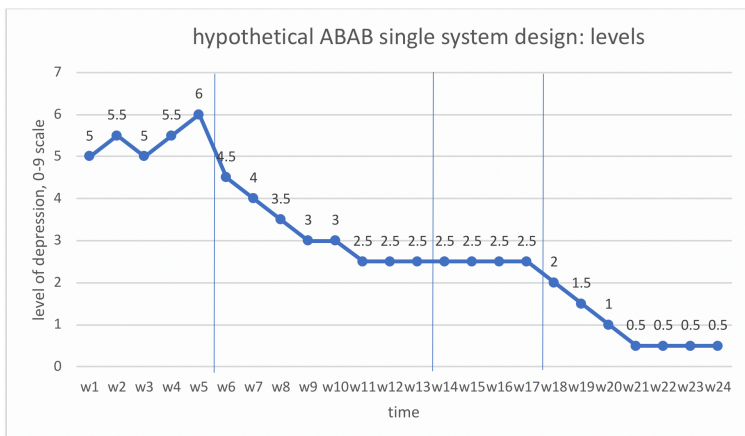
fidelity, outcome/impact, or cost evaluation. Because intervention and evaluation questions typically involve indicators of change, longitudinal designs are commonly preferred over cross-sectional designs. The remainder of the first chapter in Module 3 discussed various strategies for data gathering: naturalistic observation, participatory observation, participatory action research, and community-based participatory research (CBPR). This conversation continued in Chapter 3 which addressed various sources of data, including administrative data, secondary data, and primary data collection efforts. In addition, you learned in this module about measurement issues related to different variable types and a study's unit of analysis. Challenges associated with measurement in intervention and evaluation research included validity and reliability for measuring change, measurement sensitivity and specificity of clinical tools used in intervention and evaluation research, and concerns about clinical versus statistical significance in results.

Quantitative designs were presented in terms of capacity for preserving internal validity with the least control attributed to pre-experimental designs, more control quasi-experimental designs, and the greatest control with experimental designs. Using Xs and Os, as well as Rs, As and Bs, you were presented with 10 commonly employed intervention and evaluation research design options. Each was presented with an example and discussion of its advantages and disadvantages. The module concluded with a discussion of participant recruitment, retention, and sampling issues involved with intervention and evaluation research.

| | | | | |
|---|----------------|------------------|----------------|----------------|
| R | O ₁ | X ₁ | O ₂ | O ₃ |
| R | O ₁ | X _{TAU} | O ₂ | O ₃ |

Module 4. Understanding Basic Inferential Statistics Used in Social Work Intervention Research.

Our fourth module built on what was learned in the fourth module of our first course—statistical analyses. Intervention research often involves univariate descriptive, bivariate, and hypothesis testing (inferential) statistics. We revisited the univariate analytic tools for describing central tendency features of data (mean, median) and data distribution (range, variance, standard deviation). We then reviewed how a single-sample *t*-test might help answer an intervention or evaluation research question concerning changes in values—zero difference in scores or scores at a known threshold value. Next we reviewed several bivariate, hypothesis testing analyses for their applicability in intervention and evaluation research: independent samples *t*-test, analysis of variance (Anova), and chi-square. An entirely new statistical test was introduced in this module: the paired *t*-test for analyzing longitudinal data. This was necessitated by the paired, non-independent nature of the two “groups” of data—all being from the same individuals at two different points in time. Non-parametric and logistic regression strategies were briefly introduced, as well. Finally, you learned different ways to analyze and interpret data collected through use of a single system design. These strategies involved graphing the data and assessing change in outcome variables associated with changes in intervention conditions: changes in levels, trends, variability, means and medians, overlap, and latency/immediacy of change. In addition, you learned about the “nonoverlap of all pairs” and “two-standard deviation band” methods of hypothesis testing with single system data. Single system analysis skills were demonstrated through the hypothetical example with depression data.



Module 5. Presenting Intervention Evidence and Course Conclusion.

This brings us to our final module in the two-course sequence. The content learned in Module 5 of our prior course applies to how social work professionals present evidence to different audiences. In addition, the skills learned throughout the two courses about the relevance and uses of evidence in practice apply. Practical tools for presenting evidence to audiences were introduced. By successfully completing this two-course sequence, you have better prepared yourself as a social work professional and to contribute to the profession's continued development. This brings us back to a quote presented in the first module of the first course, attributed to Sir Isaac Newton:

“What we know is a drop, what we don’t know is an ocean.”



A Concluding Note About the Future

As you celebrate concluding this two-course sequence, and perhaps prepare for graduation from your social work program, it is important to consider how your understanding and appreciation of research and statistics will continue to develop. The Social Work Code of Ethics includes an ethical principle stating that:

“Social workers practice within their areas of competence and develop and enhance their professional expertise”
(<https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English>)

This principle is at the heart of the expectation that you will continue to engage in lifelong learning and professional continuing education activities throughout your career. The Code of Ethics clearly states one of our ethical responsibilities as professionals as:

4.01.b Social workers should strive to become and remain

proficient in professional practice and the performance of professional functions. Social workers should critically examine and keep current with emerging knowledge relevant to social work. Social workers should routinely review the professional literature and participate in continuing education relevant to social work practice and social work ethics.

To help us achieve this standard, the National Association of Social Workers (NASW), state chapters of NASW, agencies, your *alma mater*, and practice-area specialty groups host conferences, workshops, and continuing education training opportunities. By attending these events and continuing to locate and review professional literature, you remain informed about emerging evidence to inform practice; you may even be involved in training others based on the expertise you develop. As a result, you are working to meet the expectation expressed in the Code of Ethics stating that:

4.01.c Social workers should base practice on recognized knowledge, including empirically based knowledge, relevant to social work and social work ethics.



Stop and Think



Take a moment to complete the following activity.

As you finish the two-course sequence, it is important to recognize (and applaud) your accomplishments. Think about your own personal answers to these questions.

1. Comparing yourself to your classmates, how many and what pieces of information would make you identifiable as an individual—what would someone need to leave out of a report so that you could not be identified as a client/participant in an intervention or evaluation study?
2. Thinking back and comparing yourself at the start of the first course and the end of this second course, what are the 3 most important differences in your attitudes, knowledge, and skills regarding research and statistics as applied to social work practice?
3. Recognizing that knowledge in the profession is always emerging and evolving, what is your personal commitment to and plan for continuing professional education and lifelong learning?

Module 5 References

- Chorpita, B.F., Bernstein, A., & Daleiden, E.L. (2012). Empirically guided coordination of multiple evidence-based treatments: An illustration of relevance mapping in children's mental health services. *Journal of Consulting and Clinical Psychology*, 79(4), 470-480.
- Ferguson, H. (2017). How social workers reflect in action and when and why they don't: The possibilities and limits to reflective practice in social work. *Social Work Education*, 37(4), 415-427.
- Powell, C.S. (2011). Science with a common touch. *Discover* (Oct. 19). Retrieved from <http://discovermagazine.com/2011/oct/13-science-with-a-common-touch>