Reinforcing RCTs of Public Health Interventions: Defining and Linking Form and Function

by

Lisa Guttentag Lederer

BA Harvard University 2004

MA University of Pittsburgh 2012

Submitted to the Graduate Faculty of the

Department of Behavioral and Community Health Sciences

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Lisa Guttentag Lederer

It was defended on

March 30, 2020

and approved by

Robert Coulter, PhD, MPH, Assistant Professor of Behavioral and Community Health Sciences
Graduate School of Public Health, University of Pittsburgh

Ada Youk, PhD, Associate Professor of Biostatistics, Epidemiology, and Clinical & Translational Science
Graduate School of Public Health, University of Pittsburgh

Thesis Advisor:
Mary Hawk, DrPH, Associate Professor of Behavioral and Community Health Sciences
Graduate School of Public Health, University of Pittsburgh

Reinforcing RCTs of Public Health Interventions:
Defining and Linking Form and Function

Lisa Guttentag Lederer, MPH

University of Pittsburgh, 2020

Abstract

According to its stated core values, the American Public Health Association (APHA) is committed to policies and practices based on "evidence that demonstrates effectiveness." The meaning of this phrase, however, leaves much room for debate, and this paper's Public Health significance lies in its contribution to such a debate. Elaborating on its assertion about evidence, the APHA warns researchers "not to take statistics at face value." Accordingly, Public Health researchers have adopted a range of approaches for bolstering the statistical outputs of Randomized Controlled Trials (RCTs) -- though these results are frequently seen as the "gold standard" of medical and policy evidence. Not all of the approaches are equal, however, in their ability to integrate statistical results into actionable conclusions. This paper, after tracing the origins of the outsized prominence of RCT results in Public Health -- despite the APHA's warning -- and locating the central weakness of these results to their limited portability between settings, evaluates the approaches that have been proposed thus far. Surveying different disciplinary perspectives shows that, while many researchers have attempted to isolate and investigate the *context* of Public Health interventions, truly bolstering RCT results requires researching the operations of the interventions themselves. The optimal approach, it is concluded, demands creating a multi-level definition in which mechanisms are outlined at each level, including that of the participant.

Table of Contents

List of Tables

Preface

This thesis is a position paper inspired by a longstanding methodological debate on how to study health and social care interventions. The debate has roots in the writings of John Stuart Mill (Mill, 1846) but continues today -- even appearing, in lively form, at the 2019 Conference on Dissemination & Implementation in Health. I first became interested in it as a graduate student in the History and Philosophy of Science, but the questions involved have popped up in my cross-disciplinary work using methods from qualitative evaluations to RCTs.

This paper would not have been completed without the insightful comments of Faculty and Staff at the Mental Illness Research and Education Clinical Center at VA Pittsburgh, especially Gretchen Haas, Gloria Klima, and Keri Rodriguez; of the VA Implementation Science Research Group, particularly Christine Kowalski, Borsika Rabin, and Russ Glasgow; of Alexis Kirk and Monicà Perez-Jollez; or of my thesis committee members Robert Coulter and Ada Youk. It would not even have been started without the generous support of my supervisor at VA Pittsburgh, Adam Bramoweth; the incisive reflections of my Public Health advisor, Mary Hawk; or the emotional support of my husband Josh or the tolerance of my son Jonah, who (sometimes) allowed me to use Mama's computer for something other than watching Cookie Monster videos.

1.0 Introduction

On the American Public Health Association (APHA) webpage titled "our values," one finds the assertion that "the best policies and practices are ones based on research, with evidence that demonstrates effectiveness." Defining what type of information counts as "evidence," then, and how to "demonstrate effectiveness" -- regardless of whether this evidence is to serve as the foundation for Evidence-Based Policy, Evidence-Based Programs, or Evidence-Based Practices -- is essential for the field. This position paper argues that, in the crowded field of Public Health research approaches that aim to harness the power of Randomized Controlled Trials (RCTs) without sacrificing real-world applicability, the *function/form* approach stands out for its conceptual and practical soundness. After recounting, in Section 2, the history of the association between RCTs and "evidence that demonstrates effectiveness," in Section 3, the paper puts Public Health interventions in the same category – *complex interventions* -- as other healthcare interventions requiring qualitative investigation. In Section 4, techniques for researching such interventions by supplementing them with evidence about their context are reviewed, and their weaknesses noted; Section 5 describes two research approaches, including the *function/form* approach, aimed at defining interventions themselves. Section 6 provides further evidence that this approach can accommodate the real-world Public Health interventions. Section 7 concludes that funders should follow the Patient-Centered Outcome Research Institute in requiring researchers to describe their interventions in terms of functions and forms.

2.0 EBM, RCTs, and Public Health: A Brief History

Due to a dearth of methodological debate in Public Health, the definition of "evidence that demonstrates effectiveness" has partly been determined by historical accident. The race within many disciplines to become "evidence-based" can be traced to the *Evidence-Based Medicine* (EBM) movement launched by Gordon Guyatt and colleagues, who in 1992 introduced a "new approach to teaching the practice of medicine" by that name in the *Journal of the American Medical Association* (EBM Working Group, 1992). Shortly thereafter, a paper by the EBM Working Group in the same journal, titled "A method for grading healthcare recommendations," featured a hierarchy on which studies evaluating medical interventions could be rated for their evidential strength (Guyatt et al., 1995). The promulgation of this, and of similar hierarchies by research organizations dedicated to making "evidence-based" recommendations (e.g., The Cochrane Collaboration, The Campbell Collaboration) has continued ever since; today, such hierarchies typically place meta-analyses of Randomized Controlled Trials (RCTs) at the top, with single RCTs immediately below, and "clinical expertise" at the bottom, meaning that conclusions reached through properly conducted RCTs should be trusted far more than those reached through clinical expertise.

Like other medical fields, Public Health seeks a solid evidential basis and has aligned itself with the evidence-based medicine movement accordingly. Not long after the EBM evidence hierarchy had gained prominence in Public Health, however, researchers in that field, particularly in its community health and health promotion branches, began to see the apparent boost in objectivity delivered by EBM as a bias towards evidence from RCTs at the expense of other valuable evidence types. Ziglio, for example, representing the World Health Organisation's

International Working Group on Evaluating Health Promotion Approaches, stated in 1997 that "[s]earching for effectiveness of health promotion could be assessed . . . against criteria that include matters of equity, empowerment, sustainability, accountability, acceptability, fiscal feasibility, amongst others. This would be a very different type of evidence-based intervention from what we can borrow from evidence-based medicine" (Ziglio, 1997, p. 32). Indeed, under the EBM evidence hierarchy, an intervention suggested by qualitative data from patients to contribute to patient empowerment would not be considered "evidence-based" at all. More recently, despite agreeing with the overall values of the EBM movement, researchers have objected to the applicability of RCT evidence across community settings. In the *New England Journal of Medicine* in 2017, Frieden still found it necessary to "describe the use of RCTs and alternative (and sometimes superior) data sources from the vantage point of public health, illustrate key limitations of RCTs, and suggest ways to improve the use of multiple data sources [besides RCTs] for health decision making" (Frieden, 2017, p. 465).

