Influence of Contextual Elements on Comprehensive Evaluations of
Multi-Level Interventions: Lessons Learned and Promising Practices from
Cases in Tobacco Control and Childhood Obesity

By

Jenica Huddleston

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Public Health

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Thomas Rundall, Co-Chair
Professor Norman Constantine, Co-Chair
Professor Kathleen Metz

Fall 2010

Influence of Contextual Elements on Comprehensive Evaluations of Multi-Level
Interventions: Lessons Learned and Promising Practices from Cases in Tobacco
Control and Childhood Obesity

By Jenica Huddleston

ABSTRACT

Influence of Contextual Elements on Comprehensive Evaluations of Multi-Level Interventions: Lessons Learned and Promising Practices from Cases in Tobacco Control and Childhood Obesity

By

Jenica Huddleston

Doctor of Public Health

University of California, Berkeley

Professor Thomas Rundall, Co-Chair
Professor Norman Constantine, Co-Chair

There are large gaps in information currently about the way evaluations of multi-level interventions are conceptualized and implemented in public health practice. Existing evaluation literature provides little guidance to evaluators as they attempt to design and implement comprehensive evaluations of such interventions. Multi-level interventions that attempt to create change or impact at two or more levels are being developed and implemented with increasing frequency in several fields, including public health. However, these interventions are often not comprehensively evaluated. If evaluation is conducted, it is often piece-meal, and important elements such as data collection on all levels intervened are often lacking. The reasons behind this disconnect between the comprehensiveness of the intervention and evaluation is not clear.

This dissertation was conducted to fill a current gap in the fields of evaluation and public health. This study expands knowledge by having assessed current multi-level projects and their context, in order to better understand contextual elements described as influencing the comprehensiveness of their evaluations.

The author reviewed literature and current evaluation practice around multi-level evaluations and developed a Framework for Comprehensive Evaluation of Multi-Level Interventions (Framework). A multiple case study was conducted with four projects evaluating multi-level interventions in tobacco control and childhood obesity to further understand contextual elements influencing their evaluation. Analysis of case data took place individually for each project, across projects within the same field, and across the fields of tobacco control and childhood obesity. Case data were later compared with the Framework, to assess how contextual elements influenced evaluations of projects studied.

Cases studied provided information on strengths and challenges of their evaluations, sources of information used in the development of their evaluations, and contextual elements that influenced their evaluations.

Great overlap was found between cases and across the fields studied. These projects developed their evaluations using similar sources of information and expertise. Also, many similar elements were described as influencing the comprehensiveness of their evaluations. While the original Framework and hypotheses developed early in the study were not fully supported, case data provided valuable information on nuance around contextual elements, such as how different elements influenced evaluation development versus evaluation comprehensiveness.

Ultimately, the combination of having strong internal capacity for evaluation, access to external evaluation expertise, adequate and consistent funding for the evaluation, a logic model and other sources of information to guide the evaluation, and flexibility to modify the evaluation as needed, were found as key factors for the success and comprehensiveness of projects studied. This pattern was found across the cases and across both fields.

Field itself may not be a particularly distinguishing factor for cases studied in evaluation development, comprehensiveness, or contextual elements influencing these areas. Projects studied may be providing guidance to others based on their successes around evaluation comprehensiveness, expertise, and amount of time they have been funded and evaluating their programs, rather than their field. And it may be that childhood obesity as a newer field for multi-level interventions, and with large amounts of new funding, is looking to others, especially other programs with many years of multi-level intervention and evaluation experience for more guidance at this time.

Recommendations were developed to address challenges identified, promote promising strategies, and provide considerations for future work. Recommendations were developed to address areas including: capturing synergy, determining "who gets credit", funding issues, access to evaluators and improved training of evaluators, reporting and communication, building an evidence base, use of logic models, inclusion of external evaluation experts, building internal capacity, and flexibility to modify evaluations as needed.

Information gathered through this dissertation can be used by multiple audiences to improve evaluations of multi-level interventions in the future. Multi-level projects, such as those studied, are not carried out quickly or simply, so having multiple suggestions for improvement, to help with the development, comprehensiveness and other areas of the evaluation, can be quite helpful. The ultimate goal of this dissertation was to provide the fields of evaluation and public health with information to increase comprehensiveness and quality of evaluations for multi-level interventions in the future.

Future work beyond this dissertation can go in several directions, including advancing recommendations provided, further study of elements within the Framework, or studying contextual elements beyond those in the Framework, which may also influence evaluation comprehensiveness. Additionally, future studies could look at other cases within these fields, or fields beyond tobacco control and childhood obesity, that are carrying out multi-level interventions, to see if similar patterns are found.

DEDICATION

To Grandma Jane, for your everlasting love and support.

ACKNOWLEDGEMENTS

Thank you to my dissertation committee members, Dr. Tom Rundall, Dr. Norm Constantine, and Dr. Kathy Metz. I appreciate your thoughtful comments, suggestions that helped to push my thinking further, and reading several drafts of a long dissertation! I have learned a tremendous amount from working with each of you. Norm, I am grateful for your guidance throughout this program.

I greatly appreciate the participation of projects studied. Thank you to those interviewed, and especially the project contacts, for your time and willingness to share information.

To my DrPH 2007 cohort peeps, I have learned so much from all of you and am thankful for this great support system. What an interesting group of people that bonded in such a special way. The happy hours may have helped some too ☺ One of the best experiences of this program has been working with other students. I continue to be inspired by the insights and accomplishments of my DrPH colleagues. Daniela and Kristina, you kept me smiling and motivated when I needed it most…thank you!

I have been lucky to work with several great mentors, but I especially appreciate Dr. Leslie Cooksy for planting the seed of an idea to go back to school for my doctorate. When I was done laughing, apparently I gave it some thought, and I'm glad I did.

The late Dr. Curtis Acredolo introduced me to the field of evaluation in early 2000 and my professional life has not been the same since. His trust in me to run a major project, then his evaluation company, inspired me to work hard and live up to the respect I was given.

To my friends and family, your support, understanding, and camaraderie have meant the world to me. I feel really blessed to have people in my life that I care so deeply about and that feel the same in return. I look forward to being back among you and not buried in work!

I'd especially like to thank my parents, who set in me an expectation from an early age to shoot for the stars, wherever that may take me. I truly believe that those early expectations, along with support and guidance, can greatly influence the direction of a child's life. But I also really appreciate that you let me find my own way and supported me fully in following my own path. To my brother and sister, I know I can be in your business and a pain, but you should know that I am so proud to have you as my siblings and find great joy in watching as you are now coming into your own.

Hayes, this process would not have been the same without your support, patience, and love. I am so thankful that you came into my life when you did and cannot express how much your encouragement, humor, and grilled cheese sandwiches have meant. Thank you!

Go confidently in the direction of your dreams. Live the life you have imagined.
-Henry David Thoreau

Table of Contents

CHAPTER 1: INTRODUCTION

Multi-level interventions that attempt to create change or impact at two or more levels, such as the individual, family, school, community, or policy levels, are being developed and implemented with increasing frequency in several fields within public health. However, these interventions are often not comprehensively evaluated. If evaluation is conducted, it is often piece-meal, and important elements such as data collection on all levels intervened are often lacking. In fact, descriptions of multi-level interventions and related evaluations often fail to make clear which levels of intervention are being targeted and whether comprehensive evaluation practices for multi-level interventions are being followed. The reasons behind this disconnect between the comprehensiveness of the intervention and evaluation is not clear. Currently, little is known about the way evaluations of multi-level interventions are conceptualized and implemented in public health practice. The existing evaluation literature includes few critiques of evaluation practice with respect to multi-level interventions and provides little guidance to evaluators as they attempt to design and implement comprehensive evaluations of such interventions.

This dissertation seeks to fill some of these gaps by assessing the fields of tobacco control and childhood obesity for examples of comprehensive evaluations of multi-level interventions. This dissertation addresses the following overarching research question: *What are key contextual elements, challenges encountered and promising practices for comprehensively evaluating multi-level interventions*? It will seek to provide a better understanding of contextual elements influencing evaluation comprehensiveness and strategies used by projects that were able to successfully create a comprehensive evaluation. The intent here is not to describe a specific methodology, tool or evaluation as the best or only way to proceed. Rather, the purpose of the current dissertation is to identify elements that are present in comprehensive evaluations of multi-level interventions and contextual factors influencing their inclusion. In-depth examination of several exemplar projects sought identify influential contextual elements, lessons learned, promising practices, and challenges encountered in comprehensively evaluating their multi-level interventions.

To be clear, the focus of this dissertation is on the <u>evaluation</u> of interventions taking place on multiple levels concurrently. It will be important to describe and understand the interventions, both theoretically and practically, but the ultimate focus of this work is on the lessons learned, promising practices and challenges of comprehensively evaluating such interventions. The current study explores comprehensive evaluations of multi-level interventions. Therefore, both the interventions and their evaluations are multi-level (they intervene and assess impact across multiple levels).

This dissertation is organized around several main sections. Each section serves a different purpose, answers somewhat different research questions, and utilizes different methodologies, but all tie in to the overarching purpose and research question for the dissertation. As an overview, the dissertation covers three main areas of inquiry: 1) development and description of a Framework for Comprehensive Evaluation of Multi-Level Interventions (Framework) and a review of recent and publicly available multi-level evaluations, 2) a multiple case study with four projects, with detailed presentation of cases and analysis across cases and fields, and 3) development of lessons learned, determination of

promising practices from cases studied, and recommendations to move the fields of evaluation and public health forward. Additionally, a background section is provided to describe foundational concepts around multi-level interventions and evaluations, and the fields selected for study, which are built on throughout the dissertation.

Multiple methods were used to carry out this dissertation study, including literature reviews, document reviews and interviews. Two literature reviews were conducted to: 1) build a Framework for elements in a comprehensive evaluation of multi-level interventions, and 2) review recent evaluations of multi-level interventions for many of these elements. After potential projects were identified through this second literature review, cases were selected for study. A multiple case study was conducted to enable comparison across cases and fields, to determine particular elements influencing comprehensive evaluation of multi-level interventions. Four cases were selected for study to increase the rigor compared to studying just a single case. The multiple case study included document reviews of internal and publicly available evaluation reports and other project documents, and interviews with multiple participants from each case studied.

Following are descriptions of each chapter in this dissertation, including the purpose, individual research questions, and methods used, as appropriate. These elements are described in further detail in each of the relevant chapters.

The first (and current) chapter of this dissertation provides an introduction to the dissertation, including the purpose, research questions, methods, and a brief description of each chapter to come.

The second chapter will provide background information on multi-level interventions and evaluations, as well as a discussion about why the fields of tobacco control and childhood obesity were selected for study. Significance of this dissertation is also addressed in this chapter.

The third chapter will identify elements of a comprehensive evaluation and contextual factors which may influence the comprehensiveness of evaluations for multi-level interventions in tobacco control and childhood obesity. A working Framework for comprehensive evaluations, with further discussion of each element included, is presented. Additionally, a review of recent evaluations of multi-level interventions in both tobacco control and childhood obesity was conducted and the findings discussed. This chapter addresses the following research question: *What are elements of a comprehensive evaluation for multi-level interventions in tobacco control and childhood obesity?* Data were collected for the third chapter through two literature reviews.

The fourth chapter describes a multiple case study with four projects evaluating multi-level interventions in tobacco control and childhood obesity. These cases were studied to better understand the contextual elements influencing the comprehensiveness of their evaluations. This chapter describes the rationale for conducting the multiple case study, selection of cases and interview participants, data collection, and analysis for the cases. Data were collected for the multiple case study through interviews and document reviews with each of the selected projects. The research question of focus for the multiple case study is: *What are contextual elements influencing comprehensive evaluation of multi-level interventions in tobacco control and childhood obesity?* Each case is presented individually in the fourth chapter, to

provide background on each project and to describe contextual elements identified by project representatives as influencing the comprehensiveness of their evaluation.

The fifth chapter provides an analytic look at case data within and across the fields of tobacco control and childhood obesity. Findings from cases studied are presented as strengths and challenges to their evaluations, sources of information guiding development of their evaluations, and contextual elements identified by projects as influencing the comprehensiveness of their evaluations. This chapter also compares data from cases to the Framework developed earlier in the dissertation.

The sixth chapter develops recommendations to increase comprehensiveness and quality of evaluations for multi-level interventions in the future. Lessons learned from cases studied and contextual elements identified by participants as influencing their evaluations are presented. Promising practices were determined from cases studied and recommendations developed, by the author, with the intent of translating data gathered from the cases into information that can be used by others. The sixth chapter addresses the following research question: *What lessons learned, promising practices, and recommendations can be derived from the study of cases and assessment of contextual elements, in order to improve evaluations of multi-level interventions?*

The dissertation in completed with a brief conclusion and description of limitations to the dissertation in the seventh chapter.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

In order to provide background for this dissertation, an overview is presented on several relevant areas, including: evaluation in public health, evaluation of multi-level interventions, why tobacco control and childhood obesity were selected for focus and what multi-level interventions look like in these two fields.

Evaluation in Public Health

Definitions of evaluation are nearly as varied as the topics they seek to study. Frequently, evaluation is used to make some sort of judgment about whether a program or intervention is effective or has merit, whether a program should continue to be funded, modified or discontinued, or how limited resources would be best spent (Doll, Bartenfeld, & Binder, 2003). Evaluations are also conducted to assess program activities and changes in outcomes, and determine whether the program can be credited for these changes. Lessons can also be learned from evaluations, to inform improvement of activities and programs. Other examples are described in further detail in Figure 1.

Figure 1. Reasons to Evaluate Public Health Programs
(U.S. Department of Health and Human Services, 2005)

- To monitor progress toward the program's goals.
- To determine whether program components are producing the desired progress on outcomes.
- To permit comparisons among groups, particularly among populations with disproportionately high risk factors and adverse health outcomes.
- To justify the need for further funding and support.
- To find opportunities for continuous quality improvement.
- To ensure that effective programs are maintained and resources are not wasted on ineffective programs.

The Centers for Disease Control and Prevention (CDC)(1999b) has stated that "Public health professionals can no longer question whether to evaluate their programs; instead, the appropriate questions are: What is the best way to evaluate? What is being learned from the evaluation? And, how will lessons learned from evaluations be used to make public health efforts more effective and accountable?" (pp. 31-32). While expectations for including evaluation may be increasing, the capacity for conducting good evaluations and having an audience for them is also critical, as echoed by Stufflebeam and Shinkfield (2007) "…society has a critical need not only for competent evaluators but for evaluation-oriented decision makers as well" *(*p. 5).

Evaluation of Multi-Level Interventions

Published literature, and the current study, define a multi-level intervention as one that attempts to create change or impact at two or more levels (Nastasi & Hitchcock, 2009).

In these types of interventions, activities often take place on multiple levels concurrently, such as the individual, family, school, community, or organizational levels, among others. Implementing activities on multiple levels are intended to create a change in social norms, expectations and environments around a particular topic, with the goal of ultimately improving health or other outcomes. A previously conducted systematic literature review of community-based prevention programs described intervention programs showing the most impact were those that emphasized changing social norms (Merzel & D'Afflitti, 2003). Multi-level interventions are gaining momentum in some preventative health fields, such as tobacco control, childhood obesity, mental health, sexual health, HIV/AIDS and cardiovascular disease, among several others (Kothari, Edwards, Yanicki, Hansen-Ketchum, & Kennedy, 2007; Nastasi & Hitchcock, 2009; Nutbeam, 1998; Trickett, 2009).

The focus of many past health interventions has been around changes in knowledge and individual behavior change, rather than systems change or intervening on multiple levels simultaneously, which are different innovative ways to impact changes in health outcomes. More comprehensive multi-level programs are gaining momentum, but they are new, not well funded, and evaluation is often not included or required. As described by Trickett (2009), "individual level change is difficult to sustain in the absence of environmental change supportive of such individual efforts" (p. 2). Trickett goes on to add:

> *"Multilevel interventions can, in principle, be designed to address the complex issue of how to create social contexts supportive of individual change efforts. In addition, social problems, such a teen pregnancy or HIV/AIDS, are issues affecting multiple sectors of the community that cut across ecological levels of the community context, including not only individuals but families, school policies and programs, and health departments. A multilevel perspective promotes the development of interventions that are responsive to community concerns beyond the reach and resources of any individual or single community organization."* (p. 2)

Previous literature and individuals working within the field have argued that intervening on multiple levels concurrently is important for addressing complex health issues (Centers for Disease Control and Prevention, 1999a; Nastasi & Hitchcock, 2009; Trickett, 2009). It is therefore not unreasonable to conclude that a comprehensive evaluation of a multi-level intervention should correspond to the various levels and activities of the intervention. Why then is evaluation and the collection of data at each of the levels intervened often neglected? The reason for this is unclear and it is not readily discussed in the published literature. One possible reason is that these multi-level interventions are still somewhat new, and evaluation often comes behind the development of new interventions. Other possibilities are that they are too complex (Mark, Henry, & Julnes, 2000; Nutbeam, 1998) or resource intensive to be included or done well, (Doll et al., 2003; Griffith, 2009; Mark et al., 2000; Spoth & Greenberg, 2005) or that evaluating such efforts have not been considered important. Such issues will be investigated during the course of this project.

A particular area of complexity for evaluating multi-level interventions is trying to untangle the impact an intervention has had on particular components rather than the larger system, all of which by design are interwoven. Multi-level interventions are often evaluated to assess intertwined elements of program effects through either: 1) statistical methods using

hierarchical, multi-level or similar modeling approaches or 2) design and methodology (Yang, Shen, Cao, & Warfield, 2004). The current study will not focus on statistical models beyond this brief reference, but will seek to better understand multi-level interventions through the use of evaluation design.

It should be noted that published literature provides several terms, including "multi-level," "multi-component," "multi-strategy" and "comprehensive" interventions, often representing the same concept. Nuance does exist, particularly for some investigators, but the field as a whole does not seem to stick to rigid distinctions in definition. The current study will use the term "multi-level" to represent interventions and evaluations implemented on two or more levels (individual, community, school, policy, etc.). The current study will explore comprehensive evaluations of multi-level interventions. Therefore, both the interventions and their evaluations will be multi-level (they will seek to intervene and assess impact across multiple levels).

Why Select Tobacco Control and Childhood Obesity?

Tobacco control and childhood obesity have been selected for study for several reasons. An overarching reason is that the fields have many similarities, more than many people might imagine, while also having some distinct differences.

Tobacco control and childhood obesity are two areas of serious concern currently within public health. According to the CDC, "Cigarette smoking remains the leading preventable cause of death in the United States, accounting for approximately 1 of every 5 deaths (438,000 people) each year" (Centers for Disease Control and Prevention, 2002, 2005; U.S. Department of Health and Human Services, 2004). Childhood obesity is described by the CDC as "a serious health concern" (2009). Rates of childhood obesity in the United States have nearly doubled since 1994 in children 2-5 years and 6-11 years of age, and have greatly increased, though not quite doubled, for the 12-19 year age group over the same period of time (Centers for Disease Control and Prevention. National Center for Health Statistics, 2006; Ogden, Carroll, & Flegal, 2008). Additionally, both tobacco use and childhood obesity are included for focus in the Healthy People 2010 (U.S. Department of Health and Human Services. Office of Disease Prevention and Health Promotion, 2005). In fact, there are two areas included in Healthy People 2010 that relate to childhood obesity, one focusing on nutrition and overweight and another on physical activity and fitness. It is not unreasonable to assume that both of these areas may be included in the Healthy People 2020 areas of focus, being determined in the next year.

Another reason for selected focus is that these are two fields in which multi-level interventions have been implemented and continue to gather momentum. Both of these fields specifically utilize socioecological and related models, describing multiple levels of influence and interaction among similar various factors. Intervention activities are quite different between tobacco control and childhood obesity, but strategies for implementation and evaluation may be quite similar when the interventions are multi-level.

Local, statewide and national childhood obesity programs and funders have noted the success of tobacco control work and indicated an interest in duplicating strategies, to the extent possible, in an effort to reduce childhood obesity (Beery et al., 2007). Mueller and

colleagues noted that lessons learned in tobacco control can be utilized to move other prevention areas forward, including specific reference to the field of childhood obesity (Mueller, Luke, Herbers, & Montgomery, 2006).

Multi-level Interventions within Tobacco Control and Childhood Obesity

Innovative ways to address tobacco control and childhood obesity include intervening on multiple levels concurrently, such as the individual, family, school, community, organizational and policy levels. The goal is often to use multiple avenues to create a change in social norms, expectations and environments around tobacco or nutrition and physical activity for children, their families and communities.

For tobacco control, strategies may include: reducing access of tobacco products by youth, reducing marketing of tobacco products to youth, changes in policies restricting where smoking is allowed, cessation services for individuals wanting to quit smoking, and media campaigns to promote messages of tobacco-free communities and residents. In an effort to reduce childhood obesity, strategies may include: increasing accessibility to affordable fresh and healthy foods, increasing availability and utilization of safe places for physical activity, and improving food and activity in school and childcare settings, among others.

Theoretical models describing factors for health improvement are often described in relation to multi-level interventions. Three such theoretical models include the ecological model, socioecological model, and ecological systems theory. These theories are similar in many ways, including a focus on multiple levels of influence and the importance of context. There is much overlap in how these terms are used and described, both in the literature and out in the field.

One example of an ecological model, made famous by Urie Bronfenbrenner, describes context through multiple levels, which can influence individual or group well-being (1979). This influence can take place through both the individual and its environment. Socioecological models have also been used to describe interactions that exist between a health behavior and multiple other components, including the individual and their physical and social environments (such as their home, community, neighborhood, etc.) (Green, Richard, & Potvin, 1996; Kothari et al., 2007). Ecological Systems Theory also considers context and how various components in a system interact and impact each other (Davison & Birch, 2001). Such models help capture the complexity of the interrelated relationships between individuals and their environment, and the impact this has on health decision-making (Green et al., 1996).

Credibility for utilizing an ecological model was enhanced by the publication of an Institute of Medicine report, summarized in the following statement by Smedley and Syme (2000):

> *"The committee examined a wide range of social and behavioral research that was intended to promote health and well-being of individuals, their families, and their communities, and found an emerging consensus that research and intervention efforts should be based on an ecological model"* (p. 2).

The Institute of Medicine model-of-choice was later revised to recommend the use of a social ecological approach (Trickett, 2009). These three models at times are used interchangeably, whether intentionally or not, in the literature and during real-world discussions of theoretical bases for health interventions. Based on the Institute of Medicine's recommendation, and common usage within the field, the term "socioecological" model and levels will be utilized for the purposes of the current discussion.

An example of a multi-level intervention utilizing a socioecological model can be seen in Figure 2 below. This project seeks to reduce childhood obesity by intervening on three levels concurrently (individual, home, school/afterschool), including several different activities within each level. The intervention seeks to change the system and environment within which health decisions are made. Similar interventions, utilizing a socioecological model and intervening on multiple levels concurrently, have been used for many years in the field of tobacco control (Centers for Disease Control and Prevention, 1999a). An example within tobacco control would be a project with intervention activities focused on: individual behaviors (cessation services, education programs), homes (smoke-free houses and cars), school programs (to delay onset of use), and promoting policy changes (such as smoke-free workplace laws) (U.S. Department of Health and Human Services, 2001).

Figure 2. Socioecological Model for Childhood Obesity Intervention
adapted from: (McLeroy, Bibeau, Steckler, & Glanz, 1988; Sallis, Bauman, & Pratt, 1998)
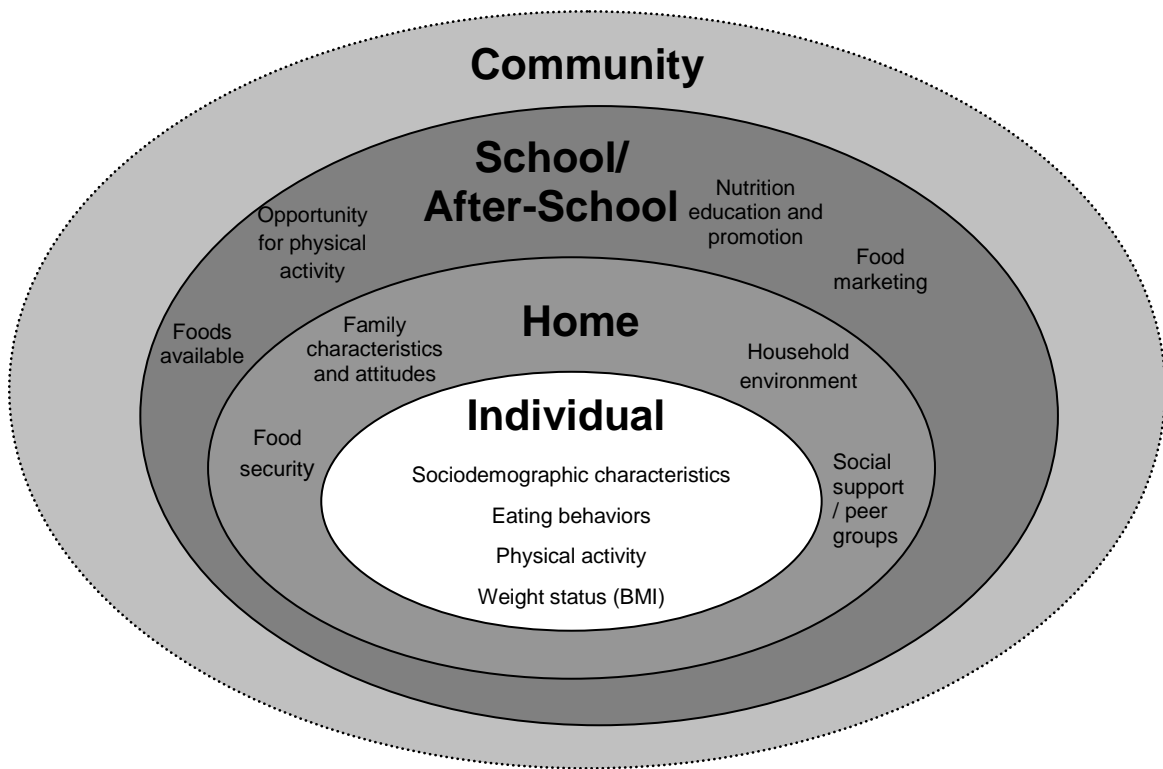
The CDC developed guidelines of "Best Practices for Comprehensive Tobacco Control Programs," which identify nine core components of a comprehensive tobacco control program (Centers for Disease Control and Prevention, 1999a). These nine components are: 1) cessation programs, 2) chronic disease programs, 3) community programs, 4) countermarketing, 5) enforcement, 6) school programs, 7) statewide programs, 8) surveillance/ evaluation, and 9) administration/management. These guidelines have been described by states using them as being a helpful planning guide to developing a comprehensive program, though further detail on implementation, activities and strategies were requested (Mueller et al., 2006). In a recent study, Mueller et al. (2006) found that many statewide tobacco control programs could not fund all "best practices" components, as laid out by the CDC, and had to choose which ones to fund. These statewide programs therefore prioritized tobacco control activities based on perceived potential impact, political acceptability, and using strategies that had been implemented previously (Mueller et al., 2006). A similar guideline for comprehensive childhood obesity programs do not seem to exist yet.

Kothari et al. (2007) describe current tobacco control projects as increasingly utilizing "multi-level and multi-strategy approaches" (p. 4). Childhood obesity projects seem further behind in implementing comprehensive interventions; there may be several reasons for this, including sources of funding and timing of attention to the issue. The field of tobacco control has moved forward greatly with research, interventions and evaluation being funded through the Master Settlement Agreement and tobacco tax allocations, at least in the United States (Mueller et al., 2006). Focus has been more prominently on tobacco control, second-hand smoke and related work for the last few decades. Concern around childhood obesity has increased dramatically within the last several years (Davison & Birch, 2001), but does not have as long a period of time in the spotlight, or a dedicated funding stream, as has been the case for tobacco control. Large amounts of funding have been prioritized recently, and research is being published, in an attempt to reduce childhood obesity, lead by organizations such as the Robert Wood Johnson Foundation (RWJF), National Institutes of Health (NIH), Institute of Medicine (IOM), and the CDC (Beery et al., 2007). The White House is getting involved as well, with its "Let's Move!" campaign highly publicized by the first lady, Michelle Obama. While great efforts are being made now related to childhood obesity, this field has not had the same amount of time of being a "hot topic" that tobacco control has had for the last few decades.

Significance

In-depth study of several projects evaluating multi-level interventions in tobacco control and childhood obesity were carried out through this dissertation, to fill a current gap in the fields of evaluation and public health. This study expands knowledge by assessing recent multi-level evaluations (of multi-level interventions) and contextual elements described by projects as influencing the comprehensiveness of their evaluations in tobacco control and childhood obesity. Lessons learned, promising practices, and recommendations from an assessment of cases studied are presented in this dissertation. The ultimate goal is to

provide the field of evaluation and public health with recommendations to increase comprehensiveness and quality of evaluations for multi-level interventions in the future.

Of most promise for advancing knowledge in this area are: 1) assessing key contextual elements, challenges encountered and promising practices for comprehensive evaluations of multi-level interventions, and 2) developing recommendations to help guide more comprehensive evaluation practices in the future.

CHAPTER 3: BUILDING A COMPREHENSIVE EVALUATION FOR MULTI-LEVEL INTERVENTIONS IN TOBACCO CONTROL AND CHILDHOOD OBESITY

Introduction

This chapter describes proposed elements for comprehensive evaluation of multi-level interventions in the fields of tobacco control and childhood obesity. Particular attention is paid to contextual elements which may influence the comprehensiveness of multi-level evaluations. Additionally, a working Framework for Comprehensive Evaluation of Multi-Level Interventions (Framework) has been developed, which is described in more detail throughout the chapter (and revisited in later chapters). A review of current and recent evaluations of multi-level interventions was conducted and findings from this review are presented in this chapter. Therefore, the main purposes of the current chapter are to: 1) identify proposed components of the working Framework and develop the Framework, and 2) identify current multi-level evaluation projects, which may be selected for further in-depth case study.

Published literature, and the current study, define a multi-level intervention as one that attempts to create change or impact at two or more levels (Nastasi & Hitchcock, 2009). In these types of interventions, activities often take place on multiple levels, such as the individual, family, school, community, organizational, or policy levels, among others. Implementing intervention activities on multiple levels concurrently are often intended to create a change in social norms, expectations and environments around a particular topic, with the goal of ultimately improving health or other outcomes. For the purposes of this study, comprehensiveness of multi-level evaluations are defined as: interventions and evaluations being multi-level, activities taking place concurrently, and evaluation activities happening, for the most part, on levels being intervened. Levels evaluated do not always correspond one-to-one with levels for desired outcome change, but there is usually great overlap, at least in evaluations being defined in this study as comprehensive. In comprehensive evaluations, process evaluation activities often take place on levels, even when outcome evaluation does not take place, or it is too soon to expect any outcome changes to be seen. For example, process measures of the number of cessation classes held, feedback from participants and media run may be used for a while before short-term or longer-term outcomes can be assessed, such as smoking quit rates. It is not necessarily realistic to expect that an evaluation needs to take place for all activities, and at all levels carried out by the project, for the evaluation to be considered comprehensive. The current study will explore comprehensive evaluations of multi-level interventions. Therefore, both the interventions and their evaluations will be multi-level (they will seek to intervene and assess impact across multiple levels), these activities are generally taking place concurrently, and evaluation is taking place on most levels being intervened.

Multi-level interventions are being developed, and more frequently implemented in several fields, but often do not include comprehensive evaluations. If evaluation is included, it is often piece-meal and important elements, such as, data collection on all levels intervened, are often lacking. The reasons behind this disconnect between the

comprehensiveness of the intervention and evaluation and potential resulting consequences are not clear. Additionally, it is unknown which levels are being addressed by multi-level projects, both through intervention and evaluation, and whether specific elements and contextual factors are linked to developing comprehensive evaluation practices. There has also been a call to the field to develop a framework for comprehensive evaluation of complex interventions (Craig et al., 2008; Warner, 2009). A consortium specifically addressing challenges to evaluating comprehensive tobacco control interventions concluded that there was a lack of methods and measurements available to adequately and accurately assess impact around such interventions (Conference Report from Institute for Global Tobacco Control at Johns Hopkins Bloomberg School of Public Health, 2002).

The current chapter begins to address these gaps by assessing the fields of tobacco control and childhood obesity for examples of comprehensive evaluations of multi-level interventions.