Even for these critics, however, it is not the evidential strength of RCT results that is problematic, but researchers' lack of consideration of the fit between study design and conclusions. Rychetnik, Frommer, Hawe, and Shiell (2002), considering whether Public Health interventions should be evaluated using the same criteria as those used in clinical practice (i.e., when treating or preventing illness in individuals, as opposed to when protecting health or preventing ill health in communities or groups) argue that "criticisms of the RCT are based on a consideration of 'classic' RCTs in which the intervention is standardized and the individual is the unit of randomization . . . [Other types of] RCTs have a long history of successful application in evaluating the effectiveness of social interventions" (Rychetnik et al., 2002, p. 121). Even harsh critics of RCTs as currently conducted have conceded that, despite serious limitations, they "provide the strongest evidence

about the causal effects of social interventions" (Bonell et al., 2012). The upshot of the methodological debate surrounding RCTs in Public Health is that, although the results flowing from RCTs should not *dictate* community health practice, RCT results -- when interpreted with caution -- provide evidence that *facilitates* practice by aiding practitioners in weighing the strengths of alternative programs and practices.

In these debates, proponents of RCTs have typically pointed to their tendency "to evenly distribut[e] known and unknown factors among control and intervention groups, reducing the potential for confounding" (Frieden, 2017, p. 32). Often left implicit has been the essential *cumulativeness* of the EBM/RCT approach: evidence about a given intervention accumulates as it is subject to repeated trials. Michie, Fixsen, Grimshaw, & Eccles, advocating the use of RCTs for evaluating "Behaviour Change" interventions (including many Public Health interventions) do note, in the section called "the advantage of reporting interventions better," that it is *replication* based on detailed RCT reports that "generates scientific knowledge" and "allows unhelpful or even harmful interventions to be avoided," but these authors fail to note that this cumulativeness is especially beneficial -- even crucial -- in clinical and health science fields, where we have incomplete knowledge of the domain, and trials are always susceptible to unexpected external influences. After all, although the hallowed $p<.05$ figure references the chances that trial results were happenstance, the more "unknowns" are lurking in a given domain, the greater the likelihood that a "significant" result was not actually due to the expected (or, in most cases, manipulated) factor. Even for relatively simple healthcare interventions, the continuity of the EBM approach is needed to ensure that an evidence base is built around a given intervention beyond the flimsy conclusions of a single trial. When evaluating Public Health interventions with RCTs, though, accumulating multiple trials of the "same" intervention, even if these trials nominally involve the

same actions on the part of practitioners, may be like aggregating apples and oranges. That is, the actual intervention – i.e., what is intervened upon or manipulated, or the independent variable – may vary between trials.

In fact, this cumulativeness problem with RCTs is felt across human service and healthcare disciplines that feature complex interventions. Depending on their disciplinary background, some researchers concerned with this issue have presumed that an Evidence-Based "core" which defines the intervention already exists – or awaits theorizing in other fields – and have thus debated *how to describe the context* surrounding this core. Others have presumed that "the" intervention that is transferred between practitioners necessarily includes core and context together and have debated *how much* local practitioners can adapt. The following sections argue that both debates are misdirected; a more fruitful debate would focus on *how to describe the boundaries of the interventions themselves* in a comprehensive yet replicable way. Then, using the example of a complex intervention that combines Public and Behavioral Health, it offers practical advice for the pursuit of such definitions.

3.0 Public Health Interventions as Complex Interventions

In fact, the number of healthcare disciplines facing a growing need for assessment strategies besides RCTs, for the reason indicated above, is steadily increasing. Ironically, perhaps due to the increasing application of Bronfrenbrenner's social-ecological model of human development to all of human health (McLaren & Hawe, 2005) healthcare interventions have grown in complexity and advanced (Dunn et al., 2019) but have simultaneously suffered from the EBM movement's restrictive view of evidence. When weighing the applicability of RCT evidence to Public Health questions, researchers could benefit from considering perspectives on evaluating complex interventions from other medical fields.

In 2000, recognizing the challenges of researching complex interventions, the Medical Research Council (MRC, the body responsible for coordinating and funding medical research in the UK) published a "Framework for the development and evaluation of RCTs for complex interventions to improve health" (updated in 2008). The MRC defines "complex interventions" as "interventions with several interacting components such as occur in health service, public health and social policy" (6) and elaborates by listing relevant dimensions:

- Number of and interactions between components within the experimental and control interventions [in trials of the intervention]
- Number and difficulty of behaviours required by those delivering or receiving the intervention
- Number of groups or organisational levels targeted by the intervention
- Number and variability of outcomes
- Degree of flexibility or tailoring of the intervention permitted (Craig et al., 2008, p. 7)

Areas of medicine like Public Health, Behavioral Health, and Health Services are replete with interventions that are complex along a number of these dimensions. For example, the intervention known as Brief Behavioral Treatment for Insomnia (BBTI), which combines

Behavioral Health and Public Health elements, was adapted from the intervention known as Cognitive-Behavioral Treatment for Insomnia (CBTI) to overcome "numerous system-, provider-, and patient-level factors [that] contribute to the gap between the high prevalence of insomnia and the relatively low use" of the latter treatment at VA hospitals (Bramoweth et al., 2018). BBTI earns a high score on all these dimensions. The treatment is a combination therapy that includes potentially interacting techniques that rely on multiple theoretical mechanisms, both neurobiological and behavioral. During the relatively brief course of treatment, the therapist must not only motivate patients to engage in multiple sessions but also to undertake difficult behavior changes that may themselves interact, such as inducing short-term sleep deprivation and avoiding their bed at all times outside a prescribed "sleep window." The desired long-term outcomes of BBTI include not only better sleep, but also improved psychiatric and even metabolic function among recipients. The healthcare organization into which BBTI is introduced must devote time and resources to train clinicians in the technique and educate Primary Care providers about its availability so that they can make appropriate referrals.

In fact, the above example highlights how the situation presented by healthcare interventions such as BBTI is even more intractable than the MRC definition of complexity allows. Further features that are intuitively *complex* include:

- Openness to influence at many levels (rather than mere *existence* at many levels – e.g. fluctuations in the patient's immediate social sphere, or in the political atmosphere that underlies funding, may have profound effects on impact)
- Number of external barriers to individual patients (e.g. burdensome travel requirements in a certain region, unfamiliarity with the concept of preventive medicine among a certain population)
- Scope of what "counts" as the intervention to healthcare organizations (e.g., an EBI described vaguely as a "multi-component, interdisciplinary intensive primary care program" might be "interdisciplinary" in that it involves cooperation between two medical departments in the same hospital, or between hospitals and research institutions); and
- Existence/entrenchment of alternative treatments in the community at large.

Before providing the above list of defining characteristics, the 2008 MRC guidelines explain their rationale for doing so, naming problems that commonly beset researchers of complex interventions attempting to build an evidence base through RCTs. These include "standardising the design and delivery of the interventions" and "the length and complexity of the causal chains linking intervention with outcome" (Craig et al., 2008, p. 6). More elegantly, a 2010 paper in the *Journal of Clinical Evaluation* by Cartwright and Munro, "The limitations of randomized controlled trials in predicting effectiveness," states the single assumption to which these problems can be traced – and suggests that they are not limited to health science fields:

> A properly conducted RCT provides evidence that the intervention works somewhere (i.e. in the trial). The decision maker, however, needs to estimate 'will it work for us?' In health and social care the underlying social and physical structures in which an intervention is devised cannot automatically be assumed to be comparable to target localities in causally relevant aspects . . . (Cartwright & Munro, 2010, p. 265)

The assumption pinpointed by Cartwright and Munro as faulty is that certain physical/social features of the setting where a given "health or social care intervention" has proven effective are similar, in some "right" but undefined way, to those where a decision-maker aims to transport the intervention. It is, in effect, the "cumulativeness" problem described in Section 2 -- the only difference being that the above quote refers to multiple settings in space, while the cumulativeness problem as described in the Introduction refers to multiple settings in time. In both cases, the factors indicating complexity named above (including both those named by the MRC and those added) would influence transfer between settings, either hampering the intervention or inflating researchers' and practitioners' impressions of its effectiveness. In implementing a complex intervention, implementers may engage in prescribed acts that they expect -- or hope -- to have certain effects, but the interventions interact with the setting, are full of changeable, and generally obscure, features that stem from the "underlying physical and social structures" referred

to by Cartwright and Munro; no number of RCTs is capable, without further evidential support, of indicating to a decision-maker whether or not a complex intervention that was beneficial in one setting -- or even many settings -- will "work for them." According to the Centre for Evidence Based Medicine, an RCT can, theoretically, provide evidence for the intervention's *efficacy* -- whether it works "under the ideal conditions of an investigation" (Last, 2001) -- conditions that hardly ever exist in the real world. But the RCT cannot, by itself, provide evidence for the intervention's *effectiveness*, or whether it works "under usual conditions of clinical care for a particular group" (ibid.).

The MRC guidelines imply that researchers can "patch up" RCTs of complex interventions, enabling practitioners in the real world to learn from them, by providing contextualizing details. The next section explores how similar advice has been presented from multiple perspectives. It then reveals how this advice is misguided.

4.0 Bolstering RCTs of Complex Interventions: Supplemental Approaches

Facing a world full of unstable social and physical structures, Cartwright and Munro argue that, for RCTs to provide the robust evidence sought by Public Health researchers, "much more evidence, and much different in kind, is required" (Cartwright & Munro, 2010, p. 4). In fact, though, the need to contextualize quantitative results from social science experiments with *qualitative* data has been recognized for over half a century. Most social scientists would say that this is a problem of *external validity*; as introduced by Campbell and Stanley in 1959, external validity "asks the question of *generalizability*: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" (their emphasis, 5). In the health sciences, achieving gains in external validity usually requires loosening control over such variables as population and setting, which often leads to corresponding decreases in *internal validity*: "Did in fact the experimental treatments make a difference in this specific experimental instance?" (Campbell & Stanley, 1959, p. 5). Because internal validity, as these authors note, is "the basic minimum without which any experiment is uninterpretable" (5) health scientists are understandably hesitant to bolster external validity in exchange, and persuading them to attend to external validity at all has required considerable effort (Green & Glasgow, 2006).

Noting that additional evidence is needed, though, is not enough. Over the past few decades, researchers have proposed a plethora of methods for producing the necessary kind of evidence, many of which have simply involved providing supplementary details or measures. The strategies proposed for "cushioning" RCTs in this way fall into two general categories: those guided by *intuition/theory* and those guided by *pre-existing lists*.

4.1 Theoretical/Intuitive Approaches

According to the MRC, the role of adding the kind of evidence that renders RCT results sufficiently robust can be fulfilled by *process evaluation*. When nested inside an experimental design like an RCT, the potential of such evaluations is wide – they "can be used to assess fidelity and quality of implementation, clarify causal mechanisms, and identify contextual factors associated with variation in outcomes" (3) – but apparently only includes "clarifying," rather than defining, the intervention's operative causal mechanisms. That is, process evaluations are intended to provide evidence that is supplementary to, but strengthens, the central assessment by RCTs.

The follow-up document to the MRC guidelines that provides guidance specific to process evaluation allows that, in determining what questions should drive such evaluations, researchers may be guided by a hodgepodge of "social science theory" and "other factors such as past experience or common sense" (Moore et al., 2015, p. 1). Problematically, however, researchers have been pointing out the foibles of such "common sense," including that of medical professionals, for decades (Engelhardt et al., 1969); the biases that randomization serves to reduce, for example, may creep into the way evaluators carve out subgroups for analysis. Furthermore, the guidance implies that, even if evaluators strictly pursue the issues highlighted by a social science theory, they may help themselves to a broad range of such theories, being careful only to "avoid focusing narrowly on inappropriate theories from a single discipline. For example, psychological theory may be useful for interventions that work at the individual level but is less useful when intervening with organisations or at wider social levels . . ." (4). Such a free-ranging choice leaves room for bias beyond what an RCT would permit. This "hodgepodge" approach may be effective for local program evaluations that are not meant to support generalized inferences, but is not appropriate when findings are reported alongside supposedly generalizable RCTs. Besides

providing the opportunities for bias noted above, process evaluations may vary by local setting according to fundamental issues, such as which factors merit investigation. Even the ideal process evaluation, unlike the *ideal* RCT, will often generate entirely non-portable lessons.