The following research question is the focus of this chapter: *What are elements of a comprehensive evaluation for multi-level interventions in tobacco control and childhood obesity?*

Data were collected for this chapter through two literature reviews, which are described in further detail in the remaining sections of this chapter. The purpose of the first literature review was to provide background information and to assist in developing the working Framework for Comprehensive Evaluation of Multi-Level Interventions, which is described in more detail throughout the chapter. The purpose of the second literature review was to determine examples of articles, documents and programs that meet the criteria for multi-level evaluations within the fields of tobacco control and childhood obesity. These documents were reviewed to better understand the levels being intervened and evaluated, types of evaluations being conducted and contextual elements influencing these multi-level evaluations. This second review also provided further insight and background information on projects that could be selected as cases for further study, which was done as a later part of this dissertation and presented in the next chapter.

Elements of a Comprehensive Evaluation of Multi-Level Interventions (Literature Review 1)

What is meant by, and influential in, developing a "comprehensive evaluation of multi-level interventions" is explored in detail in the following section. Key elements proposed as necessary for comprehensive evaluation of multi-level interventions are described, as well as contextual elements that will be explored for influence on the comprehensiveness of evaluations.

A review was conducted of literature related to evaluating multi-level interventions within and beyond tobacco control and childhood obesity, elements of comprehensive evaluations, and contextual factors, among other relevant topics. Reviewing literature in these areas was used to develop a working framework for conceptualizing elements of a comprehensive multi-level evaluation and possible contextual elements influencing the comprehensiveness of these evaluations. For this literature review, searches were conducted of published literature using keywords through search engines including PubMed, Google Scholar, and Web of Science. Keywords included combinations of: "comprehensive

evaluation," "multi-level interventions," "context," "comprehensive intervention," "evaluation of," "complex," "tobacco control," "childhood obesity," "multi-component," "ecological," and "socioecological." Articles were reviewed for applicability and further refined by health field as necessary. Specific additional articles and authors cited in relevant articles were also reviewed.

Information gathered through this review of the literature is presented below in three parts. First, is a description of elements for inclusion for a comprehensive evaluation of a simple or single-level intervention, as developed by the Centers for Disease Control and Prevention (CDC). The second section speaks more directly to the added complexity of evaluating multi-level interventions. To more appropriately capture multiple streams of intervention activities happening around the same topic, measures of intervention implementation and outcomes should be captured at each level addressed by the intervention. Additionally, measures of changing dynamics within the system or measuring changes to social norms can help to tell a more complete story of the impact a multi-level project is actually having (or not). Comprehensive multi-level evaluations seem to grapple with not only a dynamic and emergent system, one that is changing over time and involves moving parts, but also trying to measure what elements impacted which changes. Social norms can be used as an element being tracked in a changing system. Therefore, it can be considered a more specific element than measuring systems change. Capturing changes in dynamics within the system and social norms are considered important and described as part of this Framework and a comprehensive multi-level evaluation, but they are not explored in great detail in the current dissertation. It will be left to another study to give this topic the level of attention it deserves and understand this complex issue better. The focus for the rest of the dissertation is primarily on the contextual elements from the Framework. In the last section, beyond criteria to include for building a comprehensive evaluation, as previously described, are contextual elements that may influence the comprehensiveness of multi-level evaluations. These three sections build upon one another to describe comprehensively evaluating the most simple, to more complex, multi-level interventions. To comprehensively evaluate a multi-level intervention, the elements identified would be included and contextual factors taken into consideration.

These three sections describe in further detail the elements presented in the working Framework for Comprehensive Evaluation of Multi-Level Interventions included as Figure 1. As seen in this Figure, the Framework includes: elements for standard evaluation (shown at top of white box), elements to be given additional focus for multi-level interventions (shown at bottom of white box), and contextual elements influencing comprehensiveness of evaluation (shown in blue shaded boxes). Further information on each of these sections are explored in further detail below the Figure.

Figure 1: Framework for Comprehensive Evaluation of Multi-Level Interventions

Contextual Elements Influencing Comprehensiveness of Evaluation

Resources

Evaluation Requirement by Funder

Training/ Experience of Evaluator

Length of Time Carrying Out Type of Interv/ Eval

Elements for Standard Evaluation

Six "steps" for comprehensiveness from CDC Framework:
- o Engage stakeholders
- o Describe the program
- o Focus the evaluation design
- o Gather credible evidence
- o Justify conclusions
- o Ensure use and share lessons learned

Four "standards" for quality from CDC Framework:
- o Utility
- o Feasibility
- o Propriety
- o Accuracy

Elements to be given Additional Focus for Multi-level Interventions
- ➢ Evaluation / data collection included for ALL levels where intervention activities are taking place
- ➢ Special attention to specific components (again, at all levels)
  - • Measure of Intervention Implementation
    - o Particularly dose, fidelity, quality of implementation
  - • Links between project activities and short, intermediate and long-term outcomes
  - • Clearly defined and measureable outcomes
  - • Outcomes and activities are measured
- ➢ Measures of dynamic system/ changes in social norms

Intended Use of Evaluation

Access to Information to Guide Evaluation

Evaluator Access

Evaluation Capacity of Project/ Org

Agency Support

Social/ Environ. Conditions

Politics

Competing Orgs/ Interests

<u>Elements of a Standard Evaluation</u>

In 1999, the CDC developed a "Framework for Program Evaluation in Public Health" (CDC Framework), laying out "essential" elements for conducting evaluation within Public Health (Centers for Disease Control and Prevention, 1999b). The CDC Framework included six steps for effective evaluations, as well as four standards to ensure quality. As can be seen in Figure 2 below, the six steps are: 1) engage stakeholders, 2) describe the program, 3) focus the evaluation design, 4) gather credible evidence, 5) justify conclusions, and 6) ensure use and share lessons learned; and the four standards are: 1) utility, 2) feasibility, 3) propriety, and 4) accuracy. Both the steps and standards are meant to provide elements for consideration and inclusion, yet are not too detailed or prescriptive to ensure applicability and utilization in various contexts. These steps and standards are included in the working Framework for Comprehensive Evaluation of Multi-Level Interventions (Figure 1), to represent elements for a standard evaluation.

Figure 2. CDC Framework for Evaluation in Public Health
(Centers for Disease Control and Prevention, 1999b)
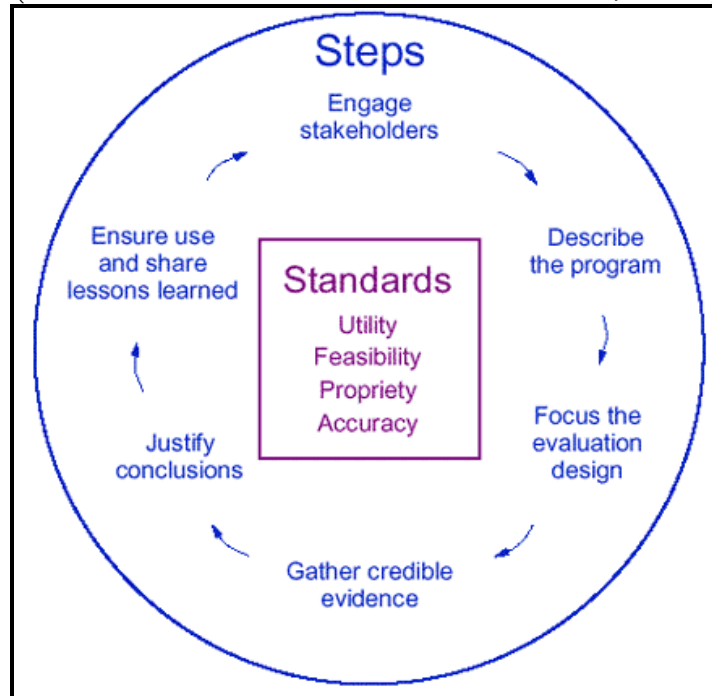
The lack of rigid rules to follow is echoed by the following description put forth by the U.S. Department of Health and Human Services (2005): "Thus, the CDC Framework approach supports the fundamental insight that there is no such thing as *the* right program evaluation. Rather, over the life of a program, any number of evaluations may be appropriate, depending on the situation" (p. 7). Differing purposes and definitions of evaluation can influence several things including methodology, conclusions drawn and the scope of evaluations conducted. The CDC Framework for evaluation describes elements considered critical to developing, implementing and disseminating effective evaluations, both

broadly as well as within specific contexts. Promotion of making the best choices possible, within acknowledged constraints of reality, resources and context, makes the CDC Framework practical and feasible. Unclear is whether these guidelines are widely used, known, and to what extent following these steps actually ensures an "effective" evaluation.

Further detail about how to apply this Framework to actual programs was developed several years later by the CDC (U.S. Department of Health and Human Services, 2005). This document primarily uses examples from programs on childhood lead poisoning prevention, provider education for immunization, and affordable home ownership to walk through the six steps of the CDC Framework. A separate document, also developed by the CDC, details how to evaluate comprehensive tobacco control programs using the CDC Framework (U.S. Department of Health and Human Services, 2001). While the focus of this document is on interventions of a comprehensive nature within tobacco control, special attention is not particularly paid to the comprehensiveness of the evaluation itself or the impact this dynamic multi-level intervention might have on evaluation planning and implementation. A parallel document, specific to evaluating comprehensive childhood obesity programs, does not currently seem to exist, though a document focusing on physical activity has been developed. Again, the focus is on more straight-forward interventions and does not focus on childhood obesity specifically or on the combination of interventions addressing both nutrition and activity together (U.S. Department of Health and Human Services, 2002).

In 2005, the CDC produced another document expanding the description of evaluations and outcomes for comprehensive tobacco control programs. Figure 3 provides seven steps set forth to plan and evaluate a comprehensive statewide tobacco control intervention. These steps describe selecting program outcomes and linking them to data collection and intervention activities. While this document is quite valuable for a variety of audiences, as was similar to previous documents, the focus was on the comprehensiveness of the intervention and program overall, more than the comprehensiveness of the evaluation. Missing in this document is a description of how pieces of this comprehensive intervention would be linked together or accounted for systematically within the evaluation. It is unclear how these various elements work together towards common or separate goals, as is usually the intent within a comprehensive or multi-level intervention. Also absent is a description of process evaluation and how systemic or social norm change would be measured or assessed in this type of comprehensive program (California Department of Public Health, 2009).

A separate document was produced in 2008 to describe the potential contribution of process evaluation for comprehensive tobacco control intervention programs. Again, the comprehensiveness of evaluating a multi-level program and how or why pieces should be linked together was not explored. In fact, a strong argument for comprehensive evaluation, particularly with process evaluation, was not particularly encouraged, as exemplified by the following statement by the CDC (U.S. Department of Health and Human Services, 2008):

> *"Also, be mindful that process evaluation of an overall comprehensive state program such as tobacco control is not typically conducted because the scope of such an evaluation would not be practical and would be extremely costly. Rather, process evaluation is typically used to evaluate a given component or activity within an overall comprehensive program"* (p .1).

Figure 3. Planning and Evaluating a State Tobacco Control Program (Starr et al., 2005)

1. Select the NTCP goal area that suits your program best.
2. Select long-term outcomes for your program.
3. Select short-term and intermediate outcomes for your program.
4. Select indicators of progress toward your selected short-term, intermediate, and long-term outcomes.
5. Select or design activities to achieve your selected outcomes.
6. Implement your selected intervention activities.
7. Evaluate your progress toward achieving your selected outcomes.

Elements to be Given Additional Consideration for a Multi-level Evaluation

The Framework for Comprehensive Evaluation of Multi-Level Interventions (Framework) is developed during this chapter (Figure 1). Beyond elements to include for a standard evaluation are additional considerations specific to evaluating a multi-level program. These elements described in the working Framework include:

➤ Evaluation / data collection included for ALL levels where intervention activities are taking place
➤ Special attention to specific components (again, at all levels)
  • Measure of Intervention Implementation—particularly dose, fidelity, and quality of implementation
  • Links between project activities and short, intermediate and long-term outcomes
  • Clearly defined and measureable outcomes
  • Outcomes and activities are measured
➤ Measures of dynamic system/ changes in social norms

Developing an evaluation for a multi-level intervention is complex for many reasons. Multi-level interventions themselves are complex and expensive; adding an evaluation to document the program's activities, implementation and outcomes adds additional complexity and expense. It may be particularly difficult when the end result will not likely be seen for many years, such as projects seeking to reduce lung cancer or childhood obesity. With this in mind, it is crucial to develop an evaluation plan that can track data on both short and long-term outcomes and provide needed information for long-term sustainability. Developing realistic expectations in funders for the pace at which changes such as these occur is also important. In the early stages of implementation, it is important for the evaluation to develop a framework for data collection, linking activities at multiple levels to data for short and long-term outcomes.

Of particular value at an early stage is documenting the process of implementing the project, including ensuring that dose and fidelity are captured (Saksvik, Nytro, Dahl-Jorgensen, & Mikkelsen, 2002; Spoth & Greenberg, 2005) and determining most and least

valuable components to guide future action. An intervention may be quite effective for participants completing the program, but if few people attend all sessions or data are not collected to measure this, it is unknown whether the true impact of the program will be captured (U.S. Department of Health and Human Services, 2001). Another scenario is that a program may show great success or failure, though if measurement of fidelity to the intended intervention is not included in the evaluation, an alternate explanation, beyond the intervention working or not, may be missed. Consistency of data collection and implementation of the intervention across various levels and sites have also been identified as important, but often challenging, components of evaluating multi-level interventions (Benn et al., 2009; Jurg, Kremers, Candel, Van der Wal, & Meij, 2006; Yang et al., 2004). Rychetnik et al. (2002) support that "evidence of adequate implementation, and other measures to monitor the evaluation process, are important regardless of whether the findings are negative or positive" (p. 122).

Evaluation can be crucial to identify any impact a project may have, as well as sustaining the project with links to shorter-term outcomes until it is realistic to look at longer-term outcomes. Nutbeam (1998) indicates that it may be "inappropriate and unrealistic" for most programs to assess long-term health and social impact, but that evaluating program implementation and short-term outcomes are necessary to start building evidence and determining potential conditions for success (p.40).

Short-term, narrowly-defined and closely controlled interventions are easier to evaluate, but likely do not accurately capture the complexity of real-world health problems (Nutbeam, 1998). As noted by Mark, Henry and Julnes (2000), "Although such multi-level assessments may increase the difficulty of making overall judgments about a program, greater problems will often result from assessments that produce overly simplified conclusions based on only one level of the stratified reality"(p. 155).

Shiell, Hawe and Gold provide a valuable discussion around the complexity of evaluating dynamic systems. They describe that evaluations often include measures of individual health outcome changes added up, but that this is not necessarily appropriate within a complex system because the emergent properties are not captured (Shiell, Hawe, & Gold, 2008). They go on to discuss how emergent properties cannot be captured as a discrete component of the system or by adding up other elements, because it is part of the system as a whole. Shiell, Hawe and Gold (2008) expand this thinking further to describe that:

> *"…evaluation of interventions aimed at changing systems requires new ways of thinking: one sensitive to ecological theory, interactions between microlevel and macrolevel variables, non-linearities, multiplier effects and the fact that individual values are shaped by the interventions we seek to evaluate and the contexts we seek to change"* (p. 1283).

This provides quite a challenge to evaluators. Solutions put forth to address such challenges include developing new tools to meet this task and measuring within short intervals to detect change as it occurs, though evaluation such as this is acknowledged to likely be expensive (Shiell et al., 2008). Benn et al. and others have suggested a need for new

designs that are sensitive to complexity in system change and other similar issues facing evaluations of multi-level interventions (Beery et al., 2007; Benn et al., 2009; Warner, 2009).

This discussion may be best summed up by these words from Rychetnik et al. *(2002):*

> *"Public health interventions tend to be complex, programmatic, and context dependent. The evidence for their effectiveness must be sufficiently comprehensive to encompass that complexity. The evaluation of evidence must distinguish between the fidelity of the evaluation process in detecting the success or failure of an intervention, and the relative success or failure of the intervention itself. Moreover, if an intervention is unsuccessful, the evidence should help to determine whether the intervention was inherently faulty (that is, failure of intervention concept or theory), or badly delivered (failure of implementation). Furthermore, proper interpretation of the evidence depends upon the availability of adequate descriptive information on the intervention and its context, so that the transferability of the evidence can be determined"* (p. 119).

While issues around measuring synergy, emergent and dynamic systems and systems change are quite interesting, they will not be explored in great detail in the current study, beyond describing that this is an area for further exploration and research. These are areas that projects evaluating comprehensive programs may struggle with, and will be identified as such if found, but not explored for further understanding of how to address this issue in the current dissertation.

Contextual Elements

Part of the working Framework developed in this chapter (Figure 1) is the inclusion of contextual elements proposed to influence the comprehensiveness of such evaluations. Contextual elements described in the working Framework include the following:

- Resources
- Evaluation Requirement by Funder
- Training/ Experience of Evaluator
- Length of Time Carrying Out Type of Intervention/ Evaluation
- Intended Use of Evaluation
- Access to Information to Guide the Evaluation
- Evaluator Access
- Evaluation Capacity of Project/Organization
- Agency Support

Broader contextual elements which may influence comprehensiveness include:

- Social/ Environmental Conditions
- Politics
- Competing Organizations/ Interests

Contextual elements are thought to influence the development and implementation of an evaluation, including its comprehensiveness. Though these factors may not be elements required for an evaluation to be comprehensive *per se*, they do likely have tremendous influence on carrying out a comprehensive evaluation. Contextual elements can influence a variety of areas, from the design selected, implementation of the intervention, which can affect the evaluation, to the collection of data, analysis conducted, and interpretations made. This is particularly the case for multi-level interventions. Such interventions are complex and dependent upon their context, therefore the evaluations should be too (Rychetnik et al., 2002). Of additional complexity is the dynamic nature of the system and how context interacts with these factors for multi-level projects. Further adding to the complexity is the idea that elements of the intervention and the context within which it is set cannot easily be separated out (Hawe, Shiell, & Riley, 2009). Though it is beyond the scope of the current study, contextual variables can be described and assessed within each level of a project's intervention and evaluation. See work by Brownson, Fielding and Maylahn for further discussion of this area (Brownson, Fielding, & Maylahn, 2009).

The remainder of this section describes further these contextual elements and how they may influence the comprehensiveness of evaluating multi-level interventions within tobacco control and childhood obesity.

Resources

Evaluation of multi-level interventions have been described as too resource intensive, as a rationale for why they are not done or not done well (Doll et al., 2003; Griffith, 2009; Mark et al., 2000; Spoth & Greenberg, 2005; U.S. Department of Health and Human Services, 2008). It is true that a comprehensive evaluation of a multi-level intervention requires a great amount of resources. Like the intervention itself, the evaluation would require resources through funding, personnel, time to carry out, and likely developing and maintaining partnerships and coordinating efforts among various parties involved. As stated by Chappell and colleagues (2006) "…comprehensive interventions that begin by addressing multiple levels require simultaneous commitment from many organizations at the outset, extensive planning and management, and substantial energy and resources…" (pp. 353-354).

Resources described by Richard, et al. (2004) as facilitating implementation of an ecological model included having funding based on the budget being prioritized over other topic areas, staff available to work on the project, especially those with experience and expertise in the area, and creating and utilizing partnerships.

It would be informative to better understand how resources influence the comprehensiveness of evaluating multi-level interventions. While resources are often a challenge, some projects are able to utilize strategies to make the most of resources available or find creative solutions to overcoming this barrier. It is unclear how much is often budgeted or how resources are allocated, for evaluating such interventions.

Evaluation Requirement by Funder

Many, but not all, funders require evaluations for projects and interventions they fund. There are many multi-level interventions being funded currently, both in tobacco

control and childhood obesity, but it is unclear how many require evaluation, what the evaluation entails, whether this is up to the funder or funded organization to determine, and whether the project is held accountable for meeting any evaluation requirements in place. Additionally, it is unknown whether projects add to any evaluation requirement set forth by their funder and under what circumstances this takes place. Some funders, such as the California Tobacco Control Program (CTCP) require at least ten percent of a project's budget be used for evaluation of their activities and outcomes. This is the same threshold set forth by the CDC for implementing comprehensive tobacco control programs. The CTCP provides guidance and technical assistance to funded agencies to develop an evaluation plan that meets their evaluation requirements (Albuquerque, Starr, Schooley, Pechacek, & Henson, 2003; Centers for Disease Control and Prevention, 1999a; Modayil, Cowling, Tang, & Roeseler).

Training/ Experience of Evaluator

 Richard, et al. (2004) described previous experience, training, and education as facilitating implementation of an ecological model in tobacco control. It would make sense that having an evaluator with training and experience would help foster the development and completion of a more comprehensive evaluation. But, it may be that other factors, such as resources, accessing helpful sources of information or the guidance of decision-makers may be as important to the ultimate comprehensiveness of an evaluation. It would be informative to better understand the influence of an evaluator's training and experience on the comprehensiveness of a multi-level evaluation.

Length of Time Carrying Out Type of Intervention/ Evaluation

 Published literature indicates that multi-level interventions and corresponding evaluations have been taking place within the field of tobacco control for a longer period than within childhood obesity. Additionally, there has been a more sustained source of funding for tobacco control work, at least in some locations, compared to childhood obesity. Both of these elements could influence the comprehensiveness of evaluations and a difference may be seen between these two fields for these reasons.

 The length of time carrying out multi-level interventions and related evaluations can influence both the individual program carrying out their work and the larger field based on knowledge made available to others.

Intended Use of Evaluation

 The purpose and intention of the evaluation can greatly influence the type of evaluation conducted, rigor and complexity of design and data collection. One size does not fit all. Patton has stated that methodological appropriateness is the "gold standard" rather than always following one specific technique, whether that is in the design or measurement tools selected (1997). While the specific methodology and purpose may vary, evaluation of multi-level projects can provide a much-needed evidence base to better understand which intervention components are particularly effective or ineffective.

The intended use of the evaluation can influence which elements are included in the evaluation and the comprehensiveness of the evaluation. It is less clear who determines evaluation use and how this influences comprehensiveness and the process through which these decisions are made.

Access to Information to Guide the Evaluation

This contextual element relates to the role that sources of information play, in the comprehensiveness of these evaluations. Projects likely have access to information to guide the development of a comprehensive intervention, such as those put forth by the CDC (1999a), but fewer resources seem to exist to guide the development of a comprehensive evaluation of such projects. Additionally, it is not clear how well known or readily available such information is to projects seeking to develop a comprehensive multi-level evaluation.

It would be valuable to learn about sources of information used by projects in developing their evaluations. This could include documents, data, people, or other sources of information that informed the evaluation for the project. Work from other fields or departments within their organization may be another source of valuable information used for their own evaluation. Such information could inform decisions around the evaluation scope and comprehensiveness, design, data collection methods, instruments selected, analysis and reporting techniques and plans for dissemination of findings.

Evaluator Access

Access to readily available and qualified evaluators could also influence the comprehensiveness of multi-level evaluations. Evaluators internal or external to the project or organization could fill this role, but having access to available evaluators is a first necessary step. Not having access to an evaluator can also impact other areas, such as the timing of when evaluation is incorporated into the overall project design and decisions and access to evaluation-related sources of information, among others. Understanding project access to evaluators, especially when needed, is important to understand for the above mentioned reasons.

Evaluation Capacity of Project/Organization

A project or organization's capacity for evaluation can also play a role in the comprehensiveness of the evaluation developed and implemented. Skills and tools necessary to carry out a complex evaluation may be required beyond that of a specific evaluator, whether internal or external. Project staff may play a critical role in organizing elements of the evaluation, creating buy-in among other team members, partners or community members, and collecting data. Limited capacity for involvement in the evaluation may influence the scope of the evaluation that can be developed and successfully completed.

Agency Support

Several studies have found various elements around agency support as linked to successful implementation of multi-level projects. These include staff acceptance, support by supervisors, and norms of the organizational environment (Benn et al., 2009; Richard et al.,

2004). These elements of agency support could influence the comprehensiveness of the evaluation on their own, or as related to with other factors such as evaluation capacity, resources and intended use of the evaluation.

Broader Contextual Elements

Broader contextual elements which may also influence the comprehensiveness of multi-level evaluations include: social and environmental conditions, politics, and competing organizations and interests.

Richard, et al. (2004) describes contextual factors facilitating implementation of an ecological model- based intervention. Primary facilitating factors included resources, agency and supervisor support, previous experience, training and education specifically around ecological approaches, and the political environment. Other contextual elements including evaluation requirement by funder, intended use of the evaluation, access to evaluators and sources of information, and evaluation capacity of the project likely influence the comprehensiveness of evaluations. The extent to which these contextual factors influence the comprehensiveness, however, is not well known. Additional exploration of this topic can provide valuable information to the field.

Hypotheses Linked to Contextual Elements in the Framework

Hypotheses were developed by the author about how certain contextual elements might influence the comprehensiveness of evaluations of multi-level interventions. A combination of early literature reviews and these hypotheses aided in the development of the Framework presented as Figure 1.

Hypothesis: All contextual elements described in the Framework influence the comprehensiveness of evaluations for multi-level interventions. Through a review of the literature, it was proposed that all contextual elements in the Framework would influence the comprehensiveness of evaluations. Contextual elements would not have been included in the Framework if it was not expected that it would influence comprehensiveness. This hypothesis was tested through data collected from four cases and comparison made with contextual elements from the Framework. This comparison is presented in Chapter 5.

Hypothesis: Requirement of evaluation by project funder influences comprehensiveness of the evaluation, regardless of field. Beyond looking at the influence on comprehensiveness for all contextual elements, as done through the first hypothesis, one contextual element was proposed to have specific influence on evaluation comprehensiveness. It was thought that of all the contextual elements, this one might particularly influence evaluation comprehensiveness, if for no other reason than it was required of projects. It was unknown what types of requirements funders put in place for evaluations they fund, whether funded projects adhere to these components, how these requirements influence comprehensiveness, whether projects add to these evaluation requirements, and under what circumstances this takes place. This hypothesis was tested through data collected from four cases and comparison made with contextual elements from the Framework. This comparison is presented in Chapter 5.

The proceeding section described three sets of elements for inclusion by a comprehensive multi-level evaluation. These three sets of elements were: elements for standard evaluation, elements to be given additional focus for multi-level interventions and contextual elements proposed to influence the comprehensiveness of these evaluations. This information was presented in graphic form as the Framework for Comprehensive Evaluation of Multi-Level Interventions (Figure 1).

In addition to the first literature review conducted to develop the Framework, a second literature review of recent multi-level evaluations was conducted. Elements from the Framework were used to assess these evaluations with the intent of finding examples of comprehensive evaluations within these fields.

Review of Multi-Level Evaluations in Tobacco Control and Childhood Obesity (Literature Review 2)

<u>Purpose</u>

The purpose of this literature review was to gain a better understanding of the current status of multi-level evaluations within the fields of tobacco control and childhood obesity.

This literature review was also intended to provide further insight and background information on projects that could be selected as cases for further study.

<u>Methods</u>

Selection Criteria

For this review, a search was conducted from March to May 2010 of published literature and internet sources using keywords through search engines including PubMed, Google Scholar, Web of Science and Google. Search terms included combinations of: "comprehensive evaluation," "multi-level interventions," "multi-level evaluation," "multi-," "context," "comprehensive intervention," "evaluation of," "complex," "tobacco control," "childhood obesity," and "multi-component." Articles were further refined by health field, and since 1995, as necessary, to further refine the search. This search yielded 547 articles and documents. Select other articles and authors cited in relevant articles were also reviewed for inclusion. Abstracts and links to documents were followed for more detailed review.

Eligibility or inclusion criteria used for this review included: intervention of tobacco control or childhood obesity project; intervention implemented on two or more levels; evaluation conducted on two or more levels; English-language document; publicly available articles, reports or documents.

The majority of articles found in the initial search were not included in the final review because while found through search terms, they did not meet inclusion criteria. The most common reason that articles were excluded was being outside the scope of the fields of tobacco control or childhood obesity (when finding articles about multi-level evaluations and interventions). Many articles were found related to multi-level interventions or evaluations in other fields (drug or alcohol studies, physical rehabilitation, nutrition, gambling, HIV/AIDS, diabetes, etc.). Also found frequently were projects within tobacco control or childhood

obesity, which focused on a single level for the intervention, evaluation, or both. The preliminary search and application of inclusion criteria were done by one person, the sole researcher on this project.

Twenty-one articles and reports that met the inclusion criteria were included in the review of recent evaluations of multi-level interventions. They are presented individually in Table 1.

Data Collection

Data were collected, when available, from articles and reports in the following areas:
- field (tobacco control or childhood obesity only);
- project scope/setting (statewide, local community, school);
- levels attempted/levels actually intervened (individual, school, community, etc.);
- levels in which data were collected/evaluated;
- methods used in evaluation;
- sampling for evaluation;
- measurement of dose/fidelity of implementation;
- evaluation conclusions drawn;
- mention of evaluation as required by funder, and whether additional evaluation added;
- resources available and potential influence (funding, evaluation capacity, timeline, etc.).

In total, 21 articles and reports were assessed as part of this review, including 11 from tobacco control and ten from childhood obesity. A spreadsheet was developed to track projects reviewed and project information on each of the elements of interest, as previously described.

Analysis

This review utilized both quantitative and qualitative analysis techniques. Simple descriptive analysis provides information on the number of projects reviewed that fell into various categories (field: tobacco control/childhood obesity; levels evaluated; scope/location of projects). Qualitative analysis was utilized for reviewing articles and documents. This was done to assess relevant descriptions, levels evaluated, and rationale for potential selection of cases studied in further detail. This qualitative component is included to better understand the projects being reviewed, their context, how the components fit together and other elements that may be important in determining projects for potential selection through in-depth case study.

Results

In total 21 articles and reports were reviewed. Eleven of these were from tobacco control and ten from childhood obesity. These articles represent 14 unique programs within

these two fields (seven from tobacco control and seven from childhood obesity) that are carrying out multi-level evaluations of multi-level interventions.

More similarity was found in the review of these articles than expected, such that equal numbers of tobacco control and childhood obesity programs were found to have meet the review criteria. The tobacco control articles were often older than those for childhood obesity. As an example, four tobacco control reports were from before the year 2004, but there was only one childhood obesity program, which was their baseline study. The scope of the programs, ranging from school and local projects to national programs was comparable between these fields.

Several elements intended for review were difficult to obtain through these published and publically available sources. These included consistent and detailed information about: dose, fidelity, contextual elements (such as funder requirement for the evaluation and resources available), levels intervened and levels intended versus actually intervened or evaluated.

Six socioecological levels were <u>evaluated</u> through the articles and programs in this review (individual, community, school, organizational, policy and media). The most frequent combination was evaluation at the individual, community and policy levels concurrently.

Multi-level evaluations reviewed covered these six levels as follows:
- Individual (18 total in all articles reviewed; 11 total by programs reviewed)
- Community (14 total in all articles reviewed; 10 total by programs reviewed)
- School (11 total in all articles reviewed; 8 total by programs reviewed)
- Organizational (2 total in all articles reviewed; 1 total by programs reviewed)
- Media (2 total in all articles reviewed; 2 total by programs reviewed)
- Policy (16 total in all articles reviewed; 10 total by programs reviewed)

Articles and programs reviewed were from the United States and beyond. Of the 14 programs included, nine were based in the United States and five were not. Programs outside the United States were from: Spain, Scotland, Canada, Netherlands and Australia.

The following tables detail articles, reports and other documents reviewed, as well as, the overarching project to which they are tied.

Table 1 describes the review of individual articles and reports within both tobacco control and childhood obesity. This table identifies the field of study (tobacco control/ childhood obesity), name of article or report reviewed, project being evaluated, levels evaluated, and a brief summary of the evaluation.

Several projects came up repeatedly through various searches during the literature review. It became apparent that the *project* would be the most appropriate level at which to determine cases for potential selection for future in-depth study.

Table 2 describes projects, which are potential cases considered for in-depth study in the fields of tobacco control. This table includes information on: project name, levels

evaluated, project overview, comments for case selection, and scope and location of the project.

The "comments for case selection" column notes reasons why a project might be selected or not for detailed study. It should be noted that beyond the contribution of the individual project, will be the combination of projects being considered for the multiple case study. Therefore, a project on its own may be a good fit for the intention of the study, but if a better grouping of projects link more closely with the goals of the multiple case study, that combination of projects may be selected instead.

Table 3 describes projects, which are potential cases, being considered for future in-depth study in the field of childhood obesity. This table also includes information on: project name, levels evaluated, project overview, comments for case selection, and scope and location of project. The difference between Table 2 and Table 3 is the field of focus.