In fact, Munro and Bloor's experience attempting to integrate a process evaluation with an RCT confirms the difficulty predicted above. In response to an apparent process evaluation craze sparked by the 2000 MRC advice, these Public Health researchers warned in 2010 that "it is important that process evaluations are not oversold: they are not a miracle ingredient" (Munro & Bloor, 2010, p. 708). They explain the perils of assuming that qualitative interviews in a process evaluation – which are designed to shed light on the factors practitioners and participants consider significant -- will straightforwardly improve future implementations of the intervention.

Munro and Bloor's trial involved an evaluation of a secondary school intervention employing "peer supporters to have informal conversations with classmates about the dangers of cigarette and cannabis smoking. Considering the process evaluation results, the researchers reflect that based on the "general agreement, among peer supporters in all the focus groups and across the different schools that cannabis was more difficult to talk about than cigarette smoking." Their comments suggest that the reluctance to discuss cannabis is a significant factor causing a reduction in the number of cannabis-related conversations; a reasonable response would seem to be re-designing the intervention to remove "the element of discretion for peer supporters that allows them to concentrate on prevention of cigarette smoking" alone (Munro & Bloor, 2010, p. 701).

Quantitative results from the RCT, however, suggest a different story: despite a significant difference between the reported number of conversations about cigarettes and the reported number of conversations about cannabis in one school receiving the intervention (23% vs. 9%) there was no such difference at another school (34% vs. 27%). Although it is not a given that the RCT results,

rather than the process evaluation results, tell the "true" story, these researchers' experience does show that even for people intimately involved in the intervention, identifying the factors an RCT would suggest to be causally relevant (such as teenagers' apparent reticence about marijuana) is not an intuitive task.

Furthermore, while the causally relevant factors in complex interventions vary between settings, the *intuitively* causally relevant factors (though they are not necessarily the same) also vary, leaving researchers and practitioners with a sea of (possibly conflicting) papers to read in preparation for implementation. Munro and Bloor predict that the project of transporting complex interventions among settings by integrating process evaluations with RCTs is doomed by practical considerations. The MRC and those who echo their advice, they argue

> are looking for an awful lot of bangs for their buck . . . these complex interventions are frequently multi-site trials, involving perhaps a score or more of clinics (or communities, or schools, etc.) and perhaps an equal number of intervention delivery teams . . . Enormous effort expended, mountains of complex data collected, and no earthly chance of making any sense of it all. (Munro & Bloor, 2010, p. 4)

Munro and Bloor's observations suggest that intuitive approaches to supplementing RCT results face insurmountable problems: the guidelines provided to implementers are too fuzzy to reliably result in datasets that are both manageable and relevant.

4.2 Pre-Existing List Approaches

Accordingly, some researchers have attempted to focus the data that accompanies RCT results by prescribing what qualitative information should be collected and reported. Notably, some such prescriptions are unrealistic. All journals published by BMC, for example, such as

Implementation Science, "strongly recommend" that authors refer to the minimum reporting guidelines listed on the website hosted by the EQUATOR network (Enhancing the Quality and Transparency Of health Research; https://www.equator-network.org/). For social and psychological interventions, the category into which most complex health interventions would fall, the checklist of items or features to be reported simply demands "sufficient details to allow replication" (Movsisyan et al., 2019, sec. eligibility criteria). Because *perfect* replication is an impossibility -- at the very least, a new trial will focus on a new population of individuals – this requirement seems unattainable.

Where this strategy has paid dividends in healthcare research is in areas characterized by "simple" interventions, such as pharmacy, where researchers have attended to a predetermined list of "external" factors such as age, sex, and disease severity. In such areas, this *listing* strategy typically reveals what contexts are appropriate for an intervention because, if important differences exist between the reactions of different individuals to a given drug, they usually align with these features.

For areas of healthcare characterized by complex interventions, however, where the number of significant features is practically infinite, no such list exists (though researchers have repeatedly attempted to construct one). Implementation Science researchers, for example, have often "carved off" the downstream, or implementation, phase and enumerated "implementation factors" that tend to be causally relevant across interventions. These lists may sometimes be appropriate in new contexts – but a standardized approach to replication does not guarantee that the *right* features – or, to use Cartwright and Munro's language, the *causally relevant* factors – will be captured. For example, historical events like a war or election may influence a behavioral health intervention by affecting the psychological well-being of the entire country, but historical

factors do not often appear on lists of commonly significant features. The long and motley list of "contextual" factors that may impact intervention success presents organizations attempting replication with a daunting task; even organizations with the time to delve into the literature and select factors relevant to their context are unlikely to have the foreknowledge to do so. Supplementing the results of RCTs of complex interventions by describing the features that are *commonly* relevant is little better than the intuitive approach of supplementing them with whatever descriptions *seem* relevant.

Instead of generating competing lists of potentially significant "external" factors, or assuming that implementers' intuition will guide them in selecting the relevant factors, implementation science researchers could build an evidence base about *what* is internal and *what* is external to a given intervention. Because it is sometimes assumed that elaborate descriptions of the "outside" of an intervention can be pulled from, pending future theorizing, to amend the definition of the "inside" (i.e., the parts of the intervention that are transferred between organizations and are not adaptable) these tasks of *describing* and *defining* are not generally seen as directly opposed. While researchers of "simple" healthcare interventions, though, can often look to biomedical theory to determine inside and outside, complex intervention researchers have no well-defined boundaries to guide them; so it is up to them to discover whether seemingly contextual features are actually essential. Taking advantage of the cumulativeness of evidence from RCTs – that is, comparing apples to apples only – requires continually conducting research to find the definition of a given complex intervention, or in more pragmatic terms, to determine what should be transferred between researchers. The final sections trace the evolution of how researchers have defined this inside of interventions, and then explain how a relatively recent approach to definition, the *core functions and forms* approach, is superior to previous approaches.

5.0 The Core Wars

References to an intervention "core" often arise in health science fields today in debates about procedures for intervention *adaptation*: the modification of components outside this core (Movsisyan et al., 2019). Though the topic of healthcare intervention adaptation only "came to the fore" in 2000, however (ibid., "eligibility criteria" section) the concept of an intervention "core" is rooted in a much earlier controversy. Researchers in human service fields outside of healthcare, fields that are traditionally more cognizant of social and communal factors, have long struggled with issues of intervention complexity. In these fields, the issue analogous to *how to define successful interventions* is *how to spread successful service programs*. In the 1980s and 1990s, a burgeoning literature existed on the *technology transfer* of human service programs between organizations. Such "technologies" are similar in many ways, and thus face nearly the same roadblocks, as do today's complex healthcare interventions; whether based in research funded by the Department of Health and Human Services -- and thus labeled "human service programs" -- or by the National Institutes of Health -- and thus labeled "healthcare interventions" – they are multi-component, rely on multiple societal levels, and ultimately aim to improve lives.

Though the technology transfer debate centered around the extent to which adaptation by local organizations should be permitted or conversely, how extensive the unchangeable "core" should be (Blakely et al., 1987). Many human service technology transfer researchers recognize the importance of pinpointing "crucial elements" (Embry, 2004) "evidence-based kernels" (Michie et al., 2009) or whatever their preferred term may be for the defining parts of a human service technology; but the term, and accompanying conception, that has ultimately taken hold is one proposed by Michie, Fixsen and colleagues (D L Fixsen et al., 2005, p. 24). Crucially, they

grounded this *core components* concept by drawing extended parallels with "Industry Research," and then extended its reach by applying it to both human service technologies and healthcare interventions. For example, in one of the earliest guidelines on Dissemination and Implementation in the NIH National Information Center (NICHSR) core components are defined as "the most essential and indispensable components of an intervention practice or program" (Blase & Fixsen, 2013, p. 3). The authors then quote an industry research journal article, elaborating that core components specify "which traits [of consumer goods, or in this case, healthcare interventions] are replicable, how these attributes are created, and the characteristics of environments in which they are worth replicating" (ibid). In a later research brief for the Office of Human Services Policy about core intervention components, Michie and Fixsen add that they are "principles" that are "operationalized" and can be either "theory-driven" or "empirically derived;" specifying further, they state that core components are those that "are intended to, or have been demonstrated through research to, positively impact the proximal outcomes that address the identified needs and that increase the likelihood that longer-term outcomes will be achieved" (Dean L Fixsen & Blase, 1993, p. 598).

Working to clarify the "core components" definition of a given intervention, i.e. what is internal to the intervention, represents a significant advance over supplementing the presumed intervention with qualitative data. This is because definitions based on core components can theoretically include a constellation of seemingly "contextual" features, enabling them to be replicated without presenting practitioners with the task of hypothesizing which features of the intervention-context complex are significant. As in the case of process evaluations, however, researchers have an impossibly broad field of candidate components to designate as *core*. Though Fixsen and colleagues stress that the list of *core components* must encompass those factors that

are theoretically relevant, and not merely tangible or salient (Michie et al., 2009) the mere fact that an intervention component (such as the provision to consumers of volunteer activities) is included in a documented theory does not preclude its inclusion in many other, conflicting theories.

Furthermore, the *core components* concept is, by Fixsen and colleagues' own admission, a near-replica of the "Arrow core" concept attached to the information economist Kenneth Arrow (an early paper of theirs asserts that "we have much to learn from industry on the subject of creating realities" for human service consumers; (Fixsen & Blase, 1993). While the neoclassical economic theories from which the *core components* concept derived are about series of *transactions*, complex health interventions are about webs of *interactions*. This difference means that while the "core components" of economic theories are *causes* that slot into linear equations, the core components in theories of healthcare interventions should be both *causes and effects*, subject to forces arising from feedback loops and nonlinearities.

Focusing on Public Health, Hawe, Shiell, and Riley (2004) thus questioned the very nature of this core. Under the *core components* conception, the social programs (or interventions) themselves were still viewed as "consisting of a number of relatively well-specified [core] program components" (Blakely et al., 1987, 255) that were separable for the purpose of measuring fidelity to the initial model*; program fidelity*, as Blakely and colleagues explain in their brief history of the fidelity-adaptation debate, "could then be defined as the number or proportion of finite program components that were implemented" (ibid.). Hawe and colleagues objected to this definition. Complexity Science, they protested, tells us that a complex system (or intervention) is more than the sum of its parts; reducing it to its components changes its identity (i.e. alters what is measured and adapted.) In "How out of control can a randomized controlled trial be?" they ask

> could a controlled trial design (which requires something to be replicable and recognisable as the intervention in each site) ever be appropriate to evaluate a

> (truly) complex intervention? The answer is yes. The crucial point lies in "what" is standardised. Rather than defining the [core] components of the intervention as standard . . . what should be defined as standard are the steps in the change process that the elements are purporting to facilitate or the key *functions* that they are meant to have . . . these [functions] could then take on different *forms* according to local context, while achieving the same objective (Hawe et al., 2004, pp. 1561–1562).

Viewing interventions through the lens of Complexity Science, as do Hawe, Shiell, and Riley, reminds us that the intervention core must include (initially) downstream factors that are hard to predict due to feedback loops and nonlinearities. In other words, expecting interventions that only include a *part* of the core components to achieve the intervention goals *in part* would be like expecting partial success from a surgical procedure that included *some* of the crucial steps (e.g., removal of the appendix) but not all of them (e.g., stitching up the surrounding skin afterward). This objection is the converse of the objection, discussed above, that researchers will often *include too many* core components in their definition: in the case of public health interventions, components that are essential to intervention success -- such as stigma-free accessibility -- are often *missed*. As Cartwright and Munro's quote about underlying social and physical structures suggests, a theory is no match for all the elusive components of a complex intervention.

Hawe, Shiell, and Riley thus present arguments based in Complexity Science for a new approach to healthcare "technology" (intervention) transfer: the *function/form* approach; yet their insightful argument leaves much to be spelled out in terms of practical application. The section below provides a more thorough exploration of their approach that goes beyond the discipline of Complexity Science.

6.0 Identifying Core Functions and Forms,

The pragmatic value of Hawe, Shiell, & Riley's (2004) functions/forms approach depends on whether it can guide researchers and implementers in their real-world endeavors. Admittedly, the approach might go against researchers' and implementers' initial EBM-based impulses, by asking them to treat two intervention implementations that differ in obvious ways as the "same" intervention, including, even, aggregating evidence for them. After all, two drugs (for example) with different physical *forms* (i.e. chemical structures) would usually be considered different.*things.*

There is no rule, however, that the core intervention functions, or what is compared in each trial, must be structural or even sensible. While the core components approach encourages researchers to seize only upon the most visible or salient aspects of an intervention, identifying core functions/forms is a comprehensive process that involves integrating the perspectives of a variety of stakeholders. Most importantly, as explained below, this approach leaves room for the multiple levels of abstraction that, as emphasized by Cartwright, characterize the practical application of scientific theories.