Table 1. Summary of Articles and Reports Reviewed

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Tobacco Control	Independent Evaluation of the CA TCP: Relationships between Program Exposure and Outcomes 1996-1998. (Rohrbach et al., 2002)	California Tobacco Control Program (CTCP)	*Individual* (cessation, s-f homes); *Policy and Community* -Exposure to SHS/ support for policies, *(*SHS- bars/ restaurants, advertising at retail outlets, youth access to tobacco products);	-Eval of statewide CTCP and intermediate/ ultimate outcomes. -Several reports came out of this longitudinal evaluation. -Telephone survey with adults (randomly selected from 18 Counties statewide) and written surveys with youth (randomly selected from 84 high schools). -Conclusions: Adults- more exposure to CTCP work linked to lower smoking prevalence, increased no-smoking policies at home and fewer violations of workplace polices. Youth- no effect found for outcomes. -CTCP may have reduced adult smoking and exposure to SHS. -External contractor for CTCP.
Tobacco Control	Independent Evaluation of the California Tobacco Control Prevention and Education Program: Waves 1, 2 & 3 (1996-2000). (Independent Evaluation Consortium, 2002)	California Tobacco Control Program (CTCP)	*Individual* (cessation, s-f homes); *Policy and Community* *(*SHS- bars/ restaurants, advertising/marketing at retail outlets, youth access to tobacco products); *Media Campaign;* *School programs;* *Data also collected on marketing by tobacco industry.*	-Final Evaluation Report for longitudinal evaluation. -Accountability for Prop 99 (tobacco tax) money. -Used Logic Model. -Assessed intermediate/ ultimate outcomes on multiple levels based on intervention at multiple levels -3 waves of data collection with 18 months between each. -Surveys, interviews, doc reviews. -18 of 58 Counties selected for inclusion. -Conclusions: Californian's exposed through at least two different levels to tobacco control messages. Higher exposure to CTCP work linked to: lower perceived access to tobacco by youth, reduced adult smoking prevalence, and increase in bans for smoking in homes. -External contractor for CTCP.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Tobacco Control	California Tobacco Control Update 2009: 20 Years of Tobacco Control in California. (California Department of Public Health, 2009)	California Tobacco Control Program (CTCP)	*Individual* (cessation, homes/ cars); *Policy and Community* (SHS- bars/ restaurants/ workplaces, outdoor areas; advertising/ marketing at retail outlets, youth access to tobacco products; *Media Campaign;*	-Good background and overview of CTCP as a whole, intervention and evaluation work. -CTCP funds ~100 local projects throughout California, plus: statewide media campaign, tobacco cessation helpline, and statewide technical support svcs. -Uses logic model to guide work -Seeking change in social norms through 4 priority areas: 1) countering pro-tobacco influences in the community (advertising/ marketing of tobacco products, event sponsorship); 2) reduced exposure to SHS (policy/ advocacy work); 3) reduced availability of tobacco (enforcing youth sales laws, eliminate sampling/ tobacco in pharmacies); 4) cessation. -Document produced by CTCP.
Tobacco Control	Long-term Compliance with California smoke-free workplace law. (Weber, Bagwell, Fielding, & Glantz, 2003)	Los Angeles County Tobacco Control and Prevention Program	This article: *Policy*- LC 6404.5 statewide and local efforts to enforce Other work by this project:: *Individual, Community, Policy, Media*	-Compliance with smoke-free policies at bars and restaurants over time. -This is one article/evaluation component that is part of a larger project for the County (which is multi-level). -This project is funded by CTCP.
Tobacco Control	The Louisiana Campaign for Tobacco-Free Living (TFL) 2007-2008 Annual Report. (Louisiana Public Health Institute, 2009)	The Louisiana Campaign for Tobacco Free Living (TFL) Program	*Individual* (knowledge, cessation); *Community* (education, media); *School* (college campuses, young people campaigns); *Policy* (SHS, tobacco tax)	-This report is annual report providing background info on project, intervention, evaluation. -Describes short-term/ long-term outcomes, process measures, importance of context. -Variety of sources of data used. -Conclusions: Reduced reported exposure to SHS in car/home/ work; cessation services used and reductions seen in tobacco consumption and increases in quit attempts and referrals; reduced youth "ever smoked" rate. -Document developed by program (TFL).

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Tobacco Control	Comprehensive Evaluation Plan for the New York Tobacco Control Program- Final Report. (Austin et al., 2003)	New York Tobacco Control Program	*Individual* (cessation, education); *Community* (media, services); *Policy* (smoking bans, pricing)	-This report is Evaluation Plan for 5-year project. -Describes multiple steps toward each ultimate program goal. -Acknowledges complexity of contextual factors and synergy across activities. -External contractor for NY State Dept of Health.
Tobacco Control	2009 Independent Evaluation Report for the New York Tobacco Control Program. (RTI International, 2009a)	New York Tobacco Control Program (NY TCP)	*Individual* (cessation, education); *Community* (media, support for tobacco control); *Policy* (smoking bans, pricing)	-Final Evaluation Report after 5 years of project. -Social norms change model. -Uses 3 evidence-based strategies: health communication, cessation interventions, and statewide and community action. -Conclusions: Interventions seem to be effective; changes are being seen in outcomes measured. Compared to the United States as a whole, over the last few years New York has seen: a faster rate of reductions in youth and adult smoking rates; higher rates of quitting smoking, attempts to quit and intentions to quit in the next 30 days; lower daily cigarette consumption. Also found were: reduced exposure to SHS (bars/ restaurants/ other workplaces), increased support for banning tobacco advertising in stores and increased support for "aggressive" tobacco control interventions. -Difficult to link changes directly to specific elements of work, but attributes changes to: high tobacco tax, comprehensive smoke-free air law, statewide and community action impacting policy and above-average funding for tobacco control work. -Detailed description of contextual factors impacting work. -External contractor for NY State Department of Health.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Tobacco Control	The impact of smokefree legislation in Scotland: results from the Scottish ITC Scotland/UK longitudinal surveys. (Hyland et al., 2009)	International Tobacco Control (ITC) Policy Evaluation Project [Part of CLEAN Project above]	*Individual* (behavior- smoking in homes, cessation), (perceptions- public support for legislation); *Community* (pubs, workplaces); *Policy*	-Evaluating impact of Scotland's comprehensive smoke-free policy(pubs, workplaces) and work in homes, cessation. -Quasi-experimental design (control group, UK without law) -Longitudinal national telephone surveys (before law and one year after in Scotland and other places in UK without law; smokers and non included). -Conclusions: reported decline in observing smoking in multiple public places in Scotland, not in rest of UK. Similar change seen in smoke-free homes and number of cigarettes both areas. No difference in cessation. Increase in support for smoke-free policies both areas. Considered policy successful at reducing SHS in various areas without negative outcomes of concern.
Tobacco Control	Scotland's Smokefree Legislation: Results from a comprehensive evaluation. (Haw, 2007)	Clean-air Legislation Evaluation (CLEAN) Collaboration	*Individual* (behavior- homes, cessation), (perceptions- public support); *Community* (pubs/ workplaces, compliance and air quality/cotinine levels); *Policy*	-Overarching evaluation of Scotland's comprehensive smoke-free policy and community work. -National surveys (exposure to SHS, smoking behavior/ norms/attitudes), Qualitative studies (multiple case studies on: bars, community behavior/culture change, smoking in homes), Health impacts (air quality measures, heart attack incidence), cotinine levels non smokers. -Conclusions: High level of support for and compliance with legislation. Improvement in air quality in bars and respiratory health of bar workers. Changing social norms around smoking overall and smoking in the home. Reduction in SHS exposure for non-smoking adults and children. Reduction in hospital heart attack admissions.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Tobacco Control	Longitudinal effects of the European smoking prevention framework approach (ESFA) project in Spanish adolescents. (Ariza et al., 2008)	European Smoking Prevention Framework Approach (ESFA) Project	*Individual* (16 session intervention, cessation program for teachers); *Community* (parents: brochures/info on activities; other areas: youth clubs, tobacco sales); *Environment/ School Policy* (school-based policies)	-Evaluation of Spanish smoking prevention program, focusing on Barcelona youth. -Quasi-experimental design (control group) -Long follow-up period (36 mos) -Addresses attrition, contamination, used multivariate analysis -Conclusions: Impacted adolescent smoking behavior (at least 3 years) with school-based program and coordinated community activities. "Acceptability, collaboration and behaviour change (smoking cessation, if they are smokers) of teachers and adults who influence them" critical for program success (p.496).
Tobacco Control	Robert Wood Johnson Foundation (funder); Research Foundation of the City University of New York (evaluator). (Robert Wood Johnson Foundation)	Tobacco Policy Change Program	*Community;* *Organizational;* *Policy*	-This project is ongoing. Evaluation sounds to be very informative, but not finished until February 2011. - Mixed-methods. Quantitative (online survey, local descriptive data) and Qualitative (document review, key informant interviews, site visits, case studies) data used for the evaluation.
Childhood Obesity	Healthy Eating, Active Communities Phase I Evaluation Findings, 2005-2008. (Samuels & Associates. UC Berkeley Center for Weight and Health. UCLA Center to Eliminate Health Disparities. Abundantia Consulting. Field Research Corporation, 2009)	Healthy Eating, Active Communities (HEAC)	*Community* (food/ physical activity, marketing of these) *School* (food/ physical activity/ environment) *Policy* (food/ physical activity environment)	-Evaluation of: changes in built environment, policy changes (organizational/ legislated), behavioral changes in children (diet/ activity). -Qualitative (focus groups, interviews, policy review, environmental assessment) and quantitative (surveys, environmental assessment) data. -Conclusions: Found changes in environment and behavior with policy changes (in schools, neighborhoods and health care sector). -Evaluation conducted by external evaluation consultant group.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Childhood Obesity	Final Report of the FFY 2008 Impact Evaluation, Sacramento (CA): California Department of Public Health, Cancer Control Branch. (Fourney, Linares, Mitchell, & Sugerman, 2009)	Network for a Healthy California	*Individual* (knowledge, perceptions, self-efficacy, behavior-food/ activity); *Community* (media, access)	-This program is funded by the United States Department of Agriculture (USDA). -Evaluate program for immediate impact (changes in fruit/ vegetable consumption, physical activity and related factors) and long-term outcomes (reduce obesity, body mass index level, related morbidity and mortality and improve health outcomes). -Evaluation of 47 funded contractors for FFY 2008. -Multiple surveys used for measures at pre and post-tests. -Many contractors found significant changes in fruit/ veg consumption. Levels of activity declined from pre-test to post-test -Described contextual challenges: high staff turnover at state and local level; inconsistent use of valid and reliable tools; and diversity of activities makes difficult to attribute changes to specific strategies. -Brings together state/ local data -Doc produced by program.
Childhood Obesity	California Obesity Prevention Plan: A Vision for Tomorrow, Strategic Actions for Today. (Department of Health Services, 2006)	California Obesity Prevention Plan	*Individual* (knowledge, behavior- activity); *Community* (knowledge, healthy environments- food/ activity, marketing); *School* (food/ drink); *Organizational* (promotion of and access to healthy food/ activity); *Policy* (schools, budget)	-Strategic Plan developed in 2006 by the CA state health department. -Coordinate and evaluation project focusing on statewide education, community work and policy around nutrition and activity. -Describes tobacco control work in California as a model for environmental change and use of policy to impact change. -Recommendations for Action around evaluation, including an inventory of existing data, incorporating standardized health indicators into statewide surveys, and addition of new sources of data (such as longitudinal studies). -Document produced by program

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Childhood Obesity	Successful Evaluation Strategies: California. (Sugerman, 2008)	Network for a Healthy California	*Individual* (knowledge, perceptions, self-efficacy, behavior-food/ activity); *Community* (media, access)	-Presentation about Evaluation for Network for a Health California. -Types of evaluations carried out for different parts of program: Process evaluation; Formative Evaluation; Impact/ Outcome Evaluations; Evaluation of Media campaign; Case studies; Evaluation planning/ models. -Different types of data used individually and compiled into annual report. -Document produced by program
Childhood Obesity	Establishing a Baseline to Evaluate Act 1220 of 2003: An Act of the Arkansas General Assembly to Combat Childhood Obesity. (Phillips, Raczynski, & Walker, 2004)	Evaluation of the Arkansas Act 1220 of 2003 to Combat Childhood Obesity Project (Act 1220 Evaluation)	*Individual* (parent/ youth knowledge, beliefs, attitudes & behavior-food/ activity); *Schools (*vending mach. restrictions, cafeteria nutrition, physical activity); *Policy* and *Committees* (overarching between levels, covering statewide, school and district-wide)	-Annual evaluation reports written since the start of the project in 2004. -Information about the history of Act 1220 and baseline data are described in the Year 1 report -Baseline BMI assessment found 38% of the Arkansas' school-aged children were overweight or at risk for overweight. Rates were much higher than previous self-reported estimated had found. -51% of parents of children who were overweight incorrectly perceived the child to be of normal weight.
Childhood Obesity	Year Six Evaluation: Arkansas Act 1220 of 2003 to Combat Childhood Obesity. (Phillips, Raczynski, & Walker, 2009)	Act 1220 Evaluation	*Individual* (parent/ youth knowledge, beliefs, attitudes & behavior-food/ activity); *Schools (*vending mach. restrictions, cafeteria nutrition, physical activity); *Policy* and *Committees* (overarching between levels, covering statewide, school and district-wide)	-after six years, found: positive changes to school environments, nutrition; family knowledge around obesity increased; increase in physical activity and decrease in vending machine purchases and soda reported by students; potential changes to activity, but not nutrition by families; no changes to student BMI. -identifies facilitators and barriers to making changes in schools.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Childhood Obesity	Evaluation of a multiple ecological level child obesity prevention program: Switch what you Do, View and Chew. (Gentile et al., 2009)	Switch what you Do, View and Chew	*Family/Individual* (knowledge, behaviors- food/ activity); *Community* (awareness, media); *School* (reinforce messages)	-Evaluation of immediate and short-term effects (decreasing children's screen time, increasing fruit and vegetable consumption, increasing physical activity). -Control group included. -Children and their parents from 10 schools in two states (Minnesota, Iowa) randomly assigned to group. -Measures at baseline, post, 6 mos. -Effects from small to large seen by group for different outcomes. -University researchers conducted evaluation and wrote report. External to project.
Childhood Obesity	Challenges in assessing the implementation and effectiveness of physical activity and nutrition policy interventions as natural experiments. (Ramanathan, Allison, Faulkner, & Dwyer, 2008)	Ontario Ministry of Education policies: Healthy Foods and Beverages (HF & B) and Daily Physical Activity (DPA)	*School;* *Policy*	-Evaluation of governmental school-based nutrition and physical activity policies. -Describes challenges to evaluating policy interventions (lack of baseline data, wide variation in implementation of policies. -Natural experiments may seem "messy" but they are practical and underused for value they contribute (p. 292). -University researchers conducted evaluation and wrote report. External to project.

FIELD	ARTICLE/ REPORT REVIEWED (CITATION)	PROJECT NAME	LEVELS EVALUATED	BRIEF SUMMARY OF EVALUATION
Childhood Obesity	Evaluation of the childhood obesity prevention program Kids - 'Go for your life'. (Silva-Sanigorski et al., 2010)	Kids-" Go for your life" (K-GFYL)	*Individual/Home* (environment, behaviors); *School* (environment , food, physical activity)	-Program seeks to improve socio-cultural, policy and physical environments in school settings across the state. -Evaluate if participation in program linked to better eating/ activity compared to those not participating. -Mixed-method, cross-sectional. -No control group. -Parent surveys, environmental questionnaires, lunchbox surveys. -Outcomes: increasing fruit/ vegetable/water intake; reducing fat/ salt/ sugar foods and sweet drinks; increasing physical activity; reducing sedentary behavior (screen time). -University researchers conducted evaluation and wrote report. External to project.
Childhood Obesity	A controlled trial of a school-based environmental intervention to improve physical activity in Dutch children: JUMP-in, kids in motion. (Jurg et al., 2006)	JUMP-in, kids in motion Project	*Individual-Parents* (information, activity week), *Kids* (Card activities, Activity week); *School* (sports program, PE program, Class Moves incorporated into school lessons/day)	- Quasi-experimental design (control group) -pre/post data collection -school-based intervention (teachers/students) with parent component added (parents and parents + children) -Assessed effectiveness of physical activity and cognitive changes in kids, comparing intervention and control groups -Links to ecological framework -Process evaluation conducted and found differential implementation between schools -Conclusions: Found changes in physical activity, but not related cognitive changes in kids. Considers influence of environment as potentially more impactful than social cognitive in behavior change. Offering access to physical activity/healthy behaviors (environmental changes) may be enough— changes in cognition may not be necessary in children.

Table 2. Tobacco Control – Potential Cases for Further Study

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
California Tobacco Control Program	*Individual (*cessation, homes/ cars); *Policy and Community* (SHS- bars/ restaurants/ workplaces, outdoor areas; advertising/ marketing at retail outlets, youth access to tobacco products; *Media Campaign;* *Schools* (more work previously, less now)	-Provides oversight and coordination for funded projects: 61 local health departments (LLAs), ~40 competitive grantees. -Also: statewide media campaign, tobacco cessation helpline, and statewide technical support services. -Uses logic model to guide evaluation. -Seeking change in social norms through 4 priority areas: 1) countering pro-tobacco influences in the community (advertising/ marketing of tobacco products, event sponsorship); 2) reduced exposure to SHS (policy/ advocacy work); 3) reduced availability of tobacco (enforcing youth sales laws, eliminate sampling/ tobacco in pharmacies); 4) cessation. -Funded through tobacco tax revenue.	<u>Why Select:</u> -Requires evaluation of all funded projects and conducts statewide evaluation -Considered leader in tobacco control work -County and statewide perspectives/ data/ evaluations available -Use of Logic Models to guide evaluations -Previous working relationship/ contacts with both state and county projects <u>Why Not Select:</u> -Since a "leader" this program has been studied by others (internal/ external) -May not be able to assess contextual role of "funder requirements" in evaluations, since they are often the funder. -Previous working relationship with project	Statewide California, United States

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
Local Lead Agencies (LLAs) funded by the California Tobacco Control Program (CTCP)	*Individual;* *Community;* *Policy;* *Media Campaign*	-61 health departments throughout California are funded by CTCP for tobacco control work. -focus on social norm change through policy and working at several levels/ through several strategies concurrently.	<u>Why Select:</u> -All projects are required by CTCP to develop comprehensive intervention/evaluation plans in 3-year cycles -Previous working relationship/ contacts with both state and LLA projects (may help with buy-in) <u>Why Not Select:</u> -Scope may not match other projects selected -Previous working relationship with several projects	County/ City Health Departments throughout California California, United States
Louisiana Campaign for Tobacco Free Living (TFL) Program	*Individual* (knowledge, cessation); *Community* (education, media); *Schools* (college campuses, young people campaigns); *Policy* (SHS, tobacco tax)	-Program goals: 1) Preventing initiation of tobacco use among youth and young adults; 2) Promoting quitting among adults and youth; 3) Eliminating exposure to secondhand smoke; 4) Identifying and eliminating tobacco-related health disparities; 5) Promoting statewide coordination of all Louisiana tobacco prevention and control programs. -Statewide SHS policy (Louisiana Smoke-Free Air Act, 2007) -Funds community grants for CBOs throughout state (~20/year)	<u>Why Select:</u> -Work at statewide level and local projects funded. 1) Can provide comparison (within same project or with another similar project); and 2) more than one perspective available, if wanted. <u>Why Not Select:</u> -May not be able to assess contextual role of "funder requirements" in evaluations, since they are often the funder. -May be too similar (story/politics/ contextual elements impacting) to other statewide TC programs	Statewide Louisiana, United States

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
New York Tobacco Control Program (NY TCP)	*Individual* (cessation, education); *Community* (media, support for tobacco control); *Policy* (smoking bans, pricing)	-Uses social norms change model. -3 evidence-based strategies: health communication, cessation interventions, and statewide and community action. -Changes seen in outcomes measured. -Attributes changes: high tobacco tax, comprehensive smoke-free air law, statewide and community action impacting policy and above-average funding for tobacco control work. -Detailed description of contextual factors impacting work. -Funds community partnerships/ contractors throughout the state.	Why Select: -Work at statewide level and local projects funded -Evaluation reports readily available (may be more willing to participate/ share info) -Impact of context discussed/ being considered in evaluations Why Not Select: -May not be able to assess contextual role of "funder requirements" in evaluations, since they are often the funder. -May be too similar (story/politics/ contextual elements impacting) to other statewide TC programs	Statewide New York, United States
European Smoking Prevention Framework Approach (ESFA) Project	*Individual* (16 session intervention, cessation program for teachers); *Community* (parents: brochures/info on activities; other areas: youth clubs, tobacco sales); *Environment/ School Policy* (school-based policies)	-Evaluation of Spanish smoking prevention program, focusing on Barcelona youth. - Comprehensive school-based project to change smoking behavior of youth through changes in parents and teachers, policies and youth environments.	Why Select: -More rigorous design than most (quasi-experimental, long follow-up period) Why Not Select: -Smaller scope than some of the other projects- school district in Barcelona	Local- City School Districts Barcelona, Spain

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
Clean-air Legislation Evaluation (CLEAN) Collaboration	*Individual* behavior- homes, cessation, perceptions- public support; *Community*- pubs, workplaces (support, compliance and air quality/cotinine levels); based in intervention by *Policy* in areas mentioned above	-Evaluation of impact of Scotland's comprehensive smoke-free policy on: pubs, workplaces, homes, cessation, public support. -Large coalition involved in evaluation- many partners noted in project documents. -Data collected through: National surveys (exposure to SHS, smoking behavior/ norms/attitudes), Qualitative studies (multiple case studies on: bars, community behavior/culture change, smoking in homes), Health impacts (air quality measures, heart attack incidence), cotinine levels non smokers.	<u>Why Select:</u> -Multiple types of data collected/methods used over time to assess impact of legislation (both before and after passed). -International project (outside United States) may introduce more variation in context/ politics, etc.) <u>Why Not Select:</u> -Nationwide scope is larger than most other projects -International project (outside United States) may introduce more variation in context/ politics, etc.)	National Scotland
Tobacco Policy Change Program; Robert Wood Johnson Foundation	*Community;* *Organizational;* *Policy*	-This project is ongoing. Evaluation sounds to be multi-level and very informative, but not done until February 2011. -Tobacco Policy Change Program is a national initiative of the Robert Wood Johnson Foundation (RWJF).	<u>Why Select:</u> -Cutting-edge project and timing (still in progress) <u>Why Not Select:</u> -This project is not yet completed (February 2011). -Nationwide scope is larger than most other projects	National United States

Table 3. Childhood Obesity – Potential Cases for Further Study

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
Healthy Eating, Active Communities (HEAC)	*Community* (food/ physical activity, marketing of these); *School* (food/ physical activity); *Policy* (food/ physical activity environment)	-HEAC program funded by The California Endowment. -Focus on improving children's environment to reduce obesity. -Partnerships between CBOs, schools, public health departments to create change. -Statewide work and funds (6) community projects. -Evaluation of: changes in built environment, policy changes (organizational/ legislated), behavioral changes in children (diet/ activity). -Mixed methods. -Logic models for program overall and each individual site.	Why Select: -Leader in the field (intervention, evaluation, organization and evaluation contractor) -Work at state level and funds local/ regional projects Why Not Select: -May not be able to assess contextual role of "funder requirements" in evaluations, since they are often the funder.	Statewide/ Communities within state California, United States
Network for a Healthy California: Champions for Change Project	*Individual* (knowledge, perceptions, self-efficacy, behavior-food/ activity); *Community* (media, access)	-Program funded by the United States Department of Agriculture (USDA). -Program focuses on immediate impact (changes in fruit/ vegetable consumption, physical activity, related factors) and long-term outcomes (reduce obesity, body mass index level, related morbidity and mortality and improved health outcomes). -Types of evaluations carried out for different parts of program: Process evaluation; Formative Evaluation; Impact/ Outcome Evaluations; Evaluation of Media campaign; Case studies; Evaluation planning/ models. -Data used individually/ compiled annually.	Why Select: -Described impact of context on evaluation, data collection and implementation of intervention (Annual Report). -Link made to previous tobacco control work in California (modeled after this). May be interesting to see how "translated" from one field to other -Work at state level and funds local/ regional projects Why Not Select: -Unclear ability to assess contextual role of "funder requirements" in evaluations, since they are both funder and funded project.	Statewide California, United States

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
Evaluation of the Arkansas Act 1220 of 2003 to Combat Childhood Obesity Project (Act 1220 Evaluation)	*Individual* (parent/ youth knowledge, beliefs, attitudes & behavior-food/ activity); S*chools (*vending mach. restrictions, cafeteria nutrition, physical activity); *Policy* and *Committees* (overarching between levels, covering statewide, school and district-wide)	-State legislation mandated changes in schools, development of committees and reporting to parents of their child's BMI annually for all children in Arkansas public schools. -program seeks to impact schools environments, parental knowledge, and nutrition and activity of Arkansas children and their families -The evaluation is funded separately, through the RWJF	<u>Why Select:</u> -Leader in the field -Different source of funding than many other projects considered (external to intervention and RWJ foundation) -Separate entity from intervention <u>Why Not Select:</u> -Funding and evaluation separate from intervention different than some other projects	Statewide Arkansas, United States
Switch what you Do, View and Chew	*Family/Individual* (knowledge, behaviors- food/ activity); *Community* (awareness, media); *School* (reinforce messages)	-This is a program of the National Institute on Media and the Family, a non-profit organization. -Evaluation of immediate and short-term effects (decreasing children's screen time, increasing fruit and vegetable consumption, increasing physical activity). -Children and their parents from 10 schools in two states (Minnesota, Iowa) randomly assigned. -Effects from small to large seen by group for different outcomes.	<u>Why Select:</u> -Addresses some of the challenges with these evaluations <u>Why Not Select:</u> -Smaller scope than some of the other projects-only ten schools in two different states	Schools with two states Minnesota/ Iowa, United States
Ontario Ministry of Education policies: Healthy Foods and Beverages (HF & B) and Daily Physical Activity (DPA)	*School;* *Policy*	-Evaluation of governmental school-based nutrition and physical activity policies. -Describes challenges to evaluating policy interventions (lack of baseline data, variation in implementation of policies	<u>Why Select:</u> -International project (outside United States) may introduce more variation in context/ politics, etc. <u>Why Not Select:</u> -International project (outside United States) may introduce more variation in context/ politics, etc.	Province Ontario, Canada

PROJECT NAME	LEVELS EVALUATED	PROJECT OVERVIEW	COMMENTS FOR CASE SELECTION	SCOPE/ LOCATION
JUMP-in, kids in motion Project	*Individual-Parents* (information, activity week); *Kids* (Card activities, Activity week); *School* (sports program, PE program, Class Moves incorporated into school lessons/day)	-School-based intervention (teachers/students) with parent component added (parents and parents + children). -Assessed effectiveness of physical activity and cognitive changes in kids. -Links to ecological framework -Process evaluation found different implementation between schools. -Conclusions: Offering access to physical activity/healthy behaviors (environmental changes) may be enough—changes in cognition may not be necessary in children.	<u>Why Select:</u> -Included process evaluation and found difference based on implementation variation by school -International project (outside United States) may introduce more variation in context/ politics, etc. <u>Why Not Select:</u> -Small scope (two school districts) -International project (outside United States) may introduce more variation in context/ politics, etc.	Schools in two districts Amsterdam, Netherlands
Kids-" Go for your life" (K-GFYL)	*Individual/Home* (environment, behaviors); *School* (environment , food, physical activity)	-Program and evaluation funded by the Department of Health, Victoria and Department of Education & Early Childhood Development. -Program seeks to improve socio-cultural, policy and physical environments in school settings across the state. -Evaluate if participation in program linked to better eating/ activity compared to those not participating.	<u>Why Select:</u> -International project (outside United States) may introduce more variation in context/ politics, etc. <u>Why Not Select:</u> -Evaluation not yet completed or at least reported -International project (outside United States) may introduce more variation in context/ politics, etc.	Statewide Victoria, Australia

<u>Potential Cases for Detailed Study</u>

From the review of recently evaluated multi-level projects came a variety of potential cases for further detailed study. Taking into consideration information that was publically available and published about these projects and evaluations, feasibility (in resources, access and timeline), and opportunities to make comparisons among projects, several options for selection of case pairings were developed.

Beyond reviewing the literature, a multiple case study was carried out by the author to investigate in further detail contextual elements that influenced the comprehensiveness of the multi-level evaluations for projects studied. Whereas previous literature reviews provided

information on the larger picture of comprehensively evaluating multi-level interventions, the study of specific cases will focus in much greater detail on contextual elements described by projects as influencing the comprehensiveness of these evaluations.

Four cases were selected for study, two in the field of tobacco control and two in childhood obesity. Interviews and a document review were used to collect data for each of the cases. Comparison of cases was made, when appropriate, across projects and across fields. Further detail on the case study carried out is provided in the next several chapters of this dissertation.

Presented below are several options developed by the author for selection of cases and a rationale for why this combination of cases may be beneficial for selection or could create particular challenges. These combinations are compared by the fields of tobacco control (TC) and childhood obesity (CO) and scope of the projects. The scope of the project is being used as a consideration for selection because there may be additional noise and complication that makes comparison of projects that are very different from each other more difficult. Since the study is already looking at complex projects, evaluating on multiple levels, in different environments, and across two fields, adding more differences, especially to the context within which the project sits, might make comparison more difficult. For the same reason, programs from outside the United States compared to programs in the United States might be a challenge to analyze across cases and across fields. It may be differences with the scope, political climate or geographic locations that are described, more than the contextual elements influencing the evaluation's development and comprehensiveness. It does not mean that these projects would not be selected, just that this is a factor being considered in the selection and comparison process.

Option #1- Statewide Programs
(TC) CA Department of Public Health: California Tobacco Control Program (CTCP)
(TC) New York Department of Public Health: Tobacco Control Program
(TC) Louisiana Department of Public Health: Tobacco Control Program
(CO) CA Department of Public Health: Network for a Health California Program
(CO) The California Endowment: Healthy Eating, Active Communities (HEAC)
(CO) Arkansas Act 1220 Evaluation

Rationale for Selection:
There are several reasons why selecting the above projects would be beneficial for the multiple case study. All of the projects above were identified through the document review as leading examples in their respective fields (for interventions, evaluations, organization and contractors). These projects have the highest quality evaluations, reports and collaborations that were found during this review. This would allow comparison of "success" cases, as intended for this multiple case study, to learn about promising practices. Adequate funding and prioritizing large-scale evaluations of impact and process were also described for these projects. These projects are also working at a somewhat comparable scope, as they are all evaluating statewide projects. Most of these projects fund local or regional work, so they act as a funder and funded program. Their role as funder and funded program may provide

multiple perspectives not otherwise available. Most of these projects have the intervention and evaluation carried out within the same organization.

One project differs from several others above in a couple unique ways. The Arkansas Act 1220 Evaluation project is funding through a foundation and evaluated separately from the implementation of the intervention. This project is similar to others in that it is considered a leading example in their field and conducts high quality work. This project also has adequate funding and a priority of comprehensively evaluating this statewide intervention. Selecting this project along with several others can provide a nice comparison on several of the elements of similarities and differences.

Only four projects (of the six listed above) will be selected for in-depth study, so selecting a few cases that appear most similar and one or two that seem more different would be of benefit for the multiple case study.

Potential Challenges with Selection:

The influence of the "funder requirement" may be assessed for the project serving as a funder, more than them receiving funding, during interviews, which may make assessment of this element as potentially influencing the comprehensiveness of the evaluation difficult. Another potential challenge is that these projects may be too similar to each other, which might reduce the diversity of responses provided across cases.

Option #2- Smaller Scope Projects Funded by Statewide Programs
These include local programs funded by larger statewide programs, such as CTCP-funded Local Lead Agencies (TC) and HEAC community projects (CO).

Rationale for Selection:

Study of these projects may translate well to provide guidance to other smaller projects looking to expand their evaluation. Also, these projects may enable an easier assessment of the role of "funder requirement" as originally conceived, than projects acting as both funder and funded projects.

Potential Challenges with Selection:

These small projects do not appear as well funded and are not as comprehensively evaluated, compared to the larger statewide programs. They are not necessarily considered "leaders" compared to larger programs, but do work with these larger programs and may have some benefit from that relationship. Promising practices may be harder to identify or be less influential, compared to those found with larger projects. But at the same time, translating these promising practices out into the field may be more realistic for other small projects. These smaller projects may be more impacted by their funder or local context than larger programs. These projects may also need to accommodate to expectations or requirements more than making their own decisions, which may impact how beneficial findings could be, if the projects themselves are not directly making decisions around evaluation development and comprehensiveness.

Option #3- Combination of Smaller and Larger Scope Programs
These could include a combination of smaller school-based or local programs and statewide or national programs in both tobacco control and childhood obesity.

Rationale for Selection:
The combination of different scopes may illustrate variation in programs and influences to the development and comprehensiveness of their evaluations.

Potential Challenges with Selection:
There may be too much noise or variation with the differences in four programs, two different fields, and of very different scopes.

Option #4- Statewide and Local/Regional Programs from Same Geographic Areas
(TC) CA Department of Public Health: California Tobacco Control Program (CTCP)
(TC) County Health Department (Local Lead Agency) or other project funded by CTCP
(CO) The California Endowment: Healthy Eating, Active Communities (HEAC)
(CO) Local/Regional Project funded by The California Endowment/ HEAC

Rationale for Selection:
These projects cover similar geographic areas, both statewide and locally, and all are in California. Comparisons could be made at the organizational level and local project level for both fields. This combination of cases could address questions around influence of "funder requirement" in a different way than just comparing at the larger statewide level (such as in Option #1) or just the local level (such as Option #2). This grouping provides a combination of "leaders" in their fields at the larger statewide level and somewhat less so for the local projects. Including projects from the same geographic areas may reduce the amount of noise or variation, while still enabling comparison for projects of different scopes.

Potential Challenges with Selection:
Studying this combination of cases would likely end up being an in-depth focus on two organizations or programs from both statewide and local perspectives, rather than four unique cases. The influence of funder may be assessed some, but there would still be overlap between the funder and funded project.

Option #5 – Same Two Organizations, Comparing their TC/CO Programs
This could include CA Department of Public Health: (TC/CO) Tobacco Control Program and Network for a Health California; and The California Endowment (or other): (TC/CO) [if comparable project] and Healthy Eating, Active Communities (HEAC)

Rationale for Selection:
Both organizations cover similar geographic areas and both are statewide in California, which can provide some nice comparison and reduced noise. Both groups also do work at organizational level and fund work at the local and regional level for both fields.

Potential Challenges with Selection:

Again with this option, there may be in-depth focus on two organizations, rather than four unique cases. The author has had trouble finding a project funded by The California Endowment in tobacco control that is of similar quality and scope to others being compared. Less information may be gathered about the role of funder through this combination of cases, since both play similar roles to each other.

Detail on the cases selected for further in-depth study, data collected, and findings are presented through the rest of the dissertation.

Discussion

It was more difficult than expected to find all-encompassing articles about comprehensive multi-level evaluations within these fields; comprehensive descriptions of evaluations were particularly rare among the group of articles published in peer-reviewed journals. Evaluation reports, presentations and limited publication documents and articles often provided more detailed information about such evaluations. Length limitations of journals, topics prioritized for publication, and documents prepared for specific audiences likely account for at least some of the differences found between these sources of information. It may also be that overarching evaluation documents are kept internal, while specific elements of the evaluation are developed further for publication or reports made public. For example, an article may be published about one specific component, level or method used within a comprehensive multi-level evaluation. Without the overarching document to tie these pieces together, or information noted in each article, it may be difficult to know whether the evaluation was in fact multi-level or comprehensive once the pieces were put together at the organizational or project level.

Information about dose, fidelity and contextual elements (such as funder requirement for the evaluation and resources available) were particularly difficult to obtain through these published and publically available sources.

Detailed discussions about levels were also difficult to find in many articles. In several articles it was difficult to determine both the levels intervened and those evaluated. Because the articles and reports reviewed were focused on the evaluations, often more information was provided about the levels evaluated, with only a brief description of the intervention. A search of other documents about that project and intervention often needed to take place to learn this information. This additional information was critical to determine if both the intervention and evaluation were multi-level. The distinction between levels attempted versus those actually completed, for both the intervention and evaluation, was particularly absent in these articles. Overall, information was most often provided about the levels evaluated, methods for collecting data and results, with some information on levels intervened, and rarely information presented about intended versus completed intervention or evaluation activities.

Information gathered from the review of recent evaluations was used by the author to determine projects of interest for further detailed study. Considerations for selecting cases

included the match between projects, or cases, and how these cases might be able to similarly and differently describe elements of interest. A study of several cases was carried out, as described later in this dissertation, to further understand how their multi-level evaluations were developed and contextual elements that influenced the comprehensiveness of their evaluations.

Limitations

The current chapter relied heavily on literature reviews and publically available sources of information; therefore, there are several limitations for this reason alone. Only information published in peer-reviewed journals or published online in some other format were included in the current reviews. There are likely large amounts of information in working documents or internally used information that did not make it to a publically available source. Additionally, limitations exist as to how searches are conducted and the possibility that other sources were available publically or in peer review journals, but not captured in the current review. The necessary step of limiting search terms and search engines used may have resulted in some missed information. The use of publically available information limited the amount of information known about recent evaluations conducted. If information was not in the articles or reports reviewed, and not found elsewhere, it was left unknown about that project.

Conclusion

The two literature reviews presented in this chapter were used to build a working Framework for elements in a comprehensive evaluation of multi-level interventions, as well as, review recent evaluations of multi-level interventions for many of these elements. A next step included identifying potential projects for further in-depth study.

The Framework for Comprehensive Evaluation of Multi-Level Interventions (Figure 1) described three sets of elements for inclusion: elements for standard evaluation, elements to be given additional focus for multi-level interventions, and contextual elements influencing comprehensiveness of the evaluation. These three sections were meant to build upon one another to describe comprehensively evaluating the most simple, to more complex, multi-level interventions. To comprehensively evaluate a multi-level intervention, all elements identified would be included and contextual factors taken into consideration.

A review of recent multi-level evaluations of multi-level interventions was also conducted. Articles and reports were reviewed from the fields of tobacco control and childhood obesity. Of the multi-level evaluations reviewed, the most frequent combination being evaluated was at the socioecological individual, community and policy levels concurrently.

Challenges in the collection of data, particularly in the review of recent evaluations, focused around information not readily available in publicly available sources. Several elements intended for review were difficult to obtain, and a complete picture of projects,

interventions and evaluations often unavailable, at least as searched through these publically available sources.

Information gathered from this review of recent evaluations was used in a later part of this study, to further understand how multi-level evaluations of specific projects were developed and how contextual elements influenced the comprehensiveness of their evaluations.

CHAPTER 4: A MULTIPLE CASE STUDY OF CONTEXTUAL ELEMENTS INFLUENCING COMPREHENSIVE EVALUATIONS OF MULTI-LEVEL INTERVENTIONS IN TOBACCO CONTROL AND CHILDHOOD OBESITY: A DESCRIPTIVE PRESENTATION OF CASES

Introduction

This chapter presents a multiple case study of four projects evaluating multi-level interventions in tobacco control and childhood obesity. The author studied the selected projects, or cases, to understand the contextual elements influencing their comprehensive evaluations of multi-level interventions. For the projects studied, both the interventions and their evaluations are multi-level. That is, they seek to intervene and evaluate across multiple levels.

An earlier chapter presented a working Framework for Comprehensive Evaluation of Multi-Level Interventions (see Figure 1 in Chapter 3), as well as, a review of recent evaluations of multi-level interventions for many of the elements identified in the Framework (see Tables 1-3 in Chapter 3). These reviews were conducted to assess elements to be included for comprehensive evaluation of multi-level interventions and whether these elements were found in recently conducted multi-level evaluations within tobacco control and childhood obesity.

The Framework for Comprehensive Evaluation of Multi-Level Interventions included three sections: 1) elements for a standard evaluation, 2) elements to be given additional focus for multi-level interventions, and 3) contextual elements influencing comprehensiveness of evaluation. After development of the Framework (in Chapter 3), later sections of this dissertation, including the cases studied and related analyses and conclusions, focus exclusively on only one part of the Framework, the contextual elements influencing comprehensiveness of evaluation. Contextual elements identified in the Framework include: resources, evaluation requirement by funder, training/experience of evaluator, length of time carrying out intervention/evaluation, intended use of evaluation, access to information to guide evaluation, evaluator access, evaluation capacity of project/organization, agency support, social/environmental conditions, politics, and competing organizations/interests. These contextual elements were the focus of the four cases studied, and are presented in the current chapter.

Through a review of recent evaluations (also in Chapter 3), based on publicly available documents, it was difficult to find all-encompassing articles and reports about comprehensive multi-level evaluations. In particular, information about dose, fidelity, contextual elements (such as funder requirement for the evaluation and resources available) and levels attempted for intervention and evaluation were often lacking or provided with limited detail. In several articles it was difficult to determine both the levels intervened and those evaluated. Because the articles and reports reviewed were focused on the evaluations, often more information was provided about the levels evaluated, with only a brief description of the intervention. The distinction between levels attempted versus those actually completed, for both the intervention and evaluation, were particularly absent in

these articles. The collection of data relevant to studying multi-level interventions was difficult. Several elements intended for review could not be obtained, and a complete picture of projects, interventions and evaluations were often unavailable, at least as searched through these publically available sources.

Because data of interest were not readily available, as evidenced by the review described above and in Chapter 3, it therefore made sense to study multi-level evaluations of multi-level interventions in further detail through a multiple case study with several projects. The intention of studying specific projects was to understand how their multi-level evaluations were developed and how contextual elements influenced the comprehensiveness of their evaluations. Gathering information about these cases directly enabled the collection of data otherwise not available.

The research question of focus for the multiple case study was: *What are contextual elements influencing comprehensive evaluation of multi-level interventions in tobacco control and childhood obesity?*

Data were collected through interviews and document reviews with each project studied. Analysis within and across cases and fields were conducted with these data. A descriptive presentation of each of the four cases is provided in the current chapter. Each case is presented individually, to provide background on that specific project and to describe contextual elements identified by the cases as influencing the comprehensiveness of their evaluations. The next chapter presents findings across these cases and across the two fields.

Methods

A multiple case study was conducted with projects carrying out multi-level evaluations of multi-level interventions in the fields of tobacco control and childhood obesity. Context is considered important and investigated through case study, as opposed to experimental studies where context is often intentionally ignored or something to be controlled (Miles & Huberman, 1994; Stake, 2006; Yin, 2009). It is important to understand not only specific details of individual cases, but also how several cases are similar or unique when compared (Stake, 2006). Evidence from a multiple case study is often considered more persuasive than that from a single case study (Yin, 2009).

Selection Criteria

A previously conducted review of current multi-level projects identified a variety of potential cases for further study (see Chapter 3 for more detail). Taking into consideration information that was publicly available and published about these projects and evaluations, feasibility (in resources, access and timeline), and opportunities to make comparisons among projects, several options for selection of cases were developed.

The purpose of this multiple case study was to investigate contextual elements that influenced the comprehensiveness of the multi-level evaluations studied. Contextual elements identified in the previously developed Framework (see blue shaded boxes in Figure 1, Chapter 3) were the focus for the study of these cases. These contextual elements included: resources, evaluation requirement by funder, training/experience of evaluator,

length of time carrying out intervention/evaluation, intended use of evaluation, access to information to guide evaluation, evaluator access, evaluation capacity of project/organization, agency support, social/environmental conditions, politics, and competing organizations/interests.

Case Selection

Cases were selected purposefully, with the intent, as described by Stake (2006), to "build in variety and create opportunities for intensive study" (p. 24). Variety was included in order to better understand contextual elements that played an important role in developing and implementing comprehensive evaluations by projects studied.

Specific purposeful sampling strategies were used in the selection of cases, including intensity sampling and theory-based sampling. Intensity sampling is the use of good or above average examples, which were the projects found and selected for use as cases (Patton, 2002). Projects with comprehensive evaluations of their multi-level interventions were selected as cases for study. It was assumed, and found true, that even projects with comprehensive evaluations were able to identify challenges encountered, strategies used to overcome barriers, and lessons learned. Selecting above average examples of projects provided the opportunity to learn about promising strategies and potential key elements for developing a comprehensive evaluation, which would not be as readily available from projects with a less comprehensive evaluation. An additional factor for case selection was the match with other cases, to enable comparison with each other and elements of the theoretical framework. This was carried out through theory-based purposeful sampling. Cases were selected to match each other in the scope of their work, such that all cases are large-scale, multi-million dollar, statewide projects. The cases selected had differing sources of funding, different strategies for implementing the interventions and evaluations, and had a mix of using internal and external evaluators. The combination of these factors influenced the selection of specific cases for this study.

Four projects carrying out multi-level evaluations of multi-level intervention programs were selected for participation, two projects in tobacco control and two in childhood obesity. Further discussion about why the fields of tobacco control and childhood obesity were selected for study was presented in Chapter 2. Briefly, these two fields were selected for study because: they are two fields in public health in which multi-level interventions and evaluations have been implemented and continue to gather momentum, both fields utilize socioecological and related models to describe influence and interaction among various factors, lessons learned in tobacco control have been borrowed by some in the field of childhood obesity, and both fields represent serious health concerns of great interest currently.

The author considered it important to include more than one case for each field, but the study also needed to stay within resources available, including the lead researcher/author's time, funding, and timeline. Cases were selected with the intent of making comparisons between the four cases and across both fields. Further detail about the review process conducted to find these projects initially was presented in Chapter 3. The following criteria were used in the selection of these four cases, such that all projects:

- are evaluating on multiple socioecological levels;
- are evaluating a program intervening on multiple socioecological levels;
- are leading examples in their respective fields;
- have funding and have prioritized large-scale evaluations of impact and their evaluation's process;
- are working statewide, with additional work at local and regional levels (both interventions and evaluations);
- may serve as both as a funder (of local projects) and funded program (by their funding entity).

Interview Participant Selection

Purposeful sampling was used to select interview participants. The intent was to interview key informants with the most applicable and richest sources of information, while also representing different roles, perspectives and involvement in the projects studied. Once a project agreed to participate, a project lead was identified and that person helped to determine individuals within the project with specific knowledge around the evaluation. Based on their role and expertise, specific individuals within the project were selected for participation. At the end of the each interview, participants were asked for the names of others within the project who would be useful to interview. This use of snowball sampling, as described by Patton (2002), provided a few additional individuals to interview. It also provided confirmation that the most appropriate people within the project were being interviewed. All individuals approached for interviews agreed to participate.

Purposeful sampling of cases and interview participants reduces the opportunity to statistically generalize, compared to other sampling strategies (Patton, 2002), but it should be emphasized that this is not usually the intent of case studies and was not the intent in the current study. Obtaining information and perspectives of specific information-rich projects and individuals, by using purposeful sampling strategies, worked well for the current project.

Data Collection

Interviews

Twelve semi-structured telephone interviews were conducted with decision-makers in the four cases studied. The number of interviews possible within each case varied, as there were only a limited number of people with the knowledge available to answer questions being asked. Key informants interviewed included project directors, internal evaluation staff, and external evaluation contractors. Following is a breakdown of interview participants. There were three interviews conducted for Case 1, two interviews conducted for Case 2, four interviews conducted for Case 3, and three interviews conducted for Case 4. Of these, seven individuals were project staff involved in both the intervention and evaluation and five were evaluators focused exclusively on the evaluation component of the project. It should be noted that project staff have evaluation expertise and also serve as evaluators in several of the projects studied. Information about interview participants is provided to add detail about

who the informants were, without providing enough detail to identify anyone individually within the project.

On average, interviews lasted approximately 57 minutes, with a range of interview lengths from 29 minutes to 126 minutes. Interviews were conducted between August and October 2010.

Project staff and evaluators involved in evaluating the selected cases were interviewed. The purpose of these interviews was to better understand the perspective and experience of the projects being studied. These interviews provided an opportunity to gather rich information about the context within which the project and evaluation were situated and developed. Examples of promising practices, challenges encountered and lessons learned, specific to the project's evaluation, were examined. Questions explored evaluation elements included in the multi-level project and why these elements were included. The role of the project's funder, capacity of evaluators and staff, and contextual factors were also examined, to determine the influence they had on the comprehensiveness of the project's evaluation.

An interview protocol was developed, which included interview questions and instructions for conducting the interviews. Before any interviews were conducted, exempt status approval from the Committee for Protection of Human Subjects at the University of California, Berkeley was obtained.

Document Review

Evaluation reports and other relevant project documents were reviewed prior to interviews. This review of documents provided additional information about the projects and the context for the case being studied.

Analysis

Interview Summaries

Interviews were audio-recorded and notes taken during interviews. After each interview, a period was set aside to write further detailed notes about the interview. Interview recordings were used to fill in notes and obtain direct quotes after interviews were completed. Recording of interviews helped to provide further rigor than taking notes during the interview alone. Comprehensive interview summaries were developed from notes and recordings. An empirical study comparing results obtained through summary reports versus the use of full transcripts indicated no substantial difference in conclusions drawn (O'Sullivan, Jay, & Powers, 2004). Working from a summary report versus full transcripts was required due to the limited time and money available for this research.

Coding

Before beginning analysis, a "start list" for coding was developed from the working framework and research questions, which in turn also influenced the development of the interview questions (Miles & Huberman, 1994). The start list was developed before the summary interview reports. As coding began on the summary reports, new codes were developed and added to the list, as appropriate, and originally developed codes were modified and discarded when not needed.

The researcher coded the summary reports using open-coding in ATLAS.ti, version 6, and through this process identified major themes.
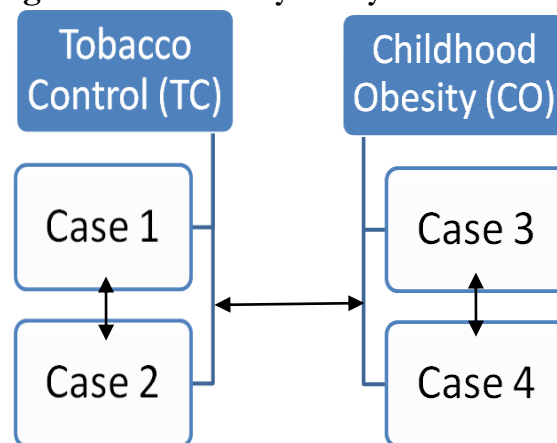
Case Study Analysis

Data from cases studied, including interview and document review data were analyzed in several ways. Interview and document review data were analyzed within the context of the case for which it was collected, as well as, within and across its field (tobacco control/childhood obesity). Three phases of the analysis are described in further detail through Figure 1 and the text following Figure 1.

Interview summaries were imported into ATLAS.ti, version 6, for coding. Each interview was coded individually, with all interviews from the same case linked together as a "family" in ATLAS.ti. Coded data from each case/family were exported from ATLAS.ti and used for the within-case analysis. Multiple respondents were included for each case and a compilation of their perspectives used as a collective representation for the case. All informants were treated as having equal weight and expertise to speak about the project when analyzing data.

Data display tables, such as those described by Miles and Huberman, are utilized to present findings from cross-case and cross-field analyses (Miles & Huberman, 1994).

Figure 1 depicts the analysis of case study data, as conducted, in three phases. A further description for each stage is presented below Figure 1.

Figure 1. Case Study Analyses Conducted

Phase 1: Analysis of data for each individual project/case. Therefore, four cases were analyzed individually, as depicted by the white boxes. These are presented individually for each case study in the current chapter.

Phase 2: Cross-case analysis within each field. Two cases were analyzed together within tobacco control and two cases within childhood obesity, as depicted by the arrows between the white boxes. These analyses are presented in the next chapter (Chapter 5).

Phase 3: Cross-field analysis comparing the fields of tobacco control and childhood obesity, as depicted by the arrow linking both sets of cases. This analysis is presented in the next chapter (Chapter 5).

Case Write-Up

   Data from interview summary reports and documents reviewed for each case were written up to represent the case for each project studied. The descriptive write-up for each case in this chapter follows the same organization, which includes the following sections:

- Project Description;
- Benefits and Complications of Evaluating on Multiple Levels;
- Contextual Elements Identified by Project Participants as Influencing (Case) Evaluation;
    - Sources of Information Guiding Development of the Evaluation
    - Resources
    - Evaluation Requirement by Funder
    - Length of Time Carrying Out Intervention/ Evaluation
    - Intended Use of Evaluation
    - Evaluator Access and Capacity
    - Evaluation Oversight and Consequence
    - Other Contextual Factors
    - Summary for Contextual Elements
- Project Strengths in the Evaluation;
- (Case) Provided Guidance to Others;
- Challenges Encountered When Evaluating the Project;
- What Project Would Have Done Differently; and
- Summary of Case.

Member checks

   After cases were written up, a copy was provided to each person interviewed, to provide "member checks." This additional step served to confirm accuracy of information presented about each project, as well as quotes attributed to each interviewee. Participants interviewed were asked to provide feedback about whether their statements were reflected accurately and whether conclusions developed were reasonable and plausible (Maxwell, 2005; Yin, 2009). Since quotes are not reported with identifying information or names, individuals were provided with the interview number linked to them, so they could check that quotes attributed to them, through their number, were accurate. The inclusion of feedback from previous participants provides a check of reliability—that what is reported is accurately what was described during interviews. Eleven of the 12 persons interviewed responded. Minor feedback was provided by several participants, to clarify terms or statements made. Respondents indicated that the write-up for their project was accurate, quotes were representative of what they had said, and several indicated they found the write-up interesting.

Case 1 - California Tobacco Control Program (CTCP)

Project Description

      The California Tobacco Control Program (CTCP) began with the enactment of the California Tobacco Tax and Health Promotion Act of 1988, or Proposition 99 (Prop 99). This legislation was passed through a ballot initiative and added a tax onto tobacco products starting in January 1989. The program's budget has varied greatly, from a high of approximately $107 million (in fiscal year 2001-2002), to a low of $37 million (in fiscal year 1995-1996) (Roeseler & Burns, 2010). For the fiscal year 2008-2009, the CTCP's budget was nearly $55 million (Roeseler & Burns, 2010). The program has received a limited amount of additional money from Master Settlement Agreement (MSA) and the CDC (California Tobacco Control Program Representative, 2010).

      Evaluation of the CTCP was mandated as part of the legislation, which required the evaluation of both local and statewide tobacco control efforts (Tang et al., 2002). Funding for the evaluation has varied greatly over the last 20 years, but in this time has averaged about 5.6% of the total budget on evaluation (California Tobacco Control Program Representative, 2010). The project currently budgets approximately $4 million per year to do surveillance and evaluation (Interview 2. California Tobacco Control Program, September 2010). CTCP's evaluation began immediately after the enabling legislation passed, in order to get baseline data (Interview 1. California Tobacco Control Program, September 2010). The evaluation of the project has changed greatly since the CTCP started, but it was included as part of the project from the beginning.

      Work of the CTCP is carried out through local projects throughout California (both health departments and other competitively funded projects), a statewide media campaign, tobacco cessation helpline, and statewide technical support services.

      Four priority areas are targeted by the CTCP through their comprehensive tobacco control program (California Department of Public Health, 2009):

- countering pro-tobacco influences in the community;
- reduced exposure to SHS;
- reduced availability of tobacco; and
- cessation.

      This project works on the above priority areas to change social norms by intervening on multiple socioecological levels concurrently. Innovative ways to address tobacco control include intervening on multiple levels concurrently, such as the individual, family, school, community, organizational, and policy levels. The goal is often to use multiple avenues to create a change in social norms, expectations and environments around tobacco for children, their families and communities (Chapter 2 provides further detail about socioecological levels). The CTCP intervenes on the following socioecological levels (with activities at each level presented in parentheses): *individual* (cessation, education, homes/cars); *community* (smoke-free workplaces and outdoor areas, retail advertising, tobacco marketing, youth access to tobacco products); and *policy* and *media*, which both cross multiple levels (California Tobacco Control Program Representative, 2010).

Evaluation for this project also takes place on multiple socioecological levels concurrently. <u>Evaluation</u> for the CTCP takes place on the following socioecological levels (measuring areas of): *individual* (cessation, smoking rate, education, homes/cars); *community* (smoke-free workplaces and outdoor areas, retail advertising, tobacco marketing, youth access to tobacco products); *policy* (smoking bans, pricing, tobacco distribution/marketing) and *media* (California Tobacco Control Program Representative, 2010).

Several different evaluations take place to measure the implementation and effectiveness of the CTCP. First, there are measures of overall state progress, using tobacco control indicators of smoking prevalence, tobacco consumption, exposure to secondhand smoke, intentions to quit, etc. (Interview 1. California Tobacco Control Program, September 2010). Next, there are measures of impact through school, media and local project interventions. Additionally, state and community projects evaluate their own work. A last element of the evaluation involves tracking the context within which the intervention takes place (Interview 1. California Tobacco Control Program, September 2010).

Early in the project's history the CTCP contracted with external evaluators for the evaluation work, both to collect various sources of data and to put them together into a cohesive whole to tell the story of the project's impact. Evaluation capacity has since been built within the project and these activities are now conducted internally. Some evaluation efforts continue to be contracted out, but the majority of the analyses are now performed by the CTCP.

To provide ongoing accountability for Prop 99 money, an Independent Evaluation was carried out from 1996 through 2000 by the Independent Evaluation Consortium, which included team members from the Gallup Organization, Stanford University and the University of Southern California. Evaluation work by this group used a logic model to assess intermediate and ultimate outcomes on multiple levels. This group produced multiple reports and publications from evaluation data collected (Independent Evaluation Consortium, 2002; Rohrbach et al., 2002).

Findings were presented in the Independent Evaluation Consortium's final report around the: tobacco marketing environment in California (tobacco sponsored events, print advertisements, bar/club promotions, retail outlets, direct mail); local tobacco control projects (countering pro-tobacco influences, reducing secondhand smoke, reducing youth access to tobacco); statewide media campaign, school-based tobacco use prevention education program; and overall impacts of the program (Independent Evaluation Consortium, 2002). A few of the overall impact conclusions from this final report included: 1) that most Californian's were exposed through at least two different levels (media, community, school) to tobacco control messages, and 2) higher exposure to CTCP activities were linked to: lower perceived access to tobacco by youth, reduced adult smoking prevalence, and increase in bans for smoking in homes (Independent Evaluation Consortium, 2002). A previous report by this group concluded that CTCP work may have reduced adult smoking and exposure to secondhand smoke (Rohrbach et al., 2002). While a comprehensive tobacco control program was indicated as important and effective, the specific elements influencing change could not be determined (Independent Evaluation Consortium, 2002).

Additional evidence about the impact of CTCP's work, from a more recent study used a simulation model of the CTCP. The study, by Levy, Hyland, Higbee, Remer & Compton (2007), indicates that this simulation model estimated that "tobacco control policies reduced smoking rates in California by an additional 25% relative to the level that they would have been if policies were kept at their 1988 level" (p.167). The article attributes this change to price increases, media policies, clean air laws, and youth access policies (Levy et al., 2007). Levy et al. (2007), further explain that "The model estimates that over 5000 lives will be saved in the year 2010 alone as a result of the CTCP and industry-initiated price increases, and that over 50,000 lives were saved over the period 1988–2010" (p.167).

Three Stages of Evaluation for the CTCP

The evaluation of the CTCP has changed dramatically over time. The earliest stage of the evaluation was mostly built upon surveillance work. But the project staff felt a disconnect between what the University researchers were measuring and what the project thought should be measured (Interview 2. California Tobacco Control Program, September 2010). "The overall evaluation structure was heavily impacted by, at least initially, trying to show that the program works…So the surveillance approach was a good idea at the beginning" (Interview 2. California Tobacco Control Program, September 2010).

Then, in the 1990's, project staff wanted to better understand the impact of various pieces of the CTCP towards moving tobacco control in California forward. The evaluation looked at each piece of the project separately, and also together to look at overall impact of the project on short-term and intermediate outcomes (Interview 2. California Tobacco Control Program, September 2010). This was done through Independent Evaluation contracts. This stage of the evaluation was multi-level and quite comprehensive, at the same time, project staff felt that if they could build internal capacity that they might overcome challenges that still existed (Interview 2. California Tobacco Control Program, September 2010).

Currently the CTCP evaluates their project internally through a combination of evaluating specific elements and assessing overall impact. "…we are evaluating each piece, trying to make each more efficient and effective in its own way, then qualitatively putting the pieces together for an evaluation overall. The pieces are put together internally now rather than contracting out" (Interview 2. California Tobacco Control Program, September 2010).

> *"There is a strong belief that the media and program go together and breaking them apart doesn't make sense. The evaluation framework was built around that concept, at least in the first two pieces. In the third, maybe we were more secure with ourselves. We didn't care, we just wanted to make each piece stronger" (Interview 2. California Tobacco Control Program, September 2010).*

The Role of Logic Models and Social Norm Change in Evaluating the CTCP

The CTCP uses a social norm change paradigm as the basis for their intervention, evaluation and logic model. Roeseler and Burns (2010) describe the CTCP's perspective on social norms as:

"The goal is to change the broad social norms around the use of tobacco, and to indirectly influence current and potential future tobacco users on a population level by creating a social environment and legal climate in which tobacco use becomes less desirable, less acceptable and less accessible" (p.i-6).

An example of the project's current logic model can be found in a recent paper by Roeseler and Burns (2010). This logic model includes the project's inputs, priority areas, outputs, and short-term, intermediate and long-term outcomes.

CTCP's logic model was described through interviews as combining not only the program's theory of what is going on and changes anticipated, but also how this relates to data collection and utilization. Benefits of the logic model are described below.

"Unless everything is derived from a common theory of change, a common logic model, if you will, the data that come back may not help you understand what's happening. You may be seeing many different pieces of results, but it is not making a mosaic, because there's no way to fit them together. You need the substrate in which to place all these pieces of information, a common model or approach. And that's hard work intellectually and it's hard work to convey that to everybody, so that everybody is on the same page and collecting data to contribute to what will be a common understanding of change" (Interview 3. California Tobacco Control Program, September 2010).

The program's first logic model was not developed until several years after the program had started. At this time the CTCP was already moving toward a focus on social norm change, which was then translated into the program's first logic model. This focus on social norm change can still be seen in the project's intervention, evaluation and logic model.

"Tobacco control people were moving toward the philosophy of social norm change and outcome-focused evaluation and not just caring about counting things done. They wanted to know what mattered and wanted to reward people who were making change happen and succeeding though the logic model pathway and not just giving a formula for dollars" (Interview 3. California Tobacco Control Program, September 2010).

"The development of that model came probably almost ten years after the program was in place. Putting into words what had been happening. Jon Lloyd penned that and made an unbelievable contribution to the field…His ability to capture what was going on and put it down in an easily understood manner allowed everyone, the program, media, everyone, to rally around what was the fundamental theory or model" (Interview 3. California Tobacco Control Program, September 2010).

Early influences from the Independent Evaluation era and the Stanford Five-City comprehensive, integrated model can still be seen in the comprehensiveness of the evaluation and logic model for the project currently.

"It has changed dramatically over time, but that framework laid the foundation for design and methods….We do this with media and local program evaluation...That influenced us heavily, now that I look back…I can see some of the touches of the evaluation model still there" (Interview 2. California Tobacco Control Program, September 2010).

A more recent contribution to the project, in terms of furthering the evaluation of the CTCP, came through a system developed initially for a different purpose. The Communities of Excellence (CX) approach, used by the CTCP, was influenced by work being done through the National Institutes of Health (NIH) around chronic disease indicators, as well as indicator efforts by the American Cancer Society and others (Interview 3. California Tobacco Control Program, September 2010). The CTCP developed and later refined this system, which now tracks indicators of change that each of their funded projects is working to address.

"It did not start out as an evaluation or planning tool, it started out as a way to recognize excellence. That's where the term comes from. Which program or grantee is doing an excellent job? We want to recognize them. What criteria should we be using to reward them?...That has turned out to be the central planning and evaluation system for local programs. That organized a lot of people around a model, what we were aiming towards, the big picture, and what our philosophy is, and what the specific outcomes are that we can focus on. And if we focus on these outcomes we know we can move the needle on social norm change. That really helped make what is a complicated program into something that you could teach or train people on. It simplified it" (Interview 3. California Tobacco Control Program, September 2010).

Benefits and Complications of Evaluating on Multiple Levels

Representatives from the CTCP were asked to describe how the project benefited by evaluating all levels of the intervention. These key informants were also asked how the project was made more complicated by evaluating all levels of the intervention.

Benefits of evaluating the CTCP with a multi-level evaluation were described by interview participants as providing accountability, eliminating areas that are not working, having a comprehensive evaluation, and reducing assumptions that may be made otherwise.