The tendency of the core components approach to under-inclusion, and of the disadvantage of squeezing all essential intervention ingredients onto a single conceptual level, has been illustrated by an actual attempt at enumerating the *core components* of CBTI and BBTI (Cognitive Behavioral Treatment or Insomnia and Brief Behavioral Treatment for Insomnia, respectiverly). For decades, *fidelity scales* have been used by psychotherapy researchers – and more recently, by human service program researchers – to "assess the adequacy of implementation of [psychotherapeutic or human service] program models" (Bond et al., 2000, p. 75; for a history of

fidelity scales, see (Mowbray et al., 2003). Fixsen and Blase vaguely insist that such measures "do not necessarily tell the whole story about what is required for effective use of an intervention in typical service settings" (Blase & Fixsen, 2013, p. 5) In practice, however, researchers often see such scales explicitly as lists of core components and neglect to go further in defining their intervention of interest.

During their national "rollout" of CBTI (Cognitive Behaioral Treatment for Insomnia) at VA medical centers, for example, Karlin et al. (2013**) constructed fidelity scales intended to measure therapist faithfulness to the core of the intervention/treatment. Components to be assessed for their presence in therapist treatment delivery include: evaluation of the patient's "sleep diary;" mention and exposition of stimulus control principles; mention and exposition of the principles of sleep restriction therapy; use of cognitive therapy and "guided discovery"; attention to patient adherence; attention to hyperarousal; assignment of homework; interpersonal effectiveness; collaboration, provision and elicitation of feedback; application of cognitive and behavioral components to te patient's particular case; and overall competency (Karlin et al., 2013). Therapists are rated on each item using a 4-point rubric. To adapt the scale for use with the briefer behavioral treatment for insomnia (BBTI) other researchers (e.g Bramoweth, *) have simply removed items, such as "use of cognitive therapy and guided discovery," referring to the therapist's direct efforts to alter patient cognitions. Thus, while the CBTI scale includes 11 items, the BBTI version includes only 9.

In contrast to componential approaches, however, the function/form approach, as elaborated by Hawe, Shiell, and Riley (2009) advocates viewing Public Health interventions as "events in [complex] systems." Adopting this view implies recognizing the intervention setting and target population, usually considered merely contextual features of the intervention, as

fundamental to its definition -- and as fundamental to the theories on which it is based, and that may arise from it. The CBTI and BBTI fidelity scales reflect intervention functions only from the perspective of the clinician, or at the provider level; this restricted perspective is reflected in the fact that in the above scales, CBTI purportedly includes more core components than does BBTI. In fact, the latter intervention—which takes into account "contextual" elements such as patient preference for brevity – is theoretically more complex. As elaborated below, however, in identifying the functions and forms of CBTI/BBTI, we would also look upward – to neurobiological theories such as theories about sleep restriction, and to social psychology theories such as theories about stigma -- and downward, to the patient-level accounts necessary to fill in gaps concerning the intervention's form (i.e. its mechanism of action). In short, the relative simplicity of the "checklist" approach to ensuring fidelity to an intervention's core often comes at the expense of under-inclusion and overinclusion of crucial components.

(Kirk et al., 2019) and (Perez Jolles et al., 2019) each exemplified the combined upward-looking and downward-looking strategy by applying it to an actual healthcare intervention. The functions and forms they used in defining their interventions of focus, as well as theoretical examples given by Hawe, Shiell, and Riley, are listed in Table I. Their choices of function and form are further detailed in the subsections below.

Table 1: Functions and Forms from the Literature

Paper	Representative Function	Representative Form
Hawe, Shiell, & Riley (2004)	Distribute information on depression tailored to local literacy, language, culture, and learning styles	Patient information kit on depression written by each site
Kirk et al. (2019)	Do not lead the conversation by mentioning hospice, start the conversation [about end-of-life care] by discussing care goals, needs, and preferences.	Framing, as prescribed by each site, of exact script/wording detailed in intervention protocol
Perez-Jolles et al. (2019)	Offer enhanced options for access to in-person care	In-person care available outside of traditional business hours

6.1 Identifying Core Intervention Functions

The role of the *form*/*function* dichotomy in the theories of various scholarly fields, from Architecture to Disability Studies, is the subject of a rich philosophical literature (Nanay, 2010); but a difference along the *concreteness-abstractness* is essential throughout. In the Hawe, Shiell, and Riley (2004) model, intervention components identified at a relatively high level of abstraction that is characteristic of researchers' theories belong in the *function* category. Hawe and colleagues, describing a fictitious intervention, name a sole function; Kirk et al. and Perez et al., who both apply the form/function framework to real-world complex interventions, each name several functions per intervention. In all cases, though, features in this category are closely tied to the intervention's motivating theory or theories. In identifying core *functions*, Perez-Jolles et al. (2019) referred to the relevant literature (e.g. journal articles) while, due to an absence of relevant

literature in their case, Kirk et al. (2019) interviewed the authors of the article introducing their chosen intervention.

In outlining a step-by-step process for identifying core forms and functions *post hoc* (i.e., after the intervention has been in practice for some time) Kirk *et al.* (2019) equate "function" with "purpose." Similarly, Perez-Jolles et al. (2019), applying the function/form framework to an ongoing intervention, state that *functions* are "the intended structural and procedural goals and purposes to reach the intervention goals" (1034). Even in scholars' interpretation of Aristotle, though, the equation of function with purpose has been criticized as too simple (Lacks & Morin, 1992); the *function* concept features feasible action. *Peace*, for example, could not be a function; *fostering constructive dialogue* could. Importantly, furthermore, *functions* evolve (Acosta et al., 2014); they maintain a degree of flexibility as is necessary for interventions arising from scientific theories that themselves constantly evolve, although intervention functions remain constant across multiple contexts.

In fact, the simultaneous existence of multiple functions, as in the Kirk et al. (2019) and Perez-Jolles et al. (2019) interventions, provides another reason – left implicit by Hawe, Shiell, and Riley (2004) -- for viewing components as a whole instead of assessing them piecemeal: functions stemming from different theories may work together synergistically. For example, BBTI combines Sleep Restriction Therapy, which relies on the neurobiology of short-term sleep deprivation, with Stimulus Control, which relies on principles of behavioral conditioning (i.e., associating the bed with a single activity: sleep). Each of these treatments has been shown to be effective individually (Movsisyan et al., 2019) but they operate best in tandem: restricting one's sleep window increases adherence to Stimulus Control treatment by reducing one's opportunity to engage in non-sleep activities (such as watching TV) in bed, while eliminating time spent awake

in bed (i.e. "lolling around") increases adherence to Sleep Restriction Therapy by reducing one's opportunity to inadvertently fall asleep. *Restricting the patient's sleep window* and *eliminating time the patient spends awake in bed* could thus be considered synergistic functions of BBTI.