A multi-level evaluation, such as that carried out by the CTCP, provides "transparency, credibility, accountability" and allows for "improvement and defending against ineffective interventions" (Interview 1. California Tobacco Control Program, September 2010). Some examples provided by the project about the benefits of evaluating on multiple levels include: "We are trying to be comprehensive in evaluating all levels and trying to be effective in doing that" (Interview 2. California Tobacco Control Program, September 2010) and "We've stopped working in areas that we found weren't useful or productive" (Interview 2. California Tobacco Control Program, September 2010). "We think we have most things covered. But if we left things out, we would have to make assumptions

about what was causing things or what was not working. We at least have evaluation touching most pieces" (Interview 2. California Tobacco Control Program, September 2010).

Complications of carrying out a multi-level evaluation of a multi-level project were identified, including the complexity that comes with the large scope of the evaluation, making sense of the data, and assessing potential synergistic effects among multiple levels of intervention.

The complexity of the project overall, and putting multiple pieces of the evaluation together into a comprehensive whole, provides a particular challenge for the evaluation. "There had to be an evaluation that measured the entire program's progress" (Interview 1. California Tobacco Control Program, September 2010). This becomes more complicated when the project's work spans multiple socioecological levels, including sometimes multiple strategies within each level, and is carried out by a variety of community and statewide projects. Of particular challenge for the CTCP is "…trying to look at the synergistic effects. Looking at school, media and local programs—do you get synergistic effects?" (Interview 1. California Tobacco Control Program, September 2010).

There is both breadth and depth to the intervention, and likewise, the evaluation of the project (Interview 3. California Tobacco Control Program, September 2010). "The CTCP is like a big elephant. There are a lot of parts to the elephant" (Interview 3. California Tobacco Control Program, September 2010).

> *"Obviously if you have many different sources of data that are being collected, it is more complicated…but not just the data collection and funding, and all the moving parts of that, and managing all that, but when you have data come back from all these different evaluations, you have to have some means of making sense of it all. That's probably harder than funding a comprehensive multi-level evaluation, is making sense of it" (Interview 3. California Tobacco Control Program, September 2010).*

> *"In the beginning of the Program, stakeholders consistently wanted to measure and compare the effectiveness of the school, media, and local program interventions. CTCP staff have always promoted the concept that it's the comprehensive and synergistic effect of the intervention that have contributed to the declines in the key indicators and promoted anti-tobacco socials norms. Thus there has been resistance to this idea of finding out which one of these things works best. Rather, the idea is to find out how each is working and to use that information to improve performance. There were a lot of complaints during the first decade of the program over the use of the social norm change approach and that there wasn't enough emphasis on the provision of direct cessation services or "youth prevention" (Interview 1. California Tobacco Control Program, September 2010).*

Contextual Elements Identified by Project Participants as Influencing CTCP's Evaluation

Project representatives were asked to describe contextual elements influencing the evaluation for the CTCP. In addition to an opportunity for the project to describe other contextual elements influencing the evaluation, questions were asked about specific contextual elements. Multiple contextual elements were described by project representatives

as having influenced the evaluation for this project, both in the development of the evaluation and the comprehensiveness.

The following contextual elements were asked about with interview participants and are described in the following section: Sources of Information Guiding Development of the Evaluation, Resources, Evaluation Requirement by Funder, Length of Time Carrying Out Intervention/ Evaluation, Intended Use of Evaluation, Evaluator Access and Capacity, Evaluation Oversight and Consequence, and Other Contextual Factors.

Sources of Information Guiding Development of the Evaluation

Guidance for developing CTCP's evaluation has come from a variety of sources over the last twenty years or so. As described previously, the CTCP has gone through three distinct stages of their evaluation. Some of the sources of information providing specific guidance in the development of the evaluation include the legislation, advice from various groups, published literature, and the institutional knowledge of evaluation consultants.

The legislation impacted the development of the evaluation, such that specific criteria were included and maintained throughout the course of the program.

Advice was sought from multiple groups over the years, too many to capture comprehensively here, but some of the most influential are described here. An advisory committee, the Evaluation Task Force, made up of experts in tobacco control research, policy, practice and other areas, provided guidance for the evaluation (Interview 3. California Tobacco Control Program, September 2010). The Evaluation Task Force provides advice, rather than oversight.

> *"We can take it or leave it, but we are interested in getting feedback. There isn't politics or grandstanding going on, that sometimes happens in public meetings. The goal of this group is to help us…Multi-level and comprehensiveness has always been an interest of the group…That group has been excellent for us to have" (Interview 2. California Tobacco Control Program, September 2010).*

Additionally, the CDC and National Cancer Institute's ASSIST program were consulted to get ideas for the evaluation (Interview 1. California Tobacco Control Program, September 2010). Others playing a role included university based evaluators, local stakeholders, and specific individuals with expertise in tobacco control and multi-level evaluations (Interview 1. California Tobacco Control Program, September 2010).

The Western States Evaluators Group has been particularly helpful for advice and continued support in recent years (Interview 1. California Tobacco Control Program, September 2010; Interview 2. California Tobacco Control Program, September 2010). This group is made up of tobacco control evaluators at a similar level of sophistication, and doing work similar to, the CTCP (Interview 2. California Tobacco Control Program, September 2010). It has been a valuable group to share ideas about data collection instruments, data, and other evaluation-related topics (Interview 1. California Tobacco Control Program, September 2010).

Published literature around the ASSIST program and Stanford Five-City project were described by a project staff as influential in the development of the CTCP's evaluation (Interview 1. California Tobacco Control Program, September 2010).

With the Independent Evaluations came a wealth of experience and institutional knowledge. Contractors had experience developing and carrying out comprehensive evaluations on other multi-level projects, in tobacco control and other chronic disease prevention areas.

> *"We drew on our own experiences and publications of work we had done ourselves in doing comprehensive community evaluations, starting in the 70's with the three community study and five city project, comprehensive heart disease prevention programs funded by NIH. We drew on our own experiences of how to design a comprehensive integrated program. We had a lot of involvement from our colleagues in health communication, behavioral scientists, public health, etc. etc. We used a lot of health communication models to talk about the presumed causal chain that we were trying to measure. We were trying to deal with modeling a very complex intervention in a free living population and seeing what the impact of that intervention was on important outcomes, intermediate and longer term outcomes"* (Interview 3. California Tobacco Control Program, September 2010).

The CTCP did not rely on the guidance of other departments within their own organization, currently titled the California Department of Public Health, but rather went to outside sources for advice and feedback related to the project's evaluation. All participants for this case described sources other than CTCP's home organization as providing guidance. "We really went to places like CDC and NCI for help, not internally…The field of Alcohol and Drugs didn't really apply because they do things so differently. They had a different intervention and focused on individual behavior change" (Interview 1. California Tobacco Control Program, September 2010).

> *"On the program side, there were a lot of people that moved from HIV over to the program, at the beginning, but I'm not sure they pulled any of the evaluation pieces really. The program is influenced a little bit from that, with the advocacy and rallying cry, but not evaluation"* (Interview 2. California Tobacco Control Program, September 2010).

> *"We drew a lot on our own experiences, on behavioral science, on health communications. Public health really didn't have a lot to contribute at that time. I think the California program, frankly, has changed public health more than the other way around. So, it was really from these other fields that we were drawing from"* (Interview 3. California Tobacco Control Program, September 2010).

It appears that other tobacco control programs within the Western States Evaluators Group may be in a similar situation as well.

"I would say more recently, that group is where I am drawing from, not from within our organization. We are dealing with the same subject topic, but I think they are in the same situation we are. They are not drawing from their other programs. Probably because tobacco control in their states are better funded and further along…That's where most of the experience is being drawn from right now, in terms of evaluation" (Interview 2. California Tobacco Control Program, September 2010).

Resources

The most valuable resources to the evaluation of this project were identified by project staff members as the funding available, accessibility and advice by others, and at one time, outside researchers and evaluators (Interview 1. California Tobacco Control Program, September 2010; Interview 2. California Tobacco Control Program, September 2010). By far, funding was described by project informants as being the most valuable resource available. Money helps to not only fund evaluation activities, but also provides additional staff and allows evaluation work to be contracted out when needed (Interview 2. California Tobacco Control Program, September 2010).

"Straight up, funding helps. Nobody has the funding we have to do evaluation and surveillance. We get an absurd amount compared to everyone else in the department...Those resources allow us to do things much more comprehensively, in terms of evaluation. It allows us to look at a number of different levels for the evaluation."

In the earlier years of the project's evaluation, when Independent Evaluations were carried out, funding was also described as influencing the evaluation.

"It was well funded as compared to what they had done in the past, but probably needed more funding. We made compromises in the evaluation plan and in the proposal. We had to back into the design we came up with on the basis of what the budget limitations were. That's a huge thing. It's not something that can be ignored, is how much money there is and how expensive it is to do this kind of work, in a quality way…I can't remember what the budget was, but whatever it was, we couldn't do a complete statewide design. We selected 18 counties. It was a complicated selection process. It wasn't just selecting the top counties…Had we had more money we would have been in 58 counties, but we weren't. Did that affect the results? One will never know. The funding issue is always going to be a big contextual factor" (Interview 3. California Tobacco Control Program, September 2010).

A lesson learned, as developed by Novotny and Siegel (1996), included not only funding as necessary, but also the consistency of that funding. While they were speaking specifically about the intervention and program, it seems reasonable to link this lesson to the evaluation of the program as well. They described that:

"…successful tobacco control cannot be achieved through year-by-year funding appropriations, it requires sustained planning with reliable revenue flows over at least a five-

year period. It is impossible to create a strategically planned, long-term tobacco control program in the face of dramatic shifts in program resources and directions" (p. 69).

Evaluation Requirement by Funder

  Specific elements of the evaluation are included to meet the Prop 99 mandate, including surveillance activities and monitoring trends for particular tobacco control indicators, such as prevalence and consumption. These are done through the California Tobacco Survey, California Adult Tobacco Survey, and the Behavioral Risk Factor Surveillance Survey (Interview 2. California Tobacco Control Program, September 2010). Two other pieces have been added to the evaluation by CTCP, which are now mandated, but were not part of the original requirements. These are the evaluation of the in-school work done by the Department of Education and the Youth Tobacco Purchase Survey, which surveys stores throughout the state to determine the rate of illegal sales of tobacco to minors (Interview 2. California Tobacco Control Program, September 2010).

  Other elements of the evaluation were added to meet the broad language of evaluating the different components of the project, such as media and local programs. "It was CTCP's job to operationalize the legislative requirements and to make it successful" (Interview 1. California Tobacco Control Program, September 2010).A project representative also stated that "The enabling legislation is very broad and flexible, so most of what we do falls under there. They set out certain items that we were required to track…also talks about evaluating each of the components of the program" (Interview 1. California Tobacco Control Program, September 2010).

  Beyond the core elements of the evaluation, or those legislatively mandated, the project had flexibility to add evaluation elements and later take them out or and change them, to meet the needs of the program.

> *"Core evaluation elements track the progress of program, but then we add other components in as needed and then phase out. For example, smoking in movies isn't in the legislation, but certainly something we have authority to do. If money goes down too far, we take this out" (Interview 1. California Tobacco Control Program, September 2010).*

  The CTCP adds "a bunch of other pieces. This gets back to the comprehensive piece" (Interview 2. California Tobacco Control Program, September 2010). The expansive nature of the media and local program evaluations, work to build capacity around evaluation, and several monitoring projects and special surveys done by the CTCP, are not required by the legislation (Interview 2. California Tobacco Control Program, September 2010). "It's not just accountability, but also seeing what's going on locally and trying to move California forward through communities" (Interview 2. California Tobacco Control Program, September 2010).

  For one of the independent evaluations, an evaluation plan was developed initially by the external evaluation contractor, then along with the CTCP, and through a series of meetings and revisions, the two parties agreed upon the program's plan for evaluation (Interview 3. California Tobacco Control Program, September 2010). "The contractor helps

you refine your vision and you help the contractor move towards a place that you think they should be. It's a negotiation. It's not always pleasant, but it is necessary" (Interview 3. California Tobacco Control Program, September 2010).

Length of Time Carrying Out Intervention/Evaluation

A clear theme of the CTCP and its evaluation evolving over time came across in the interviews with project representatives. "This didn't happen overnight. It took many years to get to this point. But that's what the evaluation has evolved into" (Interview 3. California Tobacco Control Program, September 2010). A project representative also described that "The program has been evolving. It's become more sophisticated over time. We didn't know the right way to do things in the beginning" (Interview 1. California Tobacco Control Program, September 2010).

Intended Use of Evaluation

"Intended use has been really important" (Interview 1. California Tobacco Control Program, September 2010). The program works to collect data that "helps drive policy adoption and that also measures the impact of policy and system changes" (Interview 1. California Tobacco Control Program, September 2010). "It's believed that somebody will read the conclusions and try to implement something from the evaluation. That's true for CTS, the schools project, and even when we write internal reports for media" (Interview 2. California Tobacco Control Program, September 2010).

Data utilization was described by those interviewed as being complicated for the program. Not only because of the many sources of data involved, but also making sense of the data once collected (Interview 3. California Tobacco Control Program, September 2010). During one of the Independent Evaluations there was a strong push for developing an evaluation for the program that focused on the utilization of data.

> *"The evaluation, to use a Patton term, was quite utilization-focused, more so than you might think for an independent evaluation…We were very collaborative, to the point of running surveys past them and getting their input on things. It was a huge collaborative effort with the state…No one from the state asked us to change any results, but they were very involved in helping us to refine the questions we were studying and asking…I felt a commitment to be utilization-focused, because of the stakes, because of what we were dealing with here. Many millions of dollars towards programs that needed to be evaluated in a way that would help policy-makers and program leaders make informed decisions about future investments. It translates into lives. And it's an important thing" (Interview 3. California Tobacco Control Program, September 2010).*

Evaluator Access and Capacity

Early evaluation experiences taught the CTCP a lot about working with external evaluation contractors, the direction they thought the project's evaluation should head, and that ultimately, building internal capacity for evaluation was the best route for them.

The CTCP has varied the level of internal versus external evaluation work being conducted. In the beginning years of the program the evaluation was externally contracted out for independent evaluations, which were conducted by several different organizations over many years. A mismatch between evaluator goals and that of the CTCP, disconnect between what the intervention was trying to do and what the evaluation was measuring, and the complexity and expense in hiring an outside consultant played roles in the current status of the evaluation. "I think historically there's been a lack of sufficient evaluators who are qualified to look at a social norm change interventions" (Interview 1. California Tobacco Control Program, September 2010). Tang et al. (2002) describes the experience of the CTCP related to the capacity of evaluators, such that "…many evaluators are not experienced at evaluating community-level changes. They are frequently more knowledgeable and comfortable designing and conducting evaluations that measure individual changes in knowledge, attitudes, beliefs and behaviors" (p.53).

> *"A large number of the external evaluation contractors don't have a great understanding of what the program does. They don't know what really goes on in terms of local programs or what the media is supposed to be doing. Whenever we get evaluations from contractors we have to work with them. They really understand population data, but the interpretations are off sometimes, in my mind. We have to work with them extensively sometimes to get reports and documents to read in a way that makes sense. We aren't changing their interpretations, but often times they just don't understand certain things" (Interview 2. California Tobacco Control Program, September 2010).*

The CTCP learned valuable lessons from previous experiences and decided to build their own capacity and oversee their own evaluation internally. "In the beginning we thought we would just contract out the entire evaluation, but then about nine years into program we realized we needed to build internal capacity. I think that was really important" (Interview 2. California Tobacco Control Program, September 2010).

> *"I think the experience of that independent evaluation was part of the impetus for the program developing their own internal capacity…There is a comprehensive evaluation going on, but it has not been put in the hands of one independent evaluation contractor…They decided to invest in their own internal capacity. I think it was a very good decision" (Interview 3. California Tobacco Control Program, September 2010).*

Finding evaluators with the knowledge and capacity to work with the CTCP has been an ongoing challenge, but the project has developed some strategies to find them. "Engaging evaluators that understand the intervention and what it is that we are trying to do. We were lucky that way. Eventually we found some really key people. And the people that didn't get it were dropped off" (Interview 1. California Tobacco Control Program, September 2010).

> *"Recruitment is always difficult, for people who understand evaluation. On my staff we've taken a lot of epidemiologists, sociologists and psychologists and try to make them*

understand evaluation, rather than taking people who are experienced in evaluation. Building internal capacity has been difficult. I think we're getting there, but it's been a challenge" (Interview 2. California Tobacco Control Program, September 2010).

"There are people that have been involved in tobacco control for a number of years. People who dabble in a bunch of different fields don't seem to grasp social norm change stuff. It's more the number of years engaged in tobacco control" (Interview 1. California Tobacco Control Program, September 2010).

Capacity for local projects evaluating their own work has also been a challenge. The CTCP has provided technical assistance, funded a statewide Tobacco Control Evaluation Center, adopted common indicators for projects to draw from, and developed requirements for evaluations of projects they fund (requiring that each project have a local evaluator and at least 10% of the project's budget be used for evaluation) in an attempt to build capacity. For the evaluation of local interventions there are also "challenges in tying policy change, compliance, etc. with quantifiable intermediate tobacco control indicators" (Interview 1. California Tobacco Control Program, September 2010).

Evaluation Oversight and Consequence
The CTCP has ongoing oversight and an awareness of their accountability to various parties, including the Tobacco Education Research Oversight Committee (TEROC) and the public. Previously the project has dealt with legislative hearings and external scrutiny through public records requests and external evaluations (Interview 1. California Tobacco Control Program, September 2010). Tang et al. (2002) described the role of evaluation for the CTCP, such that "because the state agency is accountable for the use of public funds, a high-quality and credible comprehensive evaluation for the state program is indispensable" (p.55). "There's always been a high degree of accountability in this program, a lot of public scrutiny…Evaluation is always being done because such requirements for accountability" (Interview 1. California Tobacco Control Program, September 2010).

Originally TEROC and the legislature played a larger role in this, but it has reduced over time. Focus seems to have shifted elsewhere or be on only specific elements of the program and evaluation. "Over time, the oversight and consequence piece isn't as intense as it used to be" (Interview 2. California Tobacco Control Program, September 2010). TEROC is still concerned if elements are not done, especially legislatively mandated pieces (Interview 2. California Tobacco Control Program, September 2010).

The Branch Chief would likely be the person held responsible, if evaluation activities were not completed, but with that position not filled, responsibility would likely fall to the person overseeing the evaluation on behalf of the program.

Another group, the Evaluation Task Force, seems more interested in the evaluation of the CTCP. There is not oversight or consequences from them, but rather feedback and interest in the evaluation (Interview 2. California Tobacco Control Program, September 2010).

Other Contextual Factors

Politics, especially in the earlier years of the program, were quite influential on the program and the development of its evaluation. An active debate was underway in the 1990's about the value of a social norm change approach, both from a scientific standpoint (whether there was social norm change and if a mass media campaign could contribute to that), and a political standpoint (whether the program should spend money on media against the tobacco industry) (Interview 3. California Tobacco Control Program, September 2010).

> *"Going after the industry, was that an appropriate use of funds? And the evaluation became quite politicized because of that. So, we were pulled and pushed in multiple directions. Some people were carrying the mantle of wanting to show and prove that this mass media strategy was going to change social norms and that would reduce prevalence, disease and disability. Others were hoping that intervention didn't happen and were hoping that nothing substantive would come out of the evaluation. They would hamper the evaluation. They didn't even want to fund it. We spent a lot of time running up to Sacramento and meeting with political appointees. There were lots of politics, it was pretty intense. Those were the times we were working in when developing this" (Interview 3. California Tobacco Control Program, September 2010).*

The tobacco industry is another contextual element to always keep in mind related to this project. "Intervention isn't occurring in a vacuum. It is occurring in the context of the tobacco industry, so you have to evaluate the results in terms of what's going on within that context" (Interview 1. California Tobacco Control Program, September 2010). Novotny and Siegel (1996) report that through June 1993, "Proposition 99 reduced cigarette consumption by 802 million packs" in California, for a loss of "$1.1 billion in pre-tax sales and approximately $286 million in profits for the tobacco industry" (p.67). The CTCP was seen as a critical threat by the tobacco industry and the CTCP has been challenged through multiple ways since Proposition 99 was passed.

In addition to the tobacco industry, in the 1990's, this program also faced unexpected challenges from groups such as the California state legislation, state administration (Governor's office), and California Medical Association (Novotny & Siegel, 1996). As described by Novotny and Siegel (1996),

> *"Despite sustained public support, the tobacco control component was weakened by political actions of the tobacco industry and also by the competing efforts of organized medicine and the lack of support from the executive and legislative branches of government. Nevertheless, Proposition 99 succeeded in reducing exposure to environmental tobacco smoke, cigarette consumption, and smoking prevalence among adults in California" (p.59).*

The state civil service hiring process was described by project staff as a "huge barrier to recruiting and retaining highly skilled researchers and evaluators" (Interview 1. California Tobacco Control Program, September 2010). This has improved some, but people having to

take an exam just to get an interview and State salaries were other barriers identified (Interview 1. California Tobacco Control Program, September 2010).

Summary for Contextual Elements

Project representatives were asked to describe contextual elements influencing the development and comprehensiveness of the evaluation for the CTCP. Of particular influence on the evaluation of this project were: utilizing multiple sources of information available to the project, having adequate funding, intended use of the evaluation and data, access to evaluators with capacity to do quality work on complex evaluations, and the evolution of the project to what it is now, which has occurred over time. Additionally, the political climate, particularly in the earlier years of the program, was described as a large contextual factor impacting the CTCP and the development of its evaluation.

Project Strengths in the Evaluation

When asked what the project did best, specific to the evaluation, several strengths and strategies for success were described. These include collecting baseline data, consulting with external evaluation experts, building internal capacity, valuing evaluation, providing guidance to others, and adaptability.

Collecting baseline data at the beginning of the program may seem intuitive, but other states have not collected this data, and later regretted it. The CTCP did and is glad to have this point of comparison (Interview 1. California Tobacco Control Program, September 2010).

CTCP representatives also felt that engaging an external evaluation task force, made up of experts, to advise the program on the evaluation was quite beneficial. Also, working with specific knowledgeable external evaluators that cared greatly about the program and developed an evaluation to measure the impact of the program has helped the program reach the success it has found today. "That was incredibly valuable to us" (Interview 1. California Tobacco Control Program, September 2010).

Building internal capacity for the CTCP has been important. Originally the program contracted out the evaluation, but learned about nine years into the project that building internal capacity was necessary for them (Interview 1. California Tobacco Control Program, September 2010). They have since been sharing this lesson learned with others. "I think that was really important. When talking with other state departments, like Mental Health, I kept advising them that they need to have a robust internal capacity" (Interview 1. California Tobacco Control Program, September 2010).

Finding evaluators that understand the intervention has been a challenge, but something the project has actively worked at and eventually found success.

> *"Engaging evaluators that understand the intervention and what it is that we are trying to do. We were lucky that way. Eventually we found some really key people. And the people that didn't get it were dropped off" (Interview 1. California Tobacco Control Program, September 2010).*

Valuing evaluation is another area identified by project participants as contributing to the CTCP's success. "There's always been a huge value placed internally on evaluation. It wasn't just something because it was in the legislation that we did it, but people saw the value" (Interview 1. California Tobacco Control Program, September 2010).

The CTCP also feels it has done a good job of providing strong guidance and tools for use by projects at the local level (Interview 1. California Tobacco Control Program, September 2010). Surveys, other evaluation tools, and guidance by the CTCP have also benefitted several other state and international tobacco control programs.

The scope of the evaluation, being both multi-level and comprehensive was also described during interviews as something the CTCP has done well.

> *"I think the comprehensive multi-level piece is probably best. Someone decided along the way that we were going to do all of those pieces and that all of the pieces were important to provide feedback to the program to make it better. In my mind that happened. Whether point of sale, monitoring, surveillance, doing legislatively mandated pieces, or whatever it is, those multi-level comprehensiveness pieces is the best part. I guess what I am saying is the scope of the evaluation" (Interview 2. California Tobacco Control Program, September 2010).*

Lastly, flexibility is acknowledged as an important factor in the program's success. "We started off with a plan, but adaptability I think is key, at least in our process" (Interview 2. California Tobacco Control Program, September 2010).

CTCP Providing Guidance to Others

The CTCP has worked with the CDC, other states, and even other countries, to share intervention and evaluation advice and provide assistance related to tobacco control work (Interview 1. California Tobacco Control Program, September 2010; Interview 2. California Tobacco Control Program, September 2010; Interview 3. California Tobacco Control Program, September 2010).

Beyond just providing advice, work developed by the CTCP has been used by others in several ways. The National Adult Tobacco Survey is based on the California Tobacco Survey and the program's media evaluation is now being replicated by others (Interview 1. California Tobacco Control Program, September 2010). They have also developed a good working relationship with the Attorney General's Office. The Attorney General's Office uses data collected by the CTCP, from Youth Tobacco Purchase surveys, local tobacco projects, and the like, to help inform their investigations and negotiations with the tobacco industry (Interview 1. California Tobacco Control Program, September 2010). "They go out and collect some of their own, but our data gives them a starting point to know where the hot spots are" (Interview 1. California Tobacco Control Program, September 2010).

CTCP's evaluation has influenced other parts of the country in very meaningful ways. The quote below highlights this point:

"They knew what we were doing in California. It was state of the art. Nobody was doing anything like it, the multi-level evaluation. He got California's permission to bring the Independent Evaluator's surveys to Florida…They took those surveys and customized them for Florida and mounted a statewide evaluation along with the statewide rollout of the brand new program. They didn't have to invest years of development work because it was provided to them. That was a direct impact of that Independent Evaluation. I'm not sure if they realize the influence" (Interview 3. California Tobacco Control Program, September 2010).

<u>Challenges Encountered When Evaluating the Project</u>

While the CTCP has had many successes related to its project and evaluation, several challenges were also identified. Some of the biggest challenges related to the evaluation included: capturing the diversity of work by local projects and tying into the larger evaluation of the CTCP, expectations for data based on funding available, lack of value for evaluation, and administrative difficulties, among others.

As previously mentioned, the complexity of this multi-level intervention and evaluation provides its own set of challenges for the evaluation. Further discussion of this issue is presented in the section titled "Benefits and Complications of Evaluating on Multiple Levels."

The diversity of interventions being implemented and differences in community capacity make it difficult to evaluate the effectiveness and impact of the state and community interventions component (Interview 1. California Tobacco Control Program, September 2010). As described by Tang et al. (2002) "Tying the local program evaluations to the overall evaluation of the TCP in California has been and will continue to be difficult because of the sheer number of local programs (approximately two hundred) and their diversity" (p.52).

Understanding exposure, let alone impact, on different target groups was described as a challenge as well. It is difficult to "understand the exposure to the intervention…who was exposed, the extent to which they were exposed and intensity of exposure. In school-based evaluation that was easier…local program interventions was much more difficult" (Interview 1. California Tobacco Control Program, September 2010).

Even though the CTCP is well funded, the amount of work that can be done within the funding available can be a challenge at times as well.

"Expectations are really high. People want county and population-specific data, but we don't have funding to do that. We use the California Tobacco Survey. We get that for demographic groups every three years, but people want it more frequently. People say that telephone surveys don't work in their population, so they want us to do a door-to-door survey in their community. And we don't have enough money to do that. So money is a big issue" (Interview 1. California Tobacco Control Program, September 2010).

While current internal value on evaluation was described by project staff as a strength of the project, this was not always the case and evaluation is still not particularly valued by everyone associated with the project. An ongoing challenge for the project's evaluation is getting legislators and local projects to value evaluation and put additional funding into

evaluation activities. Two project representatives described two different reasons why evaluation is both valued and a challenge for the CTCP. The first involves a change in how evaluation is valued internally; evaluation used to be less valued, but now is considered a strength or success of the project. The second reason is a distinction between value placed on evaluation within the project versus from those outside of the project, but still involved in project decisions being made. Legislators and local projects still do not see the value in evaluation the way CTCP staff do, so there is resistance to putting extra money available towards evaluation. "The first ten years of the program, trying to get other people in the program to believe that evaluation is important has been a challenge. Evaluation was ignored for a long time" (Interview 2. California Tobacco Control Program, September 2010).

> *"Every once in a while there is a little more Prop 99 revenues than were anticipated and we're asked how we want it used and we always say that we want a piece of that money to go towards the evaluation, but people in the legislature don't want to put more money towards evaluation. They would rather put it towards intervention…There is also a lack of value on evaluation at the local level. Again, they would rather put the money into the intervention" (Interview 1. California Tobacco Control Program, September 2010).*

One interview participant described that administratively the evaluation is now more difficult (Interview 2. California Tobacco Control Program, September 2010). At one time there was only one evaluation contractor handling the evaluation. Now it is a combination of evaluation done in-house and pieces contracted out that are then put together internally.

Other challenges related to the evaluation include defensiveness and fear around evaluation findings, a lack of control groups in most instances, and in 2008 an Executive Order by the Governor that suspended all evaluation contracts, even those in the middle of data collection (Interview 1. California Tobacco Control Program, September 2010).

The project never knows what the next "big thing" is going to be or what will be important now and gone tomorrow. Such changes in the field and project naturally had impact on the evaluation. "We don't know what will be important or not. People are working on menthol right now, but if the FDA passes something, then it's gone. It can change and that's difficult" (Interview 2. California Tobacco Control Program, September 2010).

What the Project Would Have Done Differently

The CTCP reflected on things they would have done differently, specific to the evaluation. Developing internal capacity for evaluation is something the project would have done sooner, in retrospect (Interview 1. California Tobacco Control Program, September 2010). The evaluation unit has greatly expanded over the last several years. Conducting the evaluation within the project has better integrated the evaluation and program planning, but has likely slowed their dissemination of findings (Interview 3. California Tobacco Control Program, September 2010).

The CTCP would also have liked to have "published more and sooner" (Interview 1. California Tobacco Control Program, September 2010). There are a lot of barriers around

getting your data published through the state and data analyses have to get approved (Interview 1. California Tobacco Control Program, September 2010). They have recently published several articles in a supplement of the journal *Tobacco Control*. The anniversary supplement issue was described by one participant as a "heroic effort" to get publications out about the CTCP (Interview 3. California Tobacco Control Program, September 2010). "It is very, very hard for state personnel running a multi-million dollar program to have the time to write those kinds of papers versus an academic who is funded to write" (Interview 3. California Tobacco Control Program, September 2010).

A project representative also described a regret related to surveying for youth smoking prevalence. The method they chose did not work well to meet the needs of the project and was different than what others in the field ended up doing (Interview 1. California Tobacco Control Program, September 2010). A change was made, several years later, but the project would rather have had comparable data earlier on.

For a policy-based program, there is more the CTCP feels they can do to evaluate the work local projects do towards making policy changes. The project is currently working to better determine the effectiveness of community projects (Interview 2. California Tobacco Control Program, September 2010). "We don't know exactly how that works and we haven't done a very good job of showing how to make it work" (Interview 2. California Tobacco Control Program, September 2010). The CTCP is putting together a new policy tracking system to help address this new area in their evaluation.

Summary for CTCP

The CTCP seeks to change social norms around tobacco and has developed a multi-level intervention to accomplish this and a comprehensive multi-level evaluation to assess the program. An evaluation was mandated by the enabling legislation, Prop 99, but the specifics of what the evaluation would include and the extent to which a comprehensive evaluation was developed was up to the project to determine.

Evaluation of the CTCP has evolved through three distinct phases, from a focus on surveillance, to a more comprehensive multi-level evaluation conducted by external evaluators, to now carrying out the evaluation in-house.

Benefits of evaluating the project with a multi-level evaluation were described by project participants as providing accountability, eliminating areas that are not working, having a comprehensive evaluation, and reducing assumptions that may be made otherwise.

Complications of carrying out a multi-level evaluation of a multi-level project were identified, including the complexity that comes with the large scope of the evaluation, making sense of the data, and assessing potential synergistic effects among multiple levels of intervention.

Project representatives reported that the sources of information having the most influence on the development of the evaluation were the elements mandated by the enabling legislation, institutional knowledge from external evaluators, and seeking guidance from experts in evaluation and tobacco control.

Other contextual elements identified by project representatives as influencing the comprehensiveness of the evaluation included: having adequate funding and access to

evaluators with capacity to do quality work on complex evaluations, the intended use of the evaluation and data, and the evolution of the project to what it is now, which has occurred over time. The political climate, as well as the tobacco industry, particularly in the earlier years of the program, had an impact on the project and its evaluation.