Both of the above functions, moreover, sit comfortably with another potential function for BBTI that is associated more with Public Health than Behavioral Health research. This potential function, due to the location of BBTI in the Primary Care wing rather than the Behavioral Health wing, is: *decreasing stigma associated with seeking treatment for Behavioral Health.* Note that the three BBTI functions mentioned above all exist at varying levels of abstraction or universality: Neurobiological principles predict that *all patients* would be affected similarly by Sleep Restriction Therapy; the Behavioral strategy of strengthening the association between the bed and sleep will work only for *patients with consistent sleeping-places*; and the stigma associated with seeking Behavioral Health treatment has been found to be most significant, in the US, among *veteran patients* (Cartwright, 2010). Multiple levels of abstraction can thus co-exist within the function category, reflecting the multiple levels at which scientific theorizing may take place.

6.2 Identifying Core Intervention Forms

The split between form and function is pragmatic: forms, unlike functions, are constrained by the resources available to local implementers and the idiosyncrasies of the population of participants in their intervention. As Perez-Jolles et al. (2019) put it, forms are "specific strategies or activities that may be customized by local contexts that are needed to carry out core functions" (1033). Forms not only *may* be customized but *should* be customized in accordance with local factors, as studies have repeatedly shown that at least some degree of local adaptation is not only

inevitable but advantageous (Damschroder et al., 2015). To identify core *forms*, Perez-Jolles et al. (2019) culled descriptions from study reports of the actual shape taken in different contexts by their chosen intervention; Kirk et al. (2019) simply consulted the original intervention protocol. Just as forms are expected to vary widely between different context, different methods may be optimal to elucidate functions and forms, depending on the available literature, implementers' familiarity with it, etc.

An important source of information on intervention form, however, was absent from the Perez-Jolles et al. (2019) and Kirk et al. (2019) studies: neither sought out the experiences of intervention participants, whose activities outside the medical office and personal observations, after all, represented an important part of the change process meant to be initiated by the interventions. This absence may reflect an assumption that patient reports would not correspond to intervention functions in a reliable way. Once again, Cartwright's metaphors about the application of theory to the real world are apt here: she stresses the importance of "building ladders" between abstract theories and concrete realities (Machamer et al., 2000, p. 3).

An example of how to use qualitative data from intervention participants fruitfully comes from a hybrid clinical trial combining an RCT comparing BBTI and CBTI with qualitative interviews focused on implementation factors (Bramoweth et al., unpublished data) from the Consolidated Framework for Implementation Research (Harvey, 2002). In response to the questions *Did [the intervention] work?* and *Why?* -- expected to be introductory, rather than to yield reportable data – we received rich accounts detailing how the participants believed treatment components had operated. Importantly, furthermore, these accounts often fit neatly into the mechanisms hypothesized by scientists to underlie the treatments. (Note that "mechanism" is used here in the general sense of Machamer, Darden, and Craver [2000]: "entities and activities

organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Bond et al., 2000) It does not presuppose, as it seems to in Kirk et al. [2019] the higher level of theorizing typically associated with *functions*.)

For example, in their accounts of intervention successes, patients repeatedly mentioned (unprompted) the benefits of reviewing their "sleep diaries," or daily logs of sleep patterns, with providers each week. Their assessments suggest that, at least for BBTI and CBTI at the VA Pittsburgh Medical Center, *seeing documented progress* and *receiving benign explanations* for their condition contributed to patients' improvement. Representative quotes include:

> Veteran: And, I think the fact that the person sees how effective it is…
> Interviewer: Um-hum.
> Veteran: …you know, by keeping the diaries.
> Interviewer: Right.
> Veteran: That really helps because they're seein', if they look back, they see progress.

> Veteran: It really helped me to understand…
> Interviewer: Um-hum.
> Veteran: …well, if I guess if it would be anything to dislike is that I still don't sleep straight through. I still wake up [but] I kind of understand how that happens and why that happens.

These accounts fit nicely with the mechanism outlined by Harvey (Bond et al., 2000) in her Cognitive Model of Insomnia. Patients' *beliefs,* modified by the documented evidence of the sleep diaries and explanations, no longer lead to *worry* – in effect breaking the arrow in the model between those constructs – and thus no longer contribute to *selective attention and monitoring* or, ultimately, to the *misperception* that Harvey hypothesizes lies at its root. Thus although *reducing cognitive arousal* might be listed as a *function* of BBTI/CBTI, to be standardized among all sites where they are practiced, an investigation into intervention *forms* -- which might include participant interviews like the ones described above -- would likely be a Quality Improvement project taking place at the facility level.

7.0 Conclusions

The practical upshot of this paper is that Public Health researchers would benefit both from defining novel interventions in terms of functions/forms, and from viewing existing interventions in function/form terms when planning implementation. Ultimately, though, it is up to funders to encourage the widespread adoption of the function/form approach in Public Health research. The Patient-Centered Outcomes Research Institute (PCORI) has led the way in this endeavor, specifying in its methodological standards statement online that researchers conducting PCOR must

> describe the intervention and comparator under study and clearly define aspects related to core functions and forms. Core functions refer to the intended purpose(s) of the interventions. The form of the interventions includes the intended modes of delivery, providers involved, materials or tools required, dose, and frequency/intensity. The description should also explicitly indicate to whom the intervention is aimed (e.g., patient, provider, hospital, health system).

While this interpretation of Hawe and colleagues' suggested approach does not capitalize on its basis in Complex Systems -- e.g., does not encourage researchers to explore multiple perspectives in defining functions, or to highlight potential synergies between them – it prompts researchers to see beyond a restrictive "fidelity" lens.

This paper is not intended to "solve" all the dilemmas posed when Public Health researchers attempt to apply RCT evidence to complex interventions. Remaining questions include how to assess fidelity – an essential procedure in constructing an evidence base (Bond et al., 2000) – of *functions*, given that local facilities' and organizations' directly measurable embodiments of these functions will vary in *form*. Another unanswered question is: at what point does so much evidence in favor of a particular *form* exist that it is recognized as a universal *function*? For

example, the “sleep diary” provider conversations mentioned above might be irreplaceable by other concrete forms of the hypothesized general function *reducing cognitive arousal.* In that case, though, a more specific and actionable function than the above would be *documenting daily sleep patterns*. If similar data is obtained in other facilities’ qualitative interviews with BBTI/CBTI patients, studies comparing treatments with and without the *documenting* function would provide evidence in this matter.

This paper also brings multiple disciplinary perspectives to debates related to complex healthcare interventions, clarifying these debates and assessing the arguments that have been advanced. It argues that our current incomplete understanding of Evidence-Based Interventions (or Practices or Policies) requires researchers to exercise humility in investigating them and to take care in bequeathing, to future implementers, actionable knowledge grounded in the patient voice.