Strengths of the project's evaluation were identified by project staff as including: building internal evaluation capacity, having access to and consulting with evaluation experts, valuing evaluation, adaptability to change the evaluation as needed, providing guidance and tools to local projects and other states and countries, having collected baseline data, having adequate funding, and using a logic model to frame the evaluation.

Challenges to the CTCP's evaluation were described by participants as: the complexity of the project, linking the diverse work of local projects to the whole of the program, building internal capacity, expectations based on funding available, lack of value for evaluation, administrative difficulties around contracting, and not knowing the next "big thing" and how it would impact the evaluation.

As noted previously, both the valuing of evaluation and the development of internal evaluation capacity were described by project representatives as both strengths and challenges. The value placed on evaluation seems related to two different factors. First, evaluation used to be less valued within the project, but has since become a strength or success of the project. Second, evaluation was described as being valued by the project, but there is still a lack of value by legislators and local projects. This disconnect has been noticed when additional funding has become available and CTCP staff wants to use it for evaluation and other groups are resistant; they would rather use extra money for intervention activities. Internal evaluation capacity was originally not present in the CTCP. Evaluation was conducted by external evaluation contractors, but over time they realized that building evaluation capacity would be of great benefit to the project. Challenges with access, capacity, hiring and retaining internal evaluators have all been issues dealt with by the CTCP. Internal evaluation expertise now allows the project to carry out their evaluation in-house, contract out for work as needed, put various pieces of the evaluation together themselves to assess the overall impact of the project, and be a recognized leader in their field.

The CTCP is working to continue improving internal capacity for evaluation, publishing more often, and assessing in further detail the effectiveness of their local projects throughout the state.

It should be noted that all statements about elements influencing the evaluation are based on information provided by project representatives during interviews. Therefore, they are the ones attributing influence and describing the influence on their project and the evaluation.

Case 2 - New York Tobacco Control Program (NYTCP)

Project Description

The NYTCP started with the legislated Health Care Reform Act of 2000 (HCRA), which also mandated an independent evaluation of the program. Previous to the development of the NYTCP, limited tobacco control work was done through the New York State Department of Health starting in 1991 (RTI International, 2004). NYTCP's implementation began "slowly in 2000, then took off in 2002 with new personnel and development of a strategic plan. Evaluation started slowly in 2003 with awarding of the contract to RTI and development of surveillance instruments and protocols" (New York Tobacco Control Program Representative, 2010a).

This project and its evaluation are primarily funded through New York state funds, from the HCRA. The budget has varied greatly, from a high of approximately $85 million, to currently around $65 million annually (New York Tobacco Control Program Representative, 2010a; RTI International, 2009a). On average, about 7% of the program's budget is spent on evaluation, ranging from approximately $4.5 million to $6.5 million annually (New York Tobacco Control Program Representative, 2010a). Recent dramatic budget cuts have required that strategic decisions be made about how much can be spent on various evaluation activities, so that the program can meet annual reporting requirements to the legislature and still conduct a comprehensive evaluation (Interview 1. New York Tobacco Control Program, August 2010). Additional funding has been obtained by the program from the Centers for Disease Control and Prevention (CDC), since about 2000, and more recently, the American Recovery and Reinvestment Act of 2009 (ARRA), each for specific purposes (New York Tobacco Control Program Representative, 2010a). While the NYTCP is well funded compared to other state tobacco control efforts, its entire budget is only a fraction of what the tobacco industry spends on advertising and promotion, which was estimated at around $830 million annually in 2002 for New York State (RTI International, 2004).

As described in the First Annual Independent Evaluation Report (RTI International, 2004), four goal areas are used by the NYTCP to promote a comprehensive tobacco control program, including:

- Eliminating exposure to secondhand smoke;
- Decreasing social acceptability of tobacco use;
- Promoting Cessation; and
- Preventing initiation among youth.

This project works toward the above goals by intervening on multiple socioecological levels concurrently. Innovative ways to address tobacco control include intervening on multiple levels concurrently, such as the individual, family, school, community, organizational, and policy levels. The goal is often to use multiple avenues to create a change in social norms, expectations and environments around tobacco for children, their families and communities (Chapter 2 provides further detail about socioecological levels). The NYTCP intervenes on the following socioecological levels (with activities at each level

77

presented in parentheses): *individual* (cessation, education); *community* (support for tobacco control); *media* (cross-cutting, aimed at the individual to promote cessation, community to change social norms/attitudes, and improve support for policy); and *policy* (smoking bans, pricing, retail point-of sale restrictions) (New York Tobacco Control Program Representative, 2010a). Intervention activities are carried out through work by the NYTCP and their contractors throughout the state. NYTCP has contractors through five modalities, including: community partners, youth action programs, cessation centers, school policy partners, and colleges for change (New York Tobacco Control Program Representative, 2010b).

  Evaluation for this project also takes place on multiple socioecological levels concurrently. <u>Evaluation</u> for the NYTCP takes place on the following socioecological levels (measuring areas of): *individual* (cessation, education); *community* (support for tobacco control); *media*; and *policy* (smoking bans, pricing, retail point-of-sale restrictions) (New York Tobacco Control Program Representative, 2010a).

  The NYTCP externally contracted out their evaluation, which produced an evaluation plan and annual reports starting in 2003. For detailed information about the background of the program, tobacco control in New York State, the intervention, evaluation design, outcomes measured and findings, among other topics covered, refer to reports developed by the external evaluators at RTI International (Austin et al., 2003; RTI International, 2004, 2005, 2006, 2007, 2009a, 2009b).

  By 2009, the independent evaluation concluded that the NYTCP interventions seem effective. When compared to the United States as a whole, New York has seen reductions in smoking rates, higher rates of quitting smoking, attempts and intent to quit, lower daily cigarette consumption, reduced exposure to secondhand smoke (in bars, restaurants and other workplaces), increased support for banning tobacco advertising in stores and increased support for "aggressive" tobacco control interventions (RTI International, 2009a). While this evaluation cannot identify the specific elements of the intervention that have most impacted the changes seen, several factors are considered to have contributed to these changes, including: high tobacco tax, comprehensive smoke-free air law, statewide and community action impacting policy change, and above-average funding for tobacco control work (RTI International, 2009a).

<u>Benefits and Complications of Evaluating on Multiple Levels</u>

  Representative from the NYTCP were asked to describe how the program benefited by evaluating all levels of the intervention. These interview participants were also asked how the evaluation was made more complicated by evaluating all levels of the intervention.

  Benefits were described by key informants as learning things that would have otherwise been overlooked, being able to cut things out of the program that are not contributing in the way that had been expected, and better understanding influences that are at play. A few examples from program interviews making these points:

> *"You don't know where you are going to learn something. You don't know what you are going to be able to improve or what you're going to learn. You might think you are, but*

you're wrong, if you think that in advance. By doing all of it we can learn from what exists and you don't miss anything" (Interview 1. New York Tobacco Control Program, August 2010).

"There were other things we did that that we abandoned. We gained very little from doing that work, so it was stopped. You start to make strategic decisions, especially when budget starts to intervene. If you are doing something and not learning much from it, abandon it" (Interview 1. New York Tobacco Control Program, August 2010).

"The standard benefit of multi-level is understanding the context in which any one intervention is passing and how people are influenced by that. You can account for other influences that you may falsely attribute to the intervention if you are not accounting for the multiple levels of influence"(Interview 2. New York Tobacco Control Program, August 2010).

Complications described by project representatives included the amount of data being collected, organized and reported on, as well as access to that data from the beginning.

"You have multiple studies, surveys and multiple deliverables. It becomes a burden just to track all of this stuff. The dissemination of reporting becomes very complex. What do you do with all that information? It's complicated by just the nature of being multi-component and multiple levels. Many surveys and many kinds of data sources, it's a complicated process" (Interview 1. New York Tobacco Control Program, August 2010).

"…certain levels are hard to collect data from. You've got individual, community, media, statewide policy interventions, but it is hard to create one data system that captures all of that well. We did the Adult Tobacco Survey, which is one way to get at multiple levels, but you don't always get the amount of data and level of detail that you need. So you often have to collect data from the targets of the intervention, that happen at multiple levels, which often creates the need for multiple data collection systems. In that way it's complicated" (Interview 2. New York Tobacco Control Program, August 2010).

Contextual Elements Identified by Project Participants as Influencing NYTCP's Evaluation

Project representatives were asked to describe contextual elements influencing the evaluation for the NYTCP. In addition to an opportunity for the project to describe other contextual elements influencing the evaluation, questions were asked about specific contextual elements. Multiple contextual elements were described during interviews as having influenced the evaluation for this project, both in the development of the evaluation and the comprehensiveness.

The following contextual elements were asked about with interview participants and are described in the following section: Sources of Information Guiding Development of the Evaluation, Resources, Evaluation Requirement by Funder, Length of Time Carrying Out

Intervention/ Evaluation, Intended Use of Evaluation, Evaluator Access and Capacity, Evaluation Oversight and Consequence, and Other Contextual Factors.

Sources of Information Guiding Development of the Evaluation

Guidance for developing NYTCP's evaluation were described by project participants as coming from a variety of sources including a review of existing documents and literature review and use of institutional knowledge.

Publicly available documents, such as those from the CDC, were used when developing the comprehensive tobacco control program and provided guidance for the evaluation. CDC documents used included: Best Practices for Comprehensive Tobacco Control Programs, Community Guide, Tobacco Interventions, Key Outcomes Indicator book, Introduction to Program Evaluation for Comprehensive Tobacco Control Programs, and the Framework for Evaluation in Public Health (Interview 1. New York Tobacco Control Program, August 2010; Interview 2. New York Tobacco Control Program, August 2010; New York Tobacco Control Program Representative, 2010a). While many of these documents focus more on the intervention, and only briefly mention evaluation, one document in particular was noted as guiding the evaluation of the program—CDC's Framework for Evaluation in Public Health.

"We used our own judgment of mapping the evaluation on to the intervention, plus used the CDC documents as resources"(Interview 1. New York Tobacco Control Program, August 2010). Additional sources of information identified in interviews, while developing the evaluation, included a systematic review of literature and existing intuitional knowledge, as evidenced by the following statements:

> *"Because we specialize in evaluating TC programs, and we have for so long, there's a lot of institutional knowledge that we have. So we may rely less on some of the standard things that others might use. We'd like to think we are creating some of the resources for evaluation and not just drawing on them" (Interview 2. New York Tobacco Control Program, August 2010).*

> *"I have always thought in such a multi-level way that I guess I didn't think about it in any other way. By the time I started this project we had proposed a model using Social-Ecological Model as the basis for the evaluation" (Interview 2. New York Tobacco Control Program, August 2010).*

> *"It was logical to assume we needed a multi-component evaluation to look at each of the components and then be able to sort of integrate, synthesize and do all those fancy things to come up with an overall evaluation of the comprehensive program" (Interview 1. New York Tobacco Control Program, August 2010).*

In addition to relying on the expertise of their external evaluator, RTI International, the NYTCP also sought guidance from work being done by other states (California, Montana, Massachusetts) and beyond (Australia) (Interview 1. New York Tobacco Control

Program, August 2010). The NYTCP also has an advisory board, which provides feedback and input about the evaluation on an annual basis (Interview 2. New York Tobacco Control Program, August 2010). This project, however, has gained little from other departments within their own organization or other fields when developing the project's evaluation. "Our team is multi-disciplinary, so that brings different perspectives, but it's not as if we drew a lot from HIV/AIDS interventions or other health domains" (Interview 2. New York Tobacco Control Program, August 2010).

Resources

The program's budget was described during interviews and through other documents reviewed as having a large influence on the evaluation. While the NYTCP clearly considers evaluation a critical part of the project, changes in the budget in recent years have greatly influenced the evaluation of the project. "Changes in the budget is probably the biggest one. When the budget was growing we added on to the evaluation, when the budget contracted we subtracted from it. It was pretty directly tied to that" (Interview 1. New York Tobacco Control Program, August 2010).

The NYTCP has needed to become more strategic as recent budget cuts have dramatically reduced the project's budget for both intervention and evaluation activities. From fiscal year 2008-2009 to fiscal year 2009-2010, the NYTCP's budget went from approximately $84 million annually to $64 million annually (RTI International, 2009a). The evaluation budget was also cut to maintain approximately the same percentage of contribution from the total budget. Some of the strategies used by the project are highlighted below.

> *"We go through the strategic plan and evaluation scope of work and deliverables. We make strategic decisions about how much can we spend on each of these projects and what we can cut from the package to still get enough information to do a comprehensive evaluation of this big program. Without compromising our overall evaluation" (Interview 1. New York Tobacco Control Program, August 2010).*

> *"We still need to report to the legislature every year about how the program is doing. We don't know if this protected funds or not, but made an argument that reporting requirements are still in place when budget cuts were discussed" (Interview 1. New York Tobacco Control Program, August 2010).*

A participants described steps were taken to strategically use existing resources, while developing a new and comprehensive evaluation for the NYTCP in 2003.

> *"We tried to rely on existing data sources to the extent possible, rather than creating new systems, but given how long ago we did this initially and given that tobacco control is often at the forefront, we did have to create a lot of new systems. In terms of maximizing resources, we tried to rely on existing data and using that in creative ways to inform the evaluation. For example, BRFSS is used to help inform the program and that was already there and*

measured some of the things we wanted" (Interview 2. New York Tobacco Control Program, August 2010).

Evaluation Requirement by Funder

       Language about the evaluation requirement through the HCRA is brief and gives "a clear statement that they wanted a comprehensive evaluation, but no indication of how to go about doing that" (Interview 1. New York Tobacco Control Program, August 2010). Staff from the NYTCP and external evaluation contractors from RTI International worked together to develop an evaluation to meet the requirements of the legislative mandate, as well as, the needs of the program (Interview 2. New York Tobacco Control Program, August 2010). The independent evaluation contractor assessed the existing data and surveillance system and worked collaboratively with the program to identify priorities, where resources would be spent, and get ideas on the evaluation design (Interview 2. New York Tobacco Control Program, August 2010). The goal was to develop a comprehensive evaluation of the project, while using existing data and systems as much as possible.

       The current evaluation has been modified slightly with the new ARRA funding coming into the program. CDC and ARRA funding are not a large amount, compared to HCRA, but they do have their own requirements and reporting systems. Examples of how the evaluation was modified to account for ARRA requirements include adding in different follow-up timing for data collection and reporting or adding questions into the Adult Tobacco Survey to gather newly needed information (Interview 1. New York Tobacco Control Program, August 2010).

Length of Time Carrying Out Intervention/Evaluation

       This factor was not described by project representatives as being very influential on the project's evaluation. Compared to other programs, such as the CTCP, the NYTCP is a much younger program. They began with an external evaluation contractor and have kept that same system for evaluation in place. Since they are a newer program, they have been able to draw from resources developed and lessons learned by other programs, states and the field. They have, however, been able to establish themselves as a leader in the field, both for their intervention and evaluation.

Intended Use of Evaluation

       The intended use of the evaluation was stated as quite influential in the design of the evaluation for this project. "I think the intended use drove everything. I'm hoping so at least. It drove most everything" (Interview 1. New York Tobacco Control Program, August 2010). Evaluation elements were selected to answer key questions and the program developed a clear workplan, set of goals, and designed their evaluation elements around measuring progress for outcomes (Interview 1. New York Tobacco Control Program, August 2010). Another way the intended use of the evaluation influenced the program relates to communicating with others about changes being made by the NYTCP.

"It's not just about saying that this is effective or not effective. It's translating into numbers of people who are affected. That really drives home the point that it's making public health impact to the state. So that affects our approach and the way we summarize information" (Interview 2. New York Tobacco Control Program, August 2010).

Evaluator Access and Capacity

Project representatives described a lack of capacity for high quality evaluation in the field generally, but the NYTCP has had access to high quality evaluators for their specific project. "There is a lack of capacity to do high quality evaluation" (Interview 2. New York Tobacco Control Program, August 2010). New York and California Tobacco Control Programs have both been described, by themselves and others, as having a sophisticated understanding of evaluation. "It does exist, I just don't think that it's that common" (Interview 2. New York Tobacco Control Program, August 2010).

The distinction between internal and external evaluators may influence both access and capacity for high quality evaluation. There are certainly examples, both within and external to programs, to do good evaluation, "but generally it's more out than in" (Interview 2. New York Tobacco Control Program, August 2010).

A project staff member indicated having "qualified evaluators always available" based on their contract with RTI International, but also acknowledged that it was a "treat to have the capacity and the money to have a contract like this. It's been great" (Interview 1. New York Tobacco Control Program, August 2010). The competitive procurement process undertaken was considered to have worked well in linking the program with an evaluation contractor that met their needs (Interview 1. New York Tobacco Control Program, August 2010).

Evaluation Oversight and Consequence

A difference exists in the amount of oversight by funding source. The main source of funding for NYTCP's intervention and evaluation activities is HCRA. An annual report is submitted to the legislature and was described as satisfying the reporting requirement (Interview 1. New York Tobacco Control Program, August 2010). There does not appear to be additional oversight provided or required by this source. Previous feedback has informed the evaluation, but there has not been any consequence of doing one thing or another (Interview 2. New York Tobacco Control Program, August 2010). The NYTCP also has an advisory board, which provides feedback related to the project's evaluation, more than providing oversight or consequence.

A small amount of funding from ARRA, however comes with more oversight and caution. "Because it is stimulus money, there is sensitivity around it being spent wisely. Follow-up on evaluation is very important. Everyone is taking that very, very seriously" (Interview 1. New York Tobacco Control Program, August 2010).

Directors of the NYTCP, Evaluation branch and the Division would likely be those held accountable if evaluation activities were not completed (Interview 1. New York Tobacco Control Program, August 2010). Having an external evaluation contractor may provide another party to hold accountable, but responsibility would ultimately fall to the

project to adhere to statutory requirements (Interview 2. New York Tobacco Control Program, August 2010). This has not been an issue for the project.

Other Contextual Factors

Recent contextual factors described by project representatives as having influence on the program include an increase in the statewide tobacco tax in 2008 and reduced budget for the program (from approximately $84 million in 2008 to $68 million in 2009) (RTI International, 2009a).

Political forces and state-level changes have impacted the evaluation beyond reducing the budget and passing a recent tobacco tax increase. At the whim of policy makers, changes can be made at a state level and new evaluation components need to be added by the project.

> *"Things happen at the state-level that we don't always plan on. Sometimes for the good, sometimes not. The state is in such a fiscal mess, it decided to raise tobacco taxes. They didn't ask us if we wanted to raise it, though we had certainly been promoting it for a long time. But they just decided to do it. So suddenly we have to measure changes, prices paid, tax evasion. Things happen politically at the state level, laws are passed, policies pushed" (Interview 1. New York Tobacco Control Program, August 2010).*

An example provided was the Attorney General becoming interested in children's use of flavored cigarettes. The NYTCP ended up adding questions to Youth Tobacco Survey (Interview 1. New York Tobacco Control Program, August 2010).

Overall, the political climate is currently considered a friendly one for doing tobacco control. Previously there was more sensitivity around the topic, but this has greatly improved. For further detail about the political sensitivity and the related experience of the project, see the upcoming section titled "Challenges Encountered When Evaluating the Project.""We do have political support for what we're doing, both internal and external…It's a pretty good state for tobacco control" (Interview 1. New York Tobacco Control Program, August 2010).

Summary for Contextual Elements

Project representatives were asked to describe contextual elements influencing the evaluation of the NYTCP. Multiple contextual elements were identified through interviews as having influence on the development and comprehensiveness of the evaluation for this project. Key sources of information guiding the development of the evaluation included publicly available documents, such as those developed by the CDC, work by other tobacco control programs, and knowledge from their external evaluators. Changes in the project's budget, intended use of the evaluation, having the budget and access to high quality evaluators, and politics were all described by interview participants as influencing the NYTCP's evaluation. The project's length of time carrying out the intervention and evaluation, oversight and potential consequences, and information from other departments or fields, beyond tobacco control, do not seem to have influenced the evaluation to the same

extent. Evaluation was required by the funder, but what the evaluation needed to include was basically up to the NYTCP and its evaluators to determine.

Project Strengths in the Evaluation

The NYTCP was asked to describe some of the project's strengths and successes, specific to the evaluation. "I think our flexibility, being able to change on the fly and our media piece are probably what we do the best" (Interview 1. New York Tobacco Control Program, August 2010). Also, "we have tried to use all of the information that we have collected at multiple levels. So that's one sort of generic thing that we've done well" (Interview 2. New York Tobacco Control Program, August 2010). Additionally, the project participants described actively using the evaluation to better the program.

Having the flexibility to make changes along the way, as new issues arise in the state or needs change has been very beneficial to this project. The Adult Tobacco Survey can be changed quarterly to add new questions or change them as needed. "One of the best things is that it's been a very flexible process" (Interview 1. New York Tobacco Control Program, August 2010).

"Our media evaluation piece is comprehensive." Data are collected through the Adult Tobacco Survey, to get the perspective of the general public, and online panels, to get feedback about specific examples of media. The media evaluation was described during an interview as one of the more "creative" pieces to the project's evaluation (Interview 1. New York Tobacco Control Program, August 2010).

> *"I think I'm most happy with the nuanced and helpful data that we've collected on their media campaign, which has translated into a shift in strategy over time in how they approach media. I think that's one area they have felt that the evaluation has been most helpful" (Interview 2. New York Tobacco Control Program, August 2010).*

The active role of evaluation, which has also helped to improve the program, was described as something unique to the NYTCP.

> *"…the program takes evaluation seriously, they invest in it and they use it and it makes the program better. And they haven't been afraid, for the most part, to receive not always positive data and address it and make changes…A lot of states have a love/hate relationship with evaluation. They love it when the news is good and they hate it when the news is bad. It's really tempting for a program to bury bad news, because they fear for their program. It takes a certain amount of courage to say "well, that's part of the process that we are going to fail or we are going to have problems that need to be addressed, but if we address them, we'll get better." Whereas, others try to pretend that the data aren't there and eventually that comes back to bite you. New York hasn't taken that approach. I think they deserve a lot of credit for doing that. I think that's something that's important to acknowledge" (Interview 2. New York Tobacco Control Program, August 2010).*

Recently an example of social norm change was witnessed by a program representative, as described below.

> *"A few weeks ago, I said, probably the single best measure of social norm change happening in New York state, is the fact that these legislators, without hesitation, without discussion, they raised the price on cigarettes. They didn't need to think about, they didn't need to worry about it, they just did it. And that's a big change. In the past they would hem and haw. They would come to us. Should we or shouldn't we do this. Now they just do it. Of course, under pressure to raise funds for the state. The social norm is that you can raise taxes on cigarettes and it really won't bother anyone. To me, that was a huge change in social norms, which is what we talk about changing all the time"* (Interview 1. New York Tobacco Control Program, August 2010).

NYTCP Providing Guidance to Others

NYTCP is also now being looked to for their expertise in developing multi-level interventions and evaluations by other programs, departments and fields.

Lessons learned, how to summarize evaluation data, and how to approach the program's intervention and evaluation have been shared with tobacco control programs in several states, including Indiana, Vermont, Florida, and Ohio (Interview 2. New York Tobacco Control Program, August 2010). The CDC also refers other programs to the NYTCP for guidance:

> *"The standard line at CDC is to "Go talk to the people in New York and see what they're doing. CDC really doesn't lead, in the sense that they aren't at the cutting edge. They consider New York's evaluation to be at the cutting edge. They are literally always sending people here to us, to see what we're doing, both programmatically and for the evaluation"* (Interview 1. New York Tobacco Control Program, August 2010).

Within an evaluation workgroup that the project participates, half of the people in there are in obesity, so there is a lot of information shared with them. Tobacco Control within the organization is considered to "absolutely" lead the way with innovation and evaluation (Interview 2. New York Tobacco Control Program, August 2010). "We'd like to think we are creating some of the resources for evaluation and not just drawing on them" (Interview 2. New York Tobacco Control Program, August 2010).

> *"…in fact we've led the way on a lot of multi-level evaluations and now other programs are coming to us to look for advice, expertise and ideas. I'm on a chronic disease evaluation workgroup internally and it's a one-way street. We give a lot of stuff and that's the way it is. It's not a bad thing, but that's just the way it is"* (Interview 1. New York Tobacco Control Program, August 2010).

Challenges Encountered When Evaluating the Project

While the NYTCP has had many successes related to its program and evaluation, several challenges were also identified. Some of the biggest challenges related to the evaluation were described by participants as dealing with bureaucratic sensitivity and slowness, the impacts from recent budget cuts, the complexity of the evaluation, and evaluation reporting.

A great challenge to the NYTCP has been dealing with bureaucratic sensitivity and slowness. This was more of an issue in the beginning of the program, but was described as influencing the evaluation and leading to strategic action to get around this barrier and get data collected as needed. "A report would take months to get approved and there was concern about what we were saying in reports" (Interview 1. New York Tobacco Control Program, August 2010).

> *"Early on tobacco control in New York was considered to be very sensitive. Anytime we wanted to do something, like put out a new survey, there was reluctance. It took a long time to get approval for things…A lot of that has gone away now. Tobacco is less sensitive, everyone's used to it. So we can get most of our projects and surveys out the door" (Interview 1. New York Tobacco Control Program, August 2010).*

When the Clean Indoor Air Act passed in 2003, there was only a short time to plan the evaluation prior to the law taking effect. There was a lot of bureaucratic pushback slowing the process down. The Project Director at the time ended up working through some local groups and had them take the lead on some of the collection of baseline data. The project designed the surveys and analyzed the data, but data collection was done through these other groups (Interview 1. New York Tobacco Control Program, August 2010).

Several challenges that seem linked to the budget were identified by program participants, including: personnel reductions, a personnel hiring freeze within the state, and issues around contracting (Interview 1. New York Tobacco Control Program, August 2010).

The evaluation plan for the project acknowledged the complexity of contextual factors and synergy across activities was described through published documents as influencing the evaluation of the NYTCP (Austin et al., 2003). These challenges still exist. As described in the First Annual Report of the NYTCP by RTI International (2004), while literature supports the effectiveness of specific components included in the intervention, "inherent in a comprehensive design is an expectation that there are synergies that exist across program interventions. As a result, it is difficult to precisely and reliably measure the effectiveness of individual interventions" (p. 2-9).

Additional challenges, identified by project representatives specific to the evaluation of this project include: collecting relevant data, analyzing it in an appropriate way, tailoring findings and recommendations to various audiences, and dealing with the complexity of the evaluation (Interview 2. New York Tobacco Control Program, August 2010).

> *"The more complex an evaluation gets, obviously the more complex the data coming out the other end are, so different programs have different needs for how the information is summarized and that's evolved a lot over time…The program may get one level of detail, so*

they can really understand what you mean and how they can make course corrections and changes to the program. But then folks higher up in the Department of Health, who are less involved in day-to-day program management don't want as much information, but they want more than the politicians or legislators who may only read an Executive Summary. That's been a struggle. It's not as if you find that out once, like in New York State, and then you go to Indiana and do the same thing. Every state is different. That has to do with personality, level of sophistication, etc." (Interview 2. New York Tobacco Control Program, August 2010).

What the Project Would Have Done Differently

The NYTCP reflected on things they would have done differently, specific to the evaluation. Some of the areas identified included: improvement to annual evaluation reporting and dissemination of information, not having done an expensive survey that didn't provide information expected and collecting data sooner from policy decision-makers.

Concise reporting is something the project has been working to improve. The project has made changes to how annual evaluation reports are written, both in terms of length and focus. There is now a better understanding about who the audience is and the role of the annual report.

> *"I know we wouldn't write those giant reports…Those were way over the top and misguided. I guess we thought an independent evaluation of a comprehensive program has to be a couple hundred pages long. Now we're writing a thirty page report and it is welcome just as much, probably more so" (Interview 1. New York Tobacco Control Program, August 2010).*

The NYTCP is also working to improve dissemination of the large amount of information they have. "We've done some dissemination, but really not enough. We are trying to improve that this year" (Interview 1. New York Tobacco Control Program, August 2010).

The project was involved in a survey early on, that project staff wish had not been included in the first place. A news media tracking survey ended up being a large task and very costly to the NYTCP. "We didn't get a lot from it, so we were able to abandon it. That's part of the flexibility of this project" (Interview 1. New York Tobacco Control Program, August 2010).

The project is now working to collect data from policy advocacy targets, which they would have liked to have started earlier.

> *"With hindsight I would have pushed harder and done more with collecting data from the targets of policy advocacy—the local decision-makers, politicians, etc. that have influence over local tobacco control laws…As tobacco control has shifted to pushing policy change much more than ten years ago, the evaluation has to shift to get data from the targets of those policy advocacy efforts…If you only rely on whether a policy passes or not as your measure of evaluation, you don't really know if you are making interim impact. It is fairly standard to collect data on public support, but at some level, they are less relevant because you need to*

understand whether the local decision-makers that pass ordinances are becoming more supportive of your efforts. If you don't know that, you can continue to advocate for change and may be banging your head up against a wall and not even know it. It's a gap in tobacco control that we've been trying to address and I wish I did that long ago" (Interview 2. New York Tobacco Control Program, August 2010).

<u>Summary for NYTCP</u>

The NYTCP carries out a comprehensive tobacco control program through a multi-level intervention aimed at changing social norms around tobacco. A comprehensive multi-level evaluation was developed by external evaluation contractors to assess the program. An evaluation was mandated through the funding source, HCRA, but the specifics of what the evaluation would include and development of a comprehensive evaluation was up to the program and its evaluators to determine.

Benefits of evaluating the program with a multi-level evaluation were described by participants as learning things that would have otherwise been overlooked, being able to cut things out of the program that are not contributing in the way that had been expected, and better understanding influences that are at play with the program.

Complications of carrying out a multi-level evaluation of a multi-level program were identified by project participants, including the amount of data being collected, organized and reported on, as well as access to that data from the beginning.

Sources of information having the most influence on the development of the evaluation for this program were identified by project participants as: publicly available documents, such as those developed by the CDC, work by other tobacco control programs, and experiential knowledge from the external evaluators.

Other contextual elements identified by project representatives as influencing the comprehensiveness of the evaluation included: changes in the project's budget, intended use of the evaluation, having the budget and access to high quality evaluators, and political forces.

Strengths of the project's evaluation were identified as: its flexibility, the media evaluation, an active role of evaluation in improving the program, witnessing social norm change, adequate funding (until recently), and funding for and access to high quality external evaluators.

Challenges to the NYTCP's evaluation were described interview participants as the dramatic budget reductions over the last few years, bureaucratic sensitivity and slowness, and the complexity of the evaluation, reporting and dissemination.

The program is working to continue improving their reporting and dissemination and collecting data from policy advocacy targets.

It should be noted that all statements about elements influencing the evaluation are based on information provided by project representatives during interviews. Therefore, they are the ones attributing influence and describing the influence on their project and the evaluation.

Case 3 - Network for a Healthy California (Network)

Project Description

The Network for a Health California (Network) Program is funded through the United States Department of Agriculture's (USDA) Food Stamp Program, which is a food assistance and nutrition education program now titled Supplemental Nutrition Assistance Program (SNAP). SNAP is focused on helping low-income people and families buy food they need to support good health (Shenkin & Jacobson, 2010). SNAP-Ed is a program funded through SNAP focused specifically on nutrition education. A recent Institute of Medicine report described factors related to the childhood obesity and the particularly vulnerability of low-income and children of color (Institute of Medicine, 2004). Foerster et al. (2007), conclude that "increasing children's physical activity and improving their access to and consumption of healthy foods clearly are the twin pillars for obesity prevention" (p. S176).

The Network, started in 1997 became an umbrella, such that everything that could go under it was put there, because the (now titled) Department of Public Health was able to get matching federal (USDA) funds (Interview 1. Network for a Healthy California, August 2010). California's SNAP-Ed program is now funded at approximately $325 million (Interview 3. Network for a Healthy California, September 2010). California is the largest recipient of this funding, with about one-third of the nationwide costs of the program (Interview 2. Network for a Healthy California, September 2010). The money for this program comes from the USDA, through the California Department of Public Health, and to the Network. Occasionally additional grants are received to do other work, but it does not come close to the amount of money received from the USDA.

The USDA is in an interesting position because they are charged with protecting the interests of the food industry for the country, not just promoting healthy foods and behaviors. There is also great concern by the USDA that funding only be spent on low-income, food stamp-eligible populations. Work that reaches other populations is not allowable by the USDA. According to Shenkin and Jacobson (2010), while SNAP-Ed is meant to "encourage low-income consumers to choose healthier diets" the program cannot use "marketing campaigns to discourage the consumption of unhealthy foods" (p. e2). Challenges to the program in meeting these restrictions, and implications for the program's evaluation, are discussed at length throughout this case study.

The Network started with a planning grant, but no program activities, from 1997-1999; once program activities began in 1999, so did the collection of process data, though it was not very rigorous in the beginning (Network Program Representative, 2010c). Surveillance surveys in the field, however, have been taking place since around the late 1980's, well before the program began (Interview 1. Network for a Healthy California, August 2010). Much less than 10% of the program's budget goes towards the evaluation (Interview 1. Network for a Healthy California, August 2010). The Network develops and conducts most of its evaluation internally. Guidance during the development of the evaluation, and continued evaluation work has been carried out by project staff with evaluation expertise.

In the beginning, the Network was told to generate partnerships, generate a state budget that could match to federal dollars, and do social marketing. The Network was originally focused on food and fruit and vegetable consumption. This has been expanded to work on improving nutrition and improving physical activity for low income populations in California (Interview 2. Network for a Healthy California, September 2010). Working to address childhood obesity was not an original goal of the program, and team members seem to consider that they work towards it as a long-term goal, through efforts on multiple levels addressing healthy eating and promoting activity. They still consider other areas of work to be their priorities as well.

> *"Obesity prevention was tacked on as one of the goals of the Nutrition Network, but it wasn't an original goal of the program…Network gets lumped in with obesity prevention programs, because within the Department somebody needs to be working on it, so it ends up being us because we do fruit and vegetable consumption and physical activity" (Interview 1. Network for a Healthy California, August 2010).*

Foerster et al. (2007) describe that "Similar to the tobacco control movement, obesity prevention will require changes in exposure to unhealthy influences, marketing practices, and price incentives that modify the commercial conditions in which obesity flourishes" (p. S176). They go on to state that the "building blocks for this sea change are expected to include community programs, mass communications and public education, and environmental and policy approaches" (p. S176).

This program works toward the above goals by intervening on multiple socioecological levels concurrently. Innovative ways to address childhood obesity include intervening on multiple levels concurrently, such as the individual, family, school, community, organizational, and policy levels. The goal is often to use multiple avenues to create a change in social norms, expectations and environments around nutrition and physical activity for children, their families and communities (Chapter 2 provides further detail about socioecological levels). The Network intervenes on the following socioecological levels (with activities at each level presented in parentheses): *individual* (education, self-efficacy, behavior); *community* (access, retail programs); *schools*; and *media* and *partnerships* (cross-cutting between levels) (Network Program Representative, 2010a). Intervention activities are carried out through work by the Network, local projects, regional collaboratives, and contractors throughout the state.

Evaluation for this program also takes place on multiple socioecological levels concurrently. Evaluation for the Network takes place on the following socioecological levels (measuring areas of): *individual* (knowledge/perceptions, self-efficacy, behavior for food/activity); *community* (access); and *media* and *partnerships* (overarching between levels) (Network Program Representative, 2010a). The Network funds local projects, which have to fulfill reporting requirements. While there are not guidelines that projects have to follow for the amount of budget allocated for evaluation, all projects conducting impact evaluations have some amount of budget set aside for their evaluation and reporting activities (Network Program Representative, 2010c).

The Network has published about their use of the Social-Ecological Model as a framework for evaluating their program. Further information about this model can be found in a publication by Gregson et al. (2001).

The program's impact evaluation for 2008 reported on state and local data measuring immediate impact (changes in fruit/ vegetable consumption, physical activity and related factors) and described long-term outcomes (reduce obesity, body mass index level, related morbidity and mortality and improve health outcomes) (Fourney et al., 2009). This report indicated that many of the program's contractors found significant changes in fruit and vegetable consumption (Fourney et al., 2009). Contextual challenges were also identified, including high staff turnover at the state and local level, inconsistent use of valid and reliable tools, and diversity of activities making it difficult to attribute changes to specific strategies (Fourney et al., 2009).

Most recently, the Network has developed a journal supplement with many different articles addressing their work and recent accomplishments. Quantitative changes to fruit and vegetable consumption is often favored by researchers and others outside of the program, as a way to describe the success of the program, rather than other work Network staff promote around their partnerships and media activities (Interview 1. Network for a Healthy California, August 2010).

Benefits and Complications of Evaluating on Multiple Levels

Network representatives were asked to describe how this project benefited by evaluating all levels of the intervention. These key informants were also asked how the evaluation was made more complicated by evaluating all levels of the intervention.

Benefits of evaluating the program with a multi-level evaluation were described by participants as using the scope of the evaluation to better understand the program's impact and using this information to inform future work.

> *"The project works in many channels and many levels of the socioecological model. So, a broad evaluation scope enables the Network to demonstrate positive change in a variety of settings and with various audiences and it allows for triangulation of data" (Interview 2. Network for a Healthy California, September 2010).*

The program is also using the evaluation to figure out which parts are yielding the best results and trying to use that information to inform future work and funding for regional projects.

> *"We've got all these initiatives, which ones are yielding the most results in the community? Looking at policies and environmental outcomes shows me what initiatives are exciting to regions and the change being made. It shows me that farmers markets are able to make changes, but other initiatives haven't made much impact and have not worked. For the future, this will be informing how we move forward" (Interview 4. Network for a Healthy California, September 2010).*

Complications of carrying out a multi-level evaluation of a multi-level project were identified, including: putting the work of the project into a useable framework, the diversity of funded projects, linking data to geographic areas to show change, and assessing synergy, attribution and effectiveness of the program.

The scope and complexity of the program alone offers its own challenges to the evaluation of this program. "It is difficult to bring evaluation information together into a coherent whole" (Interview 2. Network for a Healthy California, September 2010).

"I have come to realize that the bigger something is, the more difficult it is to characterize what is going on" (Interview 3. Network for a Healthy California, September 2010). The Network has spent a great deal of time applying their work to already developed frameworks, looking at different logic models, and trying to get buy-in from different groups about what the program is doing, what it wants to accomplish, and specific terminology for models (Interview 1. Network for a Healthy California, August 2010). Researchers worked to be satisfied with labels for different levels in the model, but project staff did not know what to make of it (Interview 1. Network for a Healthy California, August 2010). This was a challenge. "If it's puzzling to the staff it's just going to be a huge black box to local contractors" (Interview 1. Network for a Healthy California, August 2010). This is an ongoing challenge to the Network, but the program feels that having a working framework is important to the evaluation of a multi-level project.

> *"In meetings people debated correct labels of levels in the model…the point is that you have multiple levels of influence on an individual. Wordsmithing does come up and it's hard to get agreement…but people do think it's important" (Interview 1. Network for a Healthy California, August 2010).*

The diversity of funded projects, both in doing different work and addressing different levels, has added complexity to the Network's evaluation. "Trying to pull generalizations from really diverse projects and the different levels that they influence is kind of challenging" (Interview 1. Network for a Healthy California, August 2010).

Precision with measurement and metrics around media work and fruit and vegetable consumption have been more complicated than expected. Finding a common measurement denominator for media work done has been an ongoing challenge.

> *"We actually asked our media contractors to convert gross rating points into impressions, and to estimate the amount of USDA food stamp eligible audience they think they are reaching. But media contractors don't have a lot of expertise in food stamp regulations and have to translate their information into what the USDA wants to hear. It's the language that you use. Part of the complication is getting the right information from our contractors…So, getting a lot of different ways of presenting the same information from a contractor that is used to presenting it in one field. They are experts in media, so, we have to take their expertise and translate it…When trying to compare media impressions and classroom impressions, it really doesn't compare, but we need to reconcile these. We need to*

find a common denominator to compare. It's never perfect. It's very messy. That's one of the ways things get really complicated....we have to translate back and forth between these fields" (Interview 1. Network for a Healthy California, August 2010).

"Instrumentation and metrics is another area...We've been trying to put metrics against a national action plan that was developed. Those metrics have been hard to come by because surveillance systems aren't in place. If looking at how well schools are doing, different surveys are done every three to five years...The rate of changes also isn't that much, especially if you're not putting that much money into it, you aren't going to get much change. We believe that dietary change is slow, so you need instrumentation that can detect small changes and you need to break it down by different population segments, because they behave differently" (Interview 3. Network for a Healthy California, September 2010).

Additionally, while it is appealing to show impact through mapping, data that are available and geographic mapping do not necessarily align in a way for this to be appropriate.

"Sometimes people want to do a geographic aggregation...such as program activities, media work and surveillance data going on for a project and they want to put it all together. Theoretically you could probably do that with some sort of hierarchical modeling, but program exposures don't always fit into precise geographic units…You can't get the precise measures that you might actually need to know how much exposure is going on in a county and how that influences individuals in that county, the program data are almost too messy to really be able to take that on. It's just inherent in the way the data are measured or how the data are. Precision with measurement is an issue" (Interview 1. Network for a Healthy California, August 2010).

Synergy between elements of the intervention was described by one staff member as where greater impact is achieved, but measuring it, or even knowing how to do so, has been elusive. Another area of complication, when evaluating this multi-level intervention, which was identified by several project staff members, is attributing change directly to the program.

"There hasn't been work that I'm aware of, and we've looked hard for it, around the concept of synergy among elements in intervention. Synergy is the strong integration, repetition and consistency among the elements of your intervention and that's where you begin to get greater payoff. Each area of intervention may have a small or even undetectable positive benefit, but it is the combination or synergy among different items that ends up resulting in population change. This is what the tobacco control program in California discovered. They had benefit of a lot of community research that we didn't have, but they still had to apply it on a statewide scale…Our funding scheme affected or distorted what we might have done otherwise. They gave locals money to try different things that they thought might be effective and watched what happened with evaluation. They contracted with outside evaluators. Some of that worked, a lot of it didn't work. They took what seemed to be working and did more of it" (Interview 3. Network for a Healthy California, September 2010).

"Any kind of direct attribution is completely out of the question. It is not feasible and doesn't make scientific sense because the Network is so big…It's a huge project, it's statewide. We can't get another state to act as a control. Plus, there are other nutrition programs going on throughout the state…We have ideas about how to do multi-level analysis, but we haven't done intentional comprehensive analysis of the entire network. We sought feedback from the Robert Wood Johnson Foundation, and they were stumped too. It's a question we still can't answer" (Interview 1. Network for a Healthy California, August 2010).

"It is hard to do attribution because these things work together. In fact, the design the IOM had really is very much a model where all the stakeholders are supposed to work together, so that means that only the combination will get you to the ultimate outcome…you can't really attribute the change to just yourself, you have to be willing to share credit, and that always gets dicey because funders want to be able to say it was they who caused the result. It's another level of funding group being willing to work together and share credit and still feel like their boards of directors are going to be happy with that" (Interview 3. Network for a Healthy California, September 2010).

The Network is asked a lot about their effectiveness, but they have not been able to provide an answer until recently, because they did not have a specific salient data point, to point to for change (Interview 1. Network for a Healthy California, August 2010). This is something that has changed recently, and they are publishing about in an upcoming journal issue. "The questions we have trouble answering are those we get the most. We talk about it a lot, but almost so that we can be confident that we really can't answer them" (Interview 1. Network for a Healthy California, August 2010).

Contextual Elements Identified by Project Participants as Influencing Network's Evaluation

Project representatives were asked to describe contextual elements influencing the evaluation for the Network. In addition to an opportunity for the project to describe other contextual elements influencing the evaluation, questions were asked about specific contextual elements. Multiple contextual elements were described through interviews as having influenced the evaluation for this project, both in the development of the evaluation and the comprehensiveness.

The following contextual elements were asked about with interview participants and are described in the following section: Sources of Information Guiding Development of the Evaluation, Resources, Evaluation Requirement by Funder, Length of Time Carrying Out Intervention/ Evaluation, Intended Use of Evaluation, Evaluator Access and Capacity, Evaluation Oversight and Consequence, and Other Contextual Factors.

Sources of Information Guiding Development of the Evaluation

Guidance for developing the Network's evaluation was described by participants as coming from a variety of sources including: published literature, theory, internal discussion,

funder requirements, another department within the program's organization, and feedback from external experts.

The Network developed their evaluation by doing what they "thought was best and most useful for the program" (Interview 1. Network for a Healthy California, August 2010). Since evaluation is developed and conducted internally, this was extremely influential in the development of the program's evaluation.

> *'The main criteria for what we thought was necessary to include was whatever the funder required, which before wasn't much, and then what else we thought was timely and useful. It was our own professional opinion. It was theory-based, but it was at our discretion" (Interview 1. Network for a Healthy California, August 2010).*

The program has consistently drawn from published literature and theory. This has been combined with discussions with other colleagues about multi-level models and theoretical models that are useful when working at multiple levels. In particular, the Network used a social-ecological model to develop an overarching framework to describe and guide the program's evaluation (Gregson et al., 2001).

> *"The social-ecological approach, says that in trying to impact individual behavior change, you need to surround a person, work within their own context, peers, communities, social norms and environment, which includes policy and environmental change" (Interview 3. Network for a Healthy California, September 2010).*

An overarching framework for the evaluation has been important for the program and a source of great discussion. Multiple different frameworks have been developed by using different underlying theories, models and influences. "A broad evaluation scope allows the Network to demonstrate positive change in a variety of settings and allows for triangulation of data" (Interview 2. Network for a Healthy California, September 2010).

> *"Initially it was based on the social-ecological model, behavior change theory and social marketing literature, but most recently it's been based on the Institute of Medicine's framework for childhood obesity. That is the framework for how we are looking at evaluation" (Interview 2. Network for a Healthy California, September 2010).*

Early on, this project used evaluation literature the USDA had commissioned on nutrition education, which recommended using social marketing and networks to engage multiple stakeholders to work in a coordinated effort (Interview 3. Network for a Healthy California, September 2010).

While initially there were few requirements for the evaluation by the USDA, as evaluation requirements have increased, this has guided some elements of the evaluation. The Network's evaluation was developed originally, and continues to be, more comprehensive than necessary, based on the limited requirements of the funder. In recent

years the evaluation requirements by the USDA have increased and new elements to the Network's evaluation have been added to meet these mandates.

A Network interview participant described receiving little feedback from the USDA about whether their reports, which they submit annually, have been "good, bad or indifferent" and feels that information provided to the USDA in these reports are not being used (Interview 3. Network for a Healthy California, September 2010). Feedback from the USDA, other than to meet requirements have not greatly influenced the evaluation of the program. More often, information is reported and used internally. This process has influenced the work being done by the program and the direction to move for future funding.

As resources and capacity grew, the Network was able to incorporate evaluation into more areas of the program.

> *"We very early put in place state surveys, starting with adults, then when we had the resources, teens and children. We could track on a biennial basis reported changes in diet, physical activity, knowledge, awareness, and beliefs…As resources grew we were able to establish a research and evaluation unit. Each member of the evaluation team was paired with one of the units within our organization, to be their evaluation partner. If they were developing anything from scratch, such as a toolkit, or doing a study, they had an evaluation specialist that could help them and provide leadership. We developed a fairly rich team of evaluators who do both qualitative and quantitative activities" (Interview 3. Network for a Healthy California, September 2010).*

Various sources of information were used to guide different elements of this multi-pronged evaluation. These are presented individually below.

For the impact outcome evaluation, the USDA had some documents that were useful and the program did an extensive literature search for valid instruments (Interview 2. Network for a Healthy California, September 2010).

The process evaluation was not added until 2000 and used ideas from social marketing research and tools, along with other components added by the program (Interview 2. Network for a Healthy California, September 2010).

Surveillance questions and related methodology are based on national surveillance surveys, such as the BRFSS (Interview 2. Network for a Healthy California, September 2010). Dietary data came from modifying other sources, especially for wording in the teen and children's surveys (Interview 2. Network for a Healthy California, September 2010).

Participatory research literature helped guide the training for working with contractors on how to conduct evaluation (Interview 2. Network for a Healthy California, September 2010).

For the media element, the program used standard media measurements, including impressions and standard ad recall, a well a literature search (Interview 2. Network for a Healthy California, September 2010).

For the Communities of Excellence for Nutrition, Physical Activity and Obesity (CX3), an extensive process was used to develop indicators of community health, including

multiple roundtables internally and externally with experts to develop a list of potential indicators and a grading process with others with various expertise (Interview 2. Network for a Healthy California, September 2010). Initially the Network had a long list of community indicators in the CX3, that promoted fruit and vegetable consumption, physical activity, and related factors, but only about one-third of them could be done under USDA funding (Interview 1. Network for a Healthy California, August 2010).

The Network's case study evaluations were adapted from the California Tobacco Control Program's guidelines for case studies. "It really lent itself very well" (Interview 4. Network for a Healthy California, September 2010).

Additionally, looking at policies and environmental outcomes, even to a limited extent, informs the program about initiatives that are exciting to regions and changes being made (Interview 4. Network for a Healthy California, September 2010). This information is being used by the Network to help guide future funding.

Theory has been used both in the development of interventions and the evaluation. "We do intentionally link all interventions to a behavior theory. Even the surveillance questions have a clear link to some sort of behavior theory" (Interview 1. Network for a Healthy California, August 2010). The program wanted to use evidence-based interventions, but there were not many of a large-scale available when the Network started, mostly only small-scale interventions in schools or community settings were available (Interview 3. Network for a Healthy California, September 2010). The Network used literature review and formative evaluation, to develop, test, and revise their own interventions (Interview 3. Network for a Healthy California, September 2010).

Nutrition, physical activity and food security were described as not having the same research base and being underfunded, compared to other health areas, such as cancer prevention and tobacco control (Interview 3. Network for a Healthy California, September 2010). This lack of a research base and already-developed interventions of an appropriate size was described by participants as influencing the program and its evaluation.

> *"We anticipated using proven-effective interventions from the literature, then would try to adapt to community settings, but we mostly ended up taking what they were already doing, what they felt was working, and tried to evaluate it to see if it really was having impact" (Interview 3. Network for a Healthy California, September 2010).*

The Network was also guided by its proximity to another program within their same organization, the now-titled California Department of Public Health. The Network and its evaluation were structured after the California Tobacco Control Program (CTCP), "such as collecting process data from the projects and surveillance data to detect behavior change. That's all from tobacco control" (Interview 1. Network for a Healthy California, August 2010). The CTCP also required that coalitions be formed at the local level and were charged with actively moving the program's work forward, which the Network took and adapted to nutrition and physical activity (Interview 4. Network for a Healthy California, September 2010).

The links between these two programs was partly due to the Network and CTCP originally being under the same Branch Chief. The guidance of the Branch Chief, and the readily available tobacco control program structure for evaluation, greatly influenced the development of the Network's evaluation. The content of the programs are naturally quite different, which has created some challenge to the applicability of some knowledge transfer.

> *"Where we had a few shortcomings, is that the decision to stop smoking is pretty finite, but decisions about whether to eat fruits and vegetables is something that happens a hundred times per day. Behaviorally it is really a different animal, but as far as program structure it is pretty similar" (Interview 1. Network for a Healthy California, August 2010).*

Feedback from external sources has provided additional guidance and valuable feedback for the Network's evaluation. The Network received a small grant from the Robert Wood Johnson Foundation to look at whether the evaluations being conducted were on-target for the areas in which they were working. Key informant interviews were conducted with evaluation experts to provide feedback to the Network. The evaluation experts felt that the program's surveys were good, but were not sure about the mix and process measures being used (Interview 3. Network for a Healthy California, September 2010). One of the recommendations that came out of the study, to address the diversity and complexity of the program, was that the program should "publish, publish, publish," but do so through commissioning a journal supplement rather than peer-reviewed articles (Interview 3. Network for a Healthy California, September 2010). The Network has successfully developed 17 papers for a supplement in the Journal of Nutrition Education and Behavior, to be published in Spring 2011 (Interview 3. Network for a Healthy California, September 2010).

Resources

One of the most important resources, identified by Network representatives, has been working with partners that are not funded through the USDA. The Network recently counted the number of people and organizations involved in regional collaboratives, including funded and unfunded partners, and found that nearly 1,000 people were involved, nearly half of which were unfunded partners (Interview 4. Network for a Healthy California, September 2010). These groups have had a tremendous impact, not only on the intervention, but the evaluation as well. The Network has had to be strategic, to move forward work it feels is important, but at times cannot be done using USDA funding. Unfunded partners can do work to impact environmental and policy changes, which are long-term changes the Network would like to see in communities (Interview 4. Network for a Healthy California, September 2010). The unfunded partners have been a big part of accomplishing such work. "It's a way to really leverage your resources and energy and make a bigger bang" (Interview 4. Network for a Healthy California, September 2010).

> *"USDA restrictions make it very complex for us to balance the needs of what the community wants to rock and roll on versus what our funding can do. That's why the*

partners are so important. Their funding and not being tied to USDA can make things go further faster and they can tackle some direct policy-related issues and environmental change-related issues, where our funding is more restrictive in that way. It can be a balancing act, but the unfunded partners are so important to moving the movement forward and getting important environmental changes done" (Interview 4. Network for a Healthy California, September 2010).

In some instances, regional work that has generated the most interest can only be done by projects able to find other funding. One such campaign, "soda-free summer," was funded through a foundation, including money for an evaluation (Interview 4. Network for a Healthy California, September 2010). The evaluation showed the effectiveness of this campaign and other regions became interested in implementing something similar, but only regions with other sources of funding could implement this campaign (Interview 4. Network for a Healthy California, September 2010). Based on targeting soda in particular, which violates USDA allowances for funding, other resources are needed for any work on such a campaign. Regions not able to find additional funding, which were most of them, changed the focus to "Rethink Your Drink," which could be funded, through USDA funds. Currently nine of eleven regions in California are working on a "Rethink Your Drink" campaign (Interview 4. Network for a Healthy California, September 2010).

Another valuable resource described by Network team members was access to data. The program tries to maintain active data sets and pull in data from as many sources as possible, especially anything that can generate state-specific estimates (Interview 1. Network for a Healthy California, August 2010).

The program's access to expertise outside of their team was another resource identified by Network representatives. These included working with a survey research group, media consultants, external evaluation consultants, and University professors (Interview 2. Network for a Healthy California, September 2010). The Robert Wood Johnson Foundation provided funding for a study that contributed external evaluation guidance to the program (Interview 2. Network for a Healthy California, September 2010).

Evaluation Requirement by Funder

Initially there were few evaluation requirements by the Network's funder, the USDA. The program was told to generate partnerships, generate a state budget that could match to federal dollars and do social marketing (Interview 1. Network for a Healthy California, August 2010). As described previously, it was up to the program and its staff to determine what would be the best way to evaluate the program. "Evaluation requirements have changed quite a bit…Local programs used to get more money with less reporting required" (Interview 1. Network for a Healthy California, August 2010).

"At the beginning there was almost no requirement. Very, very minimal reporting requirements. For the next batch of years there were not extremely heavy reporting requirements. But beginning this year there are very detailed reporting requirements. Which have required us to initiate extensive reporting requirements from our contractors which has

Every program that was funded by the USDA, including the Network and its various sub campaigns was put into a USDA-developed evaluation grid, to describe whether it was using process evaluation, impact evaluation, or outcome evaluation (Interview 1. Network for a Healthy California, August 2010). "A final report was also required and had certain elements that needed to be in it, but it wasn't a rigorous final report. It was a report that could be put together without detailed data collection" (Interview 2. Network for a Healthy California, September 2010). About 150 projects complete evaluation grids and final reports for the USDA (Interview 1. Network for a Healthy California, August 2010). These evaluation elements were a necessary step, from the perspective of the funder, but not particularly valuable for the program or its funded projects; at the same time, the Network was doing their own evaluation work, to which the USDA did not seem to pay much attention (Interview 1. Network for a Healthy California, August 2010).

Around 2004 the Network became really large and the USDA became concerned with outcome and impact evaluations and added new evaluation requirements to all projects over a certain amount of funding (Interview 1. Network for a Healthy California, August 2010). The USDA put in place a requirement that all evaluation projects funded by the Network for at least $400,000 would need to do a rigorous impact evaluation (plus the USDA grid and final reports). The Network set their own threshold of $350,000 and asked all of their projects, not just evaluation projects, funded for this amount or higher to fulfill the USDA requirements (Network Program Representative, 2010c). In California the $350,000 threshold is met by approximately 48 projects (Interview 1. Network for a Healthy California, August 2010). It has been helpful that this reporting requirement is only for the larger, better-funded projects, as "the smaller contractors really don't seem to have the capacity to do a standardized rigorous evaluation" (Interview 2. Network for a Healthy California, September 2010).

The latest evaluation requirement added by the USDA is the Education and Administration Reporting System (EARS), which was piloted first in 2004, and again in 2007, and required for the first year of data collection this year (Network Program Representative, 2010b). EARS requires reporting of unduplicated counts of participants, and a breakdown by participant: gender, race/ethnicity, and program participation. The Network has worked to tailor data collection and activity tracking forms the projects are already doing, (through elements previously added by the Network) into meeting the requirements for EARS, so that local projects only have to fill out one form that serves two purposes (Interview 1. Network for a Healthy California, August 2010). While the Network has spent a great deal of time trying to help the local projects make this transition, it is challenging and time intensive because Network staff had to understand each unique project enough to develop a form that would capture their information accurately (Interview 1. Network for a Healthy California, August 2010). The new reporting requirements and implications for the Network and their funded projects are described below in further detail.

"Until this year an unduplicated count of participants and the number of contacts, with no further breakdown, was something we needed to provide. There is additional project evaluation information, if we have it available, but there is no real directive as to what that had to be. However, beginning this year, in the report we have to file by mid-November, we are going to need an unduplicated count of participants with associated detailed demographic data, number of contacts, a lot more information about the types of services and the sites for services where intervention was provided, social marketing contacts by each campaign, by the poverty level of people who were reached. Quite a bit more detailed data is going to be required by our funder. At the beginning there was almost no requirement. Very, very minimal reporting requirements. For the next batch of years there were not extremely heavy reporting requirements. But beginning this year there are very detailed reporting requirements. Which have required us to initiate extensive reporting requirements from our contractors, which has been a difficult implementation issue…For example, we could include evaluation reports, which were standardized, if we had evaluations we wanted to discuss, but there wasn't a minimum. Now, if we or our local contractors conduct evaluation projects, we must report on those evaluations with a standardized format. Given that we have about 45 contractors who do evaluation projects, which we trained them to do, this year we will have to turn in 45 individual evaluation reports for those contractors. In the past we prepared only one summary report describing what they have done. It's a huge difference. But this year, knowing this was going to happen, we prepared their report form where the first 15 questions answer the reporting requirements, so we are already set up, in such a way, so that all we have to do is copy their end of year reports, print it and we have their reports" (Interview 2. Network for a Healthy California, September 2010).

The Network added additional evaluation elements, beyond those required by the USDA, these included: Semi-Annual Activity Report (SAAR), evaluation of specific projects before statewide roll-out, and surveillance. The program's intent is to have impressions for tracking progress, specific interventions evaluated before being widely disseminated, and the surveillance to look at the overall impact of the Network (Interview 1. Network for a Healthy California, August 2010).

SAAR tracks media impressions and all media work, classes and promotional events.

Evaluation of specific campaigns and interventions also take place, using an efficacy evaluation, to test all components of the intervention and assess impact (Interview 1. Network for a Healthy California, August 2010). If the program sees changes they will roll out the intervention or campaign statewide.

Through their surveillance system, the program tracks dietary practices, including fruit and vegetable consumption and related determinants, and physical activity for adults, teens and children.

Sometimes other partners collaborate with the Network to help to fund an evaluation. Additionally, contractors may evaluate additional work they are doing, and as long as it is allowable and relevant, the Network is fine with this (Interview 2. Network for a Healthy California, September 2010).

Length of Time Carrying Out Intervention/Evaluation

The Network started with a planning grant, but no program activities in 1997, once program activities began in 1999, so did the collection of process data (Network Program Representative, 2010c). Previously, very limited data were collected for the evaluation of the program. The evaluation of the program was described as being influenced by many things, but the length of time intervening and evaluating the program does not seem to have had as much influence on the comprehensiveness of the Network's evaluation, compared to other elements.

Intended Use of Evaluation

Intended use by the funder, the program, funded projects, policymakers, and others, were identified by project staff as having influenced the Network's evaluation.

Different parts of the program conduct evaluations for different purposes. Surveillance surveys can provide quick sound bites, such as changes in fruit and vegetable consumption (Interview 1. Network for a Healthy California, August 2010). These are intended to be used by policy makers and to demonstrate effectiveness of the program to the funder (Interview 1. Network for a Healthy California, August 2010). The impact evaluation was included to be responsive to the USDA (Interview 2. Network for a Healthy California, September 2010). "USDA did make it clear that they wanted to see outcomes. They didn't have specific guidelines, but it was included for the intended use of demonstrating that the programs were accomplishing something" (Interview 2. Network for a Healthy California, September 2010).

> *"In surveillance surveys some of the questions are included because they relate to current hot topic issues and can be used to inform policy decisions. Although we are not allowed to do systems, environment and policy change work, we are allowed to inform legislators when we are asked" (Interview 2. Network for a Healthy California, September 2010).*

Some data that had intended use originally lost its use, but has not been taken out by the program. "It would be good if we could let go of things when they lost their use" (Interview 1. Network for a Healthy California, August 2010). But, as described below, changing course can be a difficult task.

> *"…we started off collecting SAAR data to capture social marketing, but as the field changed and the social marketing literature became clearer, some things didn't seem as useful anymore. We haven't really moved on from that to things that are more useful. With a big organization changing it is like steering the Queen Mary. Change in how you do business is hard. It's not just you making this decision, its 50 people that also have to agree and change their forms. And their 50 contractors have to change the way they do systems. Even a small change has a huge ripple effect on the amount of time people need to spend changing. It's become really clear that this is a huge issue. Even if the change is to make people do less work" (Interview 1. Network for a Healthy California, August 2010).*

103

Making data more useful may involve asking whether certain data or questions are really needed. "It's hard to let go of the nice-to-know, for only the essential-to-know" (Interview 1. Network for a Healthy California, August 2010).

The Network tries to make information they have and collect not only useful, but also available to their funded projects and the public. This is done through GIS mapping, generating reports, sharing findings and other work (Interview 1. Network for a Healthy California, August 2010).

Evaluator Access and Capacity

Evaluation for the Network is carried out primarily by internal program staff. A combination of a tough learning curve and previous experiences with external evaluators led the program to build evaluation capacity internally.

> *"The Network is so complicated. It takes our own staff, when we hire someone, at least a year to really get a sense of how the Network functions. Hiring an evaluator is not a very good use of money. It's better if you hire an evaluator for a project that you've already scoped out. You have clear goals and can express to the evaluator what the limitations are for the methods. To get a broad brush evaluation of the big picture of the program, we would probably be interested, if we found someone we thought could do it, but between the scope of project and the federal requirements and the vast array of research methods that go into it, we just haven't been able to do that. It's a steep learning curve. It would be a lot of money just to get even a little bit of usefulness out of it" (Interview 1. Network for a Healthy California, August 2010).*

> *"Also a big issue is that an external evaluator may come up with projects that aren't allowable by USDA. We can come up with an idea and scope of a project, then contract out something specific, like focus groups, to use their expertise for that specific piece. Actually, that's a really good way to be more efficient in our work, if you can contract a lot of that stuff out" (Interview 1. Network for a Healthy California, August 2010).*

The Network feels that they have been able to build evaluation capacity internally, though it took some strategy to make this happen. Comments such as "We developed a fairly rich team of evaluators who do both qualitative and quantitative activities" (Interview 3. Network for a Healthy California, September 2010) were echoed by several project respondents. Though they do have capacity internally, at times the program does seek evaluation guidance and feedback from outside sources. "We tend to get a concept that we think is good, then run it by experts. They will help tweak it a little bit" (Interview 1. Network for a Healthy California, August 2010). Additionally, a few external consultants provide some technical assistance to local projects and special projects (Interview 2. Network for a Healthy California, September 2010).

"In our unit we have a pretty good capacity to do it, but there are ten of us, which is big for a public health program. We have experts in nutrition, psychology, sociology, and a statistician. We actively recruited for different kinds of expertise" (Interview 1. Network for a Healthy California, August 2010).

The Network has also actively sought to develop the capacity of projects that they fund. Some of the contractors doing impact evaluations have qualified evaluators (Interview 2. Network for a Healthy California, September 2010). Others have built their own evaluation capacity while working with the Network.