Bibliography

Acosta, J. D., Becker, A., Cerully, J. L., Fisher, M. P., Martin, L. T., Vardavas, R., Slaughter, M. E., & Schel, T. L. l. (2014). *Mental health stigma in the military*. Report RR426, RAND Corporation. https://www.rand.org/pubs/research_reports/RR426.html

Blakely, C. H., Mayer, J. P., Gottschaik, R. G., Sehmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The Fidelity-Adaptation Debate: Implications for the Implementation of Public Sector Social Programs. In *American Journal of Community Psychology* (Vol. 15, Issue 3).

Blase, K., & Fixsen, D. L. (2013). *Core intervention components: Identifying and operationalizing what makes programs work.* Policy Brief, U.S. Department of Health & Human Services. https://aspe.hhs.gov/report/core-intervention-components-identifying-and-operationalizing-what-makes-programs-work

Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, *2*(2), 75–87. https://doi.org/10.1023/A:1010153020697

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science and Medicine, 75*(12), 2299-2306. https://doi.org/10.1016/j.socscimed.2012.08.032

Campbell, D. T., & Stanley, J. C. (1959). *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company: Boston.

Cartwright, N. (2010). *Will this policy work for you? Predicting effectiveness better: How philosophy helps*. Presidential Address, Philosophy of Science Association.

Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, *16*(2), Issue 2, 260–266. https://doi.org/10.1111/j.1365-2753.2010.01382.x

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). *Developing and evaluating complex interventions*. Report, Medical Research Council. www.mrc.ac.uk/complexinterventionsguidance

Damschroder, L., Hall, C., Gillon, L., Reardon, C., Kelley, C., Sparks, J., & Lowery, J. (2015). The Consolidated Framework for Implementation Research (CFIR): progress to date, tools and resources, and plans for the future. *Implementation Science*, *10*(S1). https://doi.org/10.1186/1748-5908-10-s1-a12

Dunn, B. D., O'Mahen, H., Wright, K., & Brown, G. (2019). A commentary on research rigour in clinical psychological science: How to avoid throwing out the innovation baby with the research credibility bath water in the depression field. *Behaviour Research and Therapy, 120*. https://doi.org/10.1016/j.brat.2019.103417

Embry, D. D. (2004). Community-based prevention using simple, low-cost, evidence-based kernels and behavior vaccines. *Journal of Community Psychology*, *32*(5), 575–591. https://doi.org/10.1002/jcop.20020

Engelhardt, D. M., Margolis, R. A., Rudorfer, L., & Paley, H. M. (1969). Physician Bias and the Double-Blind. *Archives of General Psychiatry, 20*, 315–320.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A Synthesis of the literature. Report, Louis de La Parte Florida Mental Health Institute, The National Implementation Research Network. https://doi.org/10.1017/CBO9781107415324.004

Fixsen, Dean L, & Blase, K. A. (1993). Creating new realities: Program development and dissemination. *Journal of Applied Behavior Analysis*, *4*(26), 597–615.

Frieden, T. R. (2017). Evidence for health decision making-beyond randomized, controlled trials. *New England Journal of Medicine, 77*(5), 465–475. https://doi.org/10.1056/NEJMra1614394

Green, L. W., & Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Evaluation and the Health Professions, 29*(1), 126–153. https://doi.org/10.1177/0163278705284445

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., Cook, R. J. (1995). Users' Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. *JAMA: The Journal of the American Medical Association*, *274*(2): 1800-4. https://doi.org/10.1001/jama.1995.03530220066035

Harvey, A. G. (2002). A cognitive model of insomnia. *Behaviour Research and Therapy 40*(8), 869-93.

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, *43*(3–4), 267–276. https://doi.org/10.1007/s10464-009-9229-9

Kirk, M. A., Haines, E. R., Rokoske, F. S., Powell, B. J., Weinberger, M., Hanson, L. C., & Birken, S. A. (2019). A case study of a theory-based method for identifying and reporting core functions and forms of evidence-based interventions. *Translational Behavioral Medicine*, ibz178, https://doi.org/10.1093/tbm/ibz178

Lacks, P., & Morin, C. M. (1992). Recent Advances in the Assessment and Treatment of Insomnia. *Journal of Consulting and Clinical Psychology, 60*(4), 586-94.

Last, J. M. (2001). *A Dictionary of Epidemiology* (4th ed.). Oxford: Oxford University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. http://www.jstor.org/journals/ucpress.html.

McLaren, L., & Hawe, P. (2005). Ecological perspectives in health research. *Journal of Epidemiology and Community Health*, *59*(1), 6–14. https://doi.org/10.1136/jech.2003.018044

Michie, S., Fixsen, D., Grimshaw, J. M., & Eccles, M. P. (2009). Specifying and reporting complex behaviour change interventions: The need for a scientific method. *Implementation Science*, *4*(1). https://doi.org/10.1186/1748-5908-4-40

Mill, J. S. (1846). *A system of logic, ratiocinative and inductive, being a connected view of the principles of evidence and the methods of scientific investigation*. London: Harper & Brothers.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ (Online)*, *350*. https://doi.org/10.1136/bmj.h1258

Movsisyan, A., Arnold, L., Evans, R., Hallingberg, B., Moore, G., O'Cathain, A., Pfadenhauer, L. M., Segrott, J., & Rehfuess, E. (2019). Adapting evidence-informed complex population health interventions for new contexts: A systematic review of guidance. *Implementation Science*, *14*, 105 (2019). https://doi.org/10.1186/s13012-019-0956-5

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340. https://doi.org/10.1016/S1098-2140(03)00057-2

Munro, A., & Bloor, M. (2010). Process evaluation: The new miracle ingredient in public health research? *Qualitative Research*, *10*(6), 699–713. https://doi.org/10.1177/1468794110380522

Nanay, B. (2010). A modal theory of function. *The Journal of Philosophy*, *107*(8), 412–431.

Perez Jolles, M., Lengnick-Hall, R., & Mittman, B. S. (2019). Core Functions and Forms of Complex Health Interventions: a Patient-Centered Medical Home Illustration. *Journal of General Internal Medicine*, *34*(6), 1032–1038. https://doi.org/10.1007/s11606-018-4818-7

Rychetnik, L., Frommer, M., Hawe, P., Shiell, A., & Rychetnik, L. (2002). Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 56(2), 119-27.

Ziglio, E. (1997). How to move towards evidence-based health promotion interventions. *Promotion & Education, 4*(2), 29-33. https://doi.org/10.1177/102538239700400215