"…these are projects that had no evaluation experience pretty much at all. The main coordinator is usually the food service director, so no background in evaluation, a lay person in the truest sense of the word…These 48 projects had to learn how to do evaluations. [A project staff member] wanted these projects to not only do evaluations, but also believe in it and value it. In first year only 12 projects of 48 were able to do it…but over several years…they were given tools so that they could do their own evaluations. We didn't want the state to go in and do it for them…It's a sustainable system" (Interview 1. Network for a Healthy California, August 2010).

Evaluation Oversight and Consequence

Oversight takes place on two levels for the Network, the first is with their funder, the USDA, and the second involves their oversight of projects they fund.

Oversight from the USDA was described by project staff as being based more on fiscal monitoring than evaluation. Annually the Network creates a plan for approval by the USDA, which includes all work that will be done. Making sure that Food Stamp money is reaching the right people is the primary concern of the USDA, and whether the program is effective or not seems to be secondary (Interview 1. Network for a Healthy California, August 2010). "Fear of fraud really overshadows a lot of the evaluation requirements that we get from USDA" (Interview 1. Network for a Healthy California, August 2010).

Another component of the accountability by the Network involves making sure that projects are working within approved census tracks. The Network developed a system of mapping to determine the census tracks with enough people eligible to be doing work in that area (Interview 1. Network for a Healthy California, August 2010). Finding accurate information and getting local projects to work in those specific areas has been challenging (Interview 1. Network for a Healthy California, August 2010). "If there is a violation of that, this is very, very serious to USDA" (Interview 1. Network for a Healthy California, August 2010).

Within the Network there is a great amount of oversight and guidance provided to projects they fund. There is some variation in the amount and type of oversight and potential consequences depending on the type of project being funded by the Network. Funded projects have a Program Manager and Contract Manager, who check for progress reports, annual reports, and final reports (Interview 1. Network for a Healthy California, August 2010). Additionally, most projects have someone from the Research and Evaluation

Unit involved as well. There is a lot of oversight to make sure that activities that are supposed to occur are taking place, and activities that are not allowed by USDA are not, or are funding through another source. If work is not being completed, additional staff may become involved with the project, to make sure work is completed or approve if it is reasonable that work is not being completed. There is not usually a monetary consequence for work, such as reports, not completed, this is based on operating policies of the Network (Interview 2. Network for a Healthy California, September 2010). For special projects, contracted out to external organizations, work may be sent back for revision or the contractor may be dismissed without compensation if it is clear that they will not be able to produce the desired product (Interview 2. Network for a Healthy California, September 2010). Not surprisingly, "That would be very strong negative factor if they were to bid for another Network contract again in the future" (Interview 2. Network for a Healthy California, September 2010).

Other Contextual Factors

Additional contextual factors identified by project representatives as influencing their evaluation include: a collection of restrictions placed on them by the funder, expectations of multiple stakeholders, administrative politics, changing goals of the program, and the complexity of the food field.

A contextual factor described by all project participants as having the most influence on the program and its evaluation is the restrictions placed on the Network by its funder. USDA evaluations are based more on providing counts and fiscal monitoring than assessing if the program is effective (Interview 1. Network for a Healthy California, August 2010). Concern over work funded by Food Stamp money reaching non-eligible food stamp populations seems to guide the USDA's perspective. This has had implications for the program overall and its evaluation. Several examples of these restrictions were described during program interviews. It should be noted, that while the USDA places many restrictions on the work that the Network can do, they are also well funded through this source.

The Network must be cautious in their choice of wording and are not able to use certain words, for fear of increasing scrutiny. Certain activities are allowed and others are prohibited. Using words incorrectly can cause problems for the program, even when they are not doing activities that should give concern. An example would be funding a campaign for "Rethink Your Drink" versus "Soda-Free Summer" such to not use USDA funding for this purpose and not disparage any food item. Another example is not being able to use the word "research" and instead using "evaluation" (Interview 1. Network for a Healthy California, August 2010). Network representatives made clear that they do not do pure research and that all their work is applied and has an evaluative component to it (Network Program Representative, 2010c).

The Network is primarily a fruit and vegetable consumption program, but if using USDA funding for a school garden, the program can buy seeds, but is not able to buy shovels (Interview 1. Network for a Healthy California, August 2010). The program cannot do just physical activity, but has to add an educational component; If physical activity

movement is happening, it cannot be a "class", but instead must be called a "demonstration" (Interview 1. Network for a Healthy California, August 2010).

Another restriction, which is a prohibition, is that "we cannot disparage any foods" (Interview 1. Network for a Healthy California, August 2010). Instead, the program must use empowerment campaigns to encourage positive behavior changes, such as eating more fruits and vegetables (Interview 1. Network for a Healthy California, August 2010). "There is a very fine line between telling the facts and disparaging. We can list sugar content of various products, but we can't disparage" (Interview 4. Network for a Healthy California, September 2010).

> *"We can't say that soda will rot your teeth and make you gain weight. We can't recommend not eating fast food or not eating candy bars. Our funder, who is giving us the money, is also responsible for the economic well-being of the rest of the food industry and a huge portion is restaurants and food manufacturers" (Interview 1. Network for a Healthy California, August 2010).*

> *"Soda-free Summer" (the foundation funded portion) was seen as disparaging soda but their funding was used for "Rethink Your Drink." USDA wanted to know how would the public know that? And we told them how would we know that unless we did an evaluation? It's a circular conversation. But we do have to do the most we can and push back when we can. It takes a lot of energy. When you get into the issue of food marketing and food products, it becomes infinitely more complex" (Interview 4. Network for a Healthy California, September 2010).*

All participants described the USDA as very cautious about work being done at the environmental level, or at least their funding being used for such work. The Network is not prohibited from doing such work, but USDA restrictions make is difficult to use its funding to do this work. But it is allowed under narrow parameters. A project representative described hearing, "sure, you can do it, but just not with USDA funds" (Interview 4. Network for a Healthy California, September 2010). Community indicators available for work by the program were cut by about one-third, based on restrictions put in place by the USDA. "That definitely influences the scope of what we do. We would do more if we could" (Interview 1. Network for a Healthy California, August 2010).

> *"You can promote policy and environmental change within the context of nutrition education. It is tricky, it is very tricky, but it can be done. That's why work with the collaboratives is so important, but it's also very complicated…We stress collaboration and how far our funding can go, but they can go to the next level…It's murky and complex and has to be maneuvered very carefully. It is not a good way to run a program from a funding perspective, because it's hard to articulate. The rules, unwritten rules, are constantly changing" (Interview 4. Network for a Healthy California, September 2010).*

If the program collects data beyond what the USDA funds or wants reported, this information is used internally, to guide the program and understand, to some extent, the effectiveness of their projects.

> *"They only want to be evaluating how their funding is being used…For them to hear about outcomes related to policies or environmental changes, that's something that I don't have a way to report on, nor would they want to receive that information" (Interview 4. Network for a Healthy California, September 2010).*

It was described that the USDA was different before 2004, when the Network became much larger and more restrictions were placed on the program (Interview 2. Network for a Healthy California, September 2010). The USDA is concerned that nutrition education, funded through them, will reach people who are not low income (Interview 2. Network for a Healthy California, September 2010).

> *"Another type of funding they restrict, which affects evaluation very much, is that they won't fund surveillance or other evaluation for populations living in households earning greater than 130% of the Federal Poverty Level. That very much affects us…We can't do comparative evaluations for the CX3 project, which is neighborhood-based. We can only compare poor neighborhoods against the county average. So we can't compare poor neighborhoods to wealthy neighborhoods" (Interview 2. Network for a Healthy California, September 2010).*

The program will wait and see what happens and if the USDA changes its restrictions and perceptions around how their funding can be used, to help address childhood obesity. When the Network tried to work with the USDA around these issues, it became problematic for the program, so they are instead putting their ideas out through scientific literature and presentations (Interview 1. Network for a Healthy California, August 2010).

> *"It will be interesting to see what happens over the next few years. If overweight becomes a national security problem and with the first lady's initiative around childhood obesity. There has been a lot of attention from other health organizations as well…Their recommendations about what programs should do and what the USDA lets programs do are actually sort of divergent. We will wait for CDC to pressure USDA, not the Network" (Interview 1. Network for a Healthy California, August 2010).*

The Network and its funder are not always on the same page about priorities and how to best impact change. Part of this comes from a difference in perspectives. The Network comes from a public health perspective, whereas, the USDA comes from a philosophy around welfare and food stamps. The USDA does not agree with the idea of using their funding to changing health outcomes by surrounding an individual with environmental change (Interview 4. Network for a Healthy California, September 2010). And though their primary source of funding does not share their approach for impacting

change, Network participants described that they still believes that working at multiple levels creates necessary changes for individuals, communities and environments.

Some effort has been made to obtain funding beyond USDA, to do work restricted by USDA funding, but that the program feels is important. The program has not been particularly successful, but a "bigger barrier is that we aren't allowed to write grants on USDA money" (Interview 1. Network for a Healthy California, August 2010). A full-time program staff described having "to sign a paper semi-annually saying that I am doing 100% of my work for USDA. So, if I write a grant proposal, to do something else, I am doing that on my own time" (Interview 2. Network for a Healthy California, September 2010). And the money that comes in is not much compared to USDA funding. "It's almost easier to find a way to make what you want to do fit into USDA, even if it means leaving out components. If you can get funding to comply, it is almost a sure thing. There is not a lot of incentive to do other kinds of fundraising" (Interview 1. Network for a Healthy California, August 2010).

Since raising additional funds has been particularly challenging, working with other partners through the regional collaboratives, partners that are not funded through the USDA, has been really valuable to the program. Further information about these partnerships is presented in the above section titled: "Contextual Elements Identified by Project Participants as Influencing Network's Evaluation: Resources."

When working with multiple players and stakeholders, they may have different values, norms and expectations for what the program is supposed to accomplish (Interview 3. Network for a Healthy California, September 2010). The Network commissioned outside evaluators to interview various stakeholders, including those within the political chain of command, and found that "people's expectations were all over the place. Nobody was going to be satisfied. We had to reframe and put together realistic expectations based on resources, the environment and knowledge we had" (Interview 3. Network for a Healthy California, September 2010). Executives in the California Department of Public Health want the Network to do certain things, but it cannot because of USDA restrictions; this requires education about constraints that exist and discussions about how others might be involved to accomplish work (Interview 1. Network for a Healthy California, August 2010).

Political and administrative changes were also described as influencing the program and its evaluation. A Network representative described a difference in how the program is impacted between someone at a high level within the California Department of Public Health that is supportive versus less supportive.

> *"…if they tend to be more supportive of the Network, then we don't get as much attention, so evaluation is doing what we are going to do. But if there are people in there that are not as happy about the Network, for whatever reason, and remember, it's a huge program with a lot of money, so it does attract a lot of attention and the executives do have opinions about what they think should or shouldn't be happening. The more attention the Network gets from the executives, the more they want to see specific things in the evaluation, so that changes what we do. We have to generate certain reports or data or answer questions they want to know" (Interview 1. Network for a Healthy California, August 2010).*

The Network has shifted in its focus in goals and target audience somewhat during the course of the program. In addition to food stamp eligible people in California, and fruit and vegetable consumption, the program also works in areas such as food security and obesity prevention (Interview 1. Network for a Healthy California, August 2010).

> *"Changing goals dilutes the evaluation…it doesn't matter as much how big the program is, but more that you have a concentrated message. There has been a lot of noise around other areas, which has taken time and focus away from original goals" (Interview 1. Network for a Healthy California, August 2010).*

There is a lot of complexity in the food field. The program indicated a distinction between work they do and tobacco control, such that there is no such a thing as healthy tobacco, but there are many grey areas around food (Interview 3. Network for a Healthy California, September 2010). Plus, the food industry is a large part of the country's economy, not just agriculture, but also manufacturers, restaurants, shipping, etc. "You need to look very differently at changing something that represents about 20% of the country's economy versus something like tobacco which is a very tiny percentage" (Interview 3. Network for a Healthy California, September 2010).

"It's partly metrics, partly complexity, but estimating the cost of obesity is very complicated" (Interview 3. Network for a Healthy California, September 2010). Some may factor in lost work productivity, but may not factor in fruit and vegetable consumption; experts find it very difficult to do attributable cost or cost-effectiveness in this area (Interview 3. Network for a Healthy California, September 2010). An additional complication is that stores own the data of what is purchased there. "By law the government isn't allowed to collect that data" (Interview 3. Network for a Healthy California, September 2010).

> *"I've always been jealous of the tobacco program because they were able to gauge the success of their tobacco control program by looking at data on sales of cigarettes. Theoretically could do it in our field, if we had scanner data on sentinel soda, fruits, vegetables or other products. Somebody needs to do it, but the issue hasn't been considered important enough, nor has it been framed in that way. The field is emerging. People are very upset about the childhood obesity rate, so that becomes the metric, without looking necessarily at how you measure that metric across a population" (Interview 3. Network for a Healthy California, September 2010).*

> *"There's always a struggle about how much to evaluate something…tobacco control is much more outcome-driven, in terms of environmental changes that have resulted. It is easier to put a value on what you are doing. Here it is much harder because it is principally behavior change that we are looking at and there are a lot of limitations on how our evaluation money can be used. It makes it hard to see the overarching potential impact of it. Just the behavior change for obesity prevention, with healthy eating and physical activity, is infinitely more complex. Whereas in tobacco control, you could look at prevalence and consumption, with*

cigarettes sold, but it is hard in this field" (Interview 4. Network for a Healthy California, September 2010).

Summary for Contextual Elements

Project representatives were asked to describe contextual elements influencing the evaluation for the Network. Multiple contextual elements were identified by project staff as influencing the evaluation for the Network, both in the development of the evaluation and its comprehensiveness. Key sources of information described by project participants as guiding the development of the evaluation included published literature, theory, internal discussions, external experts, and requirements by the funder. Working with unfunded partners, and the progress they are able to make at multiple levels, was touted as greatly influencing the program and its evaluation. Having access to many sources of data, intended use by multiple audiences, building internal evaluation capacity, evaluation oversight and concerns of consequences by the funder, and changes to the funder's evaluation requirements were also described by project participants as influential to the Network's evaluation. A contextual element identified by project staff as greatly influencing the program and its evaluation are restrictions put in place by the funder. A difference in perspectives between the Network and the USDA may partly explain this disconnect.

<u>Project Strengths in the Evaluation</u>

Though the program has many challenges that it is facing, it has identified multiple strengths and successes as well. Several areas of strength, related to the evaluation, identified by project participants during interviews included: comprehensive surveillance data, well evaluated interventions before making them widely available, building of local evaluation capacity, case studies, the media evaluation and evaluating a program of such a large scope.

The Network's surveillance work was described by several key informants as a strong component of the program's evaluation. "Our surveillance surveys are good. We need to be able to generate more reports from them, but we are doing them well" (Interview 2. Network for a Healthy California, September 2010). Multiple project representatives described surveillance and other forms of data as an area of success and strength of the Network.

> *"Even though I just said the amount of data we collect and not using it was a challenge, on the other hand, we probably have some of the best data of any nutrition education program and nutrition network in the country. Some of these problems are a luxury. Other programs don't have the data at all. Having comprehensive surveillance data is really good and we've been really good at maintaining that and keeping it consistent" (Interview 1. Network for a Healthy California, August 2010).*

Another highly regarded element is the evaluation of newly developed interventions. Interventions rolled out by the state are evidence-based and evaluated before they go out to the regions (Interview 1. Network for a Healthy California, August 2010). "We have a

number of interventions that we have evaluated rigorously, that I think are quite good" (Interview 2. Network for a Healthy California, September 2010).

But it should be noted that local projects do not always, or even have to, use interventions recommended by the Network. "They have a list that they get to choose from, but there are about 500 on the list and some are better than others. The ones the state specifies, I think, are better than others because we've tested them" (Interview 1. Network for a Healthy California, August 2010).

The building of local evaluation capacity was described by multiple respondents as a great success of the program. "…our evaluation capacity building of local contractors and reporting on their findings. Both we and they are becoming very proficient with this" (Interview 2. Network for a Healthy California, September 2010).

Case studies provide rich data about the work being carried out by local projects, which might otherwise not be captured based on other reporting requirements. "Using case study reports to summarize findings about projects that take place in the same kind of venues in a variety of ways that don't lend themselves to quantitative evaluation. We've done very good formative research over the years" (Interview 2. Network for a Healthy California, September 2010).

Media tracking was also mentioned by multiple Network staff members as an element of success for the evaluation, though it has been quite a journey getting there. "We've had to learn it as we go along" (Interview 3. Network for a Healthy California, September 2010).

Lastly, some overall feedback, based on the scope of the evaluation, was provided. "I think the multi-level look at different kinds of evaluation…has yielded some good results" (Interview 4. Network for a Healthy California, September 2010). "I think we've used some innovative methods for evaluation. And, I think we're doing pretty well considering the scope, breadth and depth of what we do" (Interview 2. Network for a Healthy California, September 2010).

Network Providing Guidance to Others

Successes, lessons learned, guidance, and tools by the Network have been shared with other programs and states. Some of the Network's tools have been nationally disseminated.

The Network is held up as an example to other networks funded by USDA.

> *"In some ways it's where the other projects look to, to see what's innovative and cutting edge. But at the same time, what we're doing is so different and a comparison between California and their state doesn't always work. They don't see a lot of technology transfer, to borrow the intervention from California and use it in their state, because it's so different. But at the same time, we're looked to, to really push the envelope. It's really a mixed role" (Interview 1. Network for a Healthy California, August 2010).*

But, that mixed role often involves coming up with strategies to accomplish the Network's priorities, while trying to work within the constraints of USDA funding.

"We are also the state that gets in trouble the most. We are the state that thinks up some way around the regulations, and then USDA has to come up with a new regulation to put us back in line…We might present work at a conference and get good feedback about being innovation, but then the next year we get guidelines from USDA saying we can't do that anymore. It's made us more cautious about the kinds of things we take on and definitely about the language we use" (Interview 1. Network for a Healthy California, August 2010).

The Network provides information and guidance to other programs and states, about the work of the Network, including evaluations of specific interventions they have conducted. They share surveillance data and data collection instruments (Interview 1. Network for a Healthy California, August 2010). "I've received feedback from other networks about the usefulness about evaluation materials and validated instruments that we've shared with them" (Interview 2. Network for a Healthy California, September 2010). The program has also shared information through presentations at national and regional conferences (Interview 2. Network for a Healthy California, September 2010).

Some programs developed and evaluated by the Network have been nationally disseminated, such as "Power Play" and the Food Stamp Office Resource Kit (Interview 1. Network for a Healthy California, August 2010).

Challenges Encountered When Evaluating the Project

While the program has found many successes and provided guidance to other programs, several areas of challenge were also identified by Network participants during interviews. These include: restrictions by the USDA, getting credit for contributions to changes made, new data and reporting required by the funder, and an excess of data. Additionally, several areas were described by participants as a complication of evaluating a multi-level project, but could be considered challenges as well. These include: the scope and complexity of the program, difficulty bringing pieces together into a coherent whole, diversity of funded projects, measurement issues around media work and fruit and vegetable consumption, and assessing synergy and effectiveness of the program. These are described in further detail in the section titled "Benefits and Complications of Evaluating on Multiple Levels."

Restrictions by the USDA, such as not enabling funding to be used for work on environmental and policy levels, but the program's interest in these areas, creates unique challenges.

"It also impacts the degree to which we can ask contractors to collect data on policy and environmental change that occur in their venues. We have to ask questions indirectly and have them report very carefully to ensure that we won't get negative feedback from funder" (Interview 2. Network for a Healthy California, September 2010).

The program is not able to take credit where it is actually due. Collecting data on areas that the program has some involvement, but is careful to separate from USDA

funding, "…doesn't let us, as one would say, "toot our own horn" as much as we would like for the work that our contractors are involved in. It doesn't let us shine as much as we would be entitled to" (Interview 2. Network for a Healthy California, September 2010). "Even though we often are the organization that facilitates the ability for those things to happen, it is very difficult for us to take any credit for it" (Interview 2. Network for a Healthy California, September 2010).

> *"How do you measure or how do you claim that overall success? It's a challenge, maybe for government or for us in particular. We definitely want to have partners engaged and energized. That means giving a lot of credit to most other players. It's hard to put your arms around it and say "we did this and it helped to lay the foundation for change." and yet the ones that are claiming change are the ones that say it was a success. Often when you get other funders in there, they aren't willing to share the glory. We are often an unseen partner based on restrictions. In the future, when you have other funding opportunities, you are overlooked because you couldn't claim the impact of what you did before. It does get kind of messy out there when trying to get funding" (Interview 4. Network for a Healthy California, September 2010).*

New data required by the funder has greatly increased the amount of reporting that is done by the program. They have worked to ease the burden of data collection and reporting for local projects, but this change has been very time consuming for the program to deal with. For further discussion on this issue, please see section titled: "Contextual Elements Identified by Project Participants as Influencing Network's Evaluation: Evaluation Requirement by Funder."

While the amount of, and access to, data were described during interviews as a strength of the Network, the program also struggles with having too much data. This includes data that are outdated and not used. A further discussion of this issue is presented in the section titled "Contextual Elements Identified by Project Participants as Influencing Network's Evaluation: Intended Use of Evaluation."

> *"One of the challenges that we have is letting go of data, of the collecting of some of it. The reason this is important is because we collect probably more than we really do need. Because this is a food stamp funded program, the local projects have huge amount of paperwork that they already have to do for fiscal responsibility. When we add a lot of program stuff on top of it, it really becomes ridiculous. Some of the local contractors have full-time people who do nothing but document what they are doing. It becomes a question of utility. But, if we really did use all the data we collected, people would probably have more confidence that what they did turn in was actually useful and not just an exercise in administrative procedures" (Interview 1. Network for a Healthy California, August 2010).*

<u>What the Project Would Have Done Differently</u>

The Network identified multiple areas where they would have done things differently. These include: better dissemination of information, improving the feedback loop with local

projects, obtaining additional funding, making "success stories" more accessible, developing simpler assessment tools, changes to the process and outcome evaluations, and collecting policy data.

Dissemination of information and investing in a feedback loop were described as things the program would like to have done differently, or perhaps will do in the future. This work is not supported by the funder, however, so adding these pieces would be quite challenging. "We should produce more reports, we should publish more journal articles, and we should send updates on various projects to local projects. Where their data is going and help frame it and let them know the importance of it" (Interview 1. Network for a Healthy California, August 2010).

> *"The feedback loop part of it is, in a practical sense, optional for the funder. Not something required by the funder. The funder often wants you to move on to the next thing they require of you, once you're done with original requirements. To take extra steps and work on the project after it has concluded and disseminate information consistently is a huge amount of work. It could double the amount of work and time. To write reports and get approval through the department for dissemination could be six months, as a best case scenario, added on. It compounds workload for the next year. A lot of times that feedback loop gets left off. I know that the feedback loop is an essential part of community-based participatory evaluation and involving your participants in the evaluation results and putting the humanity back into evaluation, but that, in practicality, is very hard to implement. But that's something we should do differently and be more mindful of in the future" (Interview 1. Network for a Healthy California, August 2010).*

Obtaining funding from other sources would enable work on levels not currently addressed by the Network. The program and the primary funder do not have the same priorities, but program workload is an issue, so writing proposals and seeking other funding is a barrier as well (Interview 1. Network for a Healthy California, August 2010).

> *"…there's a big gap at the systems, environment and policy level. We are not really changing access to food, we are not addressing issues like getting higher quality food into low income areas so they can redeem their food stamps and bring in money into low income areas and get quality foods. A few projects are, but in a consistent, large-scale way, that is a huge gap in our program…That will limit the effectiveness of the project" (Interview 1. Network for a Healthy California, August 2010).*

"Success stories" are collected from funded projects every year or two, where projects describe what they thought went well. Suggestions are made for improving this component, so that they easier to use and accessible.

> *"I would like to find a way to capture those success stories in a way that would render them more useful, to compile for reporting purposes. Even something as basic as having a set of keywords that contactors would choose to go with their particular success stories. Because they*

are great stories, but you have to wade your way through, page, after page, after page, to get
any sense of what the stories are" (Interview 2. Network for a Healthy California,
September 2010).

Another area identified was developing and using simpler assessment tools to
evaluate diet and physical activity.

"I have become so much more aware of just how deficient our population is both at reading
and understanding the questions we ask, whether oral or written. And looking back at a lot
of the things we have asked, we just need to do it more simply. In the same context, I would
make sure that all survey questions and consent forms are written at the lowest literacy level
possible" (Interview 2. Network for a Healthy California, September 2010).

Unfunded partners have played a role in moving environmental and policy changes
forward. The Network plays a role in funding regional collaboratives that bring various
funded and unfunded partners together to make changes in their region. USDA restricts the
work that can be done with their funding, but unfunded partners have greatly contributed by
working in areas that USDA funded projects cannot. The specific influence of these
unfunded partners and the collaboratives has not been measured until recently and is still not
as extensive as the Network would like.

"The program has not systematically collected information on policies that have been passed
as a result of partners' involvement in the nutrition education initiatives…Even though our
funding can't claim that we caused these policies to change, many of the policies have resulted
based on these partnerships and initiatives" (Interview 4. Network for a Healthy
California, September 2010).

Regions were recently asked to provide information about policies that have been
passed and environmental changes that have resulted from the technical assistance or
nutrition education initiatives that have been carried out by the collaborative (Interview 4.
Network for a Healthy California, September 2010). It is a start, but the program would like
to do more.

"I would systematically and continuously collect information on policy and environmental
changes and get a lot more information about them…I would love to get copies of the policies
and do a much more in-depth look at exactly what these outcomes look like. To be able to
do better analysis and share with other regions and other projects, so they could see the type of
policy that yielded certain things. That would be very exciting" (Interview 4. Network for a
Healthy California, September 2010).

Participating project staff indicated other areas they would have liked to have done
the evaluation differently, including changes to the process evaluation and development of
the outcome evaluation.

"From the onset I would have wanted to have included a quantitative component and process evaluation, which wasn't done from the onset. And I would want to think harder about what that process evaluation should actually include, to be more meaningful than what we get now" (Interview 2. Network for a Healthy California, September 2010).

"I would probably involve local contractors in the process of developing outcome evaluation measures. With the caveat that the criteria have to be met and that we were going to have standardized measures. Contactors have been extremely useful in helping us develop the way we were going to collect the detailed demographic measures that had to be done this year. It was hard, it was really hard, working with them, but if we hadn't I don't think we would have the buy-in" (Interview 2. Network for a Healthy California, September 2010).

Summary for Network

The Network carries out a multi-level project aimed at changing dietary habits, such as consumption of fruits and vegetables, and physical activity in low income populations in California. A multi-level evaluation was developed by internal project staff, through many revisions to an overarching framework and use of various tools to collect data. Collection of data around activities conducted was mandated through the funding source, the USDA, with increasing requirements over the last few years. Developing a comprehensive evaluation, however, was up to the project to determine and carry out.

Benefits of evaluating the Network with a multi-level evaluation were described by participants as using the scope of the evaluation to better understand the program's impact and using this information to inform future funding.

Complications of carrying out a multi-level evaluation of a multi-level program were identified, including the scope and complexity of the program, difficulty bringing pieces together into a coherent whole, diversity of funded projects, precision of measurement, and measuring synergy and the program's effectiveness overall.

Sources of information having the most influence on the development of the evaluation for this program were described by project participants as published literature and theory combined with internal discussion, proximity to the CTCP, external evaluation feedback, and to some extent, the funder requirements.

Other contextual elements identified by project staff as influencing the comprehensiveness of the evaluation included: work by unfunded partners, access to large amounts of data, intended use of the evaluation by multiple audiences, building internal evaluation capacity, oversight by funder, politics, and complexity of the food field. Described by all participants as most influential to the comprehensiveness of the evaluation were the restrictions placed on the program by the USDA. Restrictions placed on the Network by the USDA was described by participants as having influenced the levels that they intervene and evaluate, evaluation comparisons that can be made and wording used when reporting to USDA and in media campaigns, such to not disparage any foods.

Strengths of the program's evaluation were identified by project staff as: comprehensive surveillance data, well evaluated interventions before making them widely

available, building of local evaluation capacity, case studies, media evaluation, providing guidance to other state Networks, and evaluating a program of such a large scope.

Challenges to the Network's evaluation were described by project staff as the restrictions by the program's funder, not get appropriate credit for work impact made, new data requirements by funder and having an excess of data.

It should be noted that the amount of data collected and data available were described both as a challenge and strength of the program. Multiple project representatives described surveillance and other forms of data as an area of strength for the Network. In addition to identifying data as a strength, one person also described in further detail, some challenges that exist related to data for this project. More data are collected than is used, which is a challenge to regional projects to collect and report so much data and the Network to have data that are not fully used. Determining what is really needed versus what would be "nice to have" may improve efficiency and utilization of data, but would be a difficult task.

Additionally, while the USDA places many restrictions on the work that the Network can do, they are also well funded through this source. All participating project representatives described benefits that come with adequate funding and great work that has been done through the Network, but also many examples of challenges that have come from restrictions placed on the program by the funder. Inconsistencies between priorities of the Network and the USDA, on levels to intervene and evaluate, and restrictions influencing media and community education work were brought to light during interviews. This dynamic keeps the Network doing work they feel is important, within the limits of USDA guidelines, while pushing the envelope somewhat in an attempt to further impact nutrition and activity for low-income populations in California.

In retrospect, project staff interviewed indicated that they would have improved their dissemination of information, improved the feedback loop with local projects, obtained additional funding, made "success stories" more accessible, developed simpler assessment tools, and collected more policy data. The program has recently been making strides in the areas of dissemination and collecting policy data.

It should be noted that statements about elements influencing evaluations are based on information provided by participants during interviews. Therefore, they are the ones attributing influence and describing the influence on their project and the evaluation.

Case 4 - Evaluation of the Arkansas Act 1220 of 2003 to Combat Childhood Obesity (Act 1220 Evaluation) Project

<u>Project Description</u>

The University of Arkansas for Medical Sciences' College of Public Health (COPH) is currently evaluating the implementation of Act 1220 of 2003 to Combat Childhood Obesity (Act 1220) and its various components. Phillips, Raczynski & Walker (2004) report that "Legislators felt the most effective way to address the obesity epidemic would be to focus on creating healthier environments for children through the involvement of schools, parents and communities"(p. 3). The Act 1220 Evaluation Project evaluates how Act 1220 "impacts school environments, parental knowledge of childhood obesity issues, and the nutrition and physical activity behaviors of Arkansas students and their families" (Robert Wood Johnson Foundation, 2009).

Phillips, et al. (2005) describe that "In summary, this legislation called for a coordinated, multi-level effort to combat childhood obesity" (p.5). The main provisions of Act 1220 required the following (Phillips et al., 2004):

- Annual body mass index (BMI) screenings for all public school students, with the results reported to parents;
- Restricted access to vending machines in public elementary schools;
- Disclosure of schools' contracts with food and beverage companies;
- Creation of district advisory committees made up of parents, teachers and local community leaders; and
- Creation of a Child Health Advisory Committee to recommend additional physical activity and nutrition standards for public schools.

> *"The really innovative and forward thinking pieces of it were the committees that it required be set up. There was a state-level Child Health Advisory Committee comprised of a number of representatives from stakeholder groups. Stakeholder groups were actually named in the law. It required each school district to set up a committee to advise it about additional policies. The state level committee was to meet on a regular basis and make recommendations to the state Department of Education about future policies to be implemented" (Interview 3. Act 1220 Evaluation Project, October 2010).*

Act 1220 was passed in April 2003, and enacted in July 2003, but was not implemented in schools during the 2003-2004 school year (Interview 2. Act 1220 Evaluation Project, October 2010). Baseline data were collected in Spring 2004 and Act 1220 implementation began in schools during the 2004-2005 school year (Interview 2. Act 1220 Evaluation Project, October 2010). There were delays in starting implementation while the CHAC was established and reviewing evidence to make recommendations and schools prepared to move forward with elements in the legislation.

> *"Act 1220, right after it was passed, it actually passed with very little discussion and notice. It didn't really come to anyone's attention until right after it was passed. It passed in April*

2003 and became effective in July. During that period, from April to July, the schools looked up and said "you want us to do what?" and then it began to get noticed" (Interview 3. Act 1220 Evaluation Project, October 2010).

Interestingly, Act 1220 did not provide funds for either the implementation of intervention activities or an evaluation (Act 1220 Evaluation Project Representative, 2010). Schools were provided no funding to make changes mandated through Act 1220. "It's very interesting that it happened as smoothly as it did" without funding (Interview 1. Act 1220 Evaluation Project, October 2010). COPH's evaluation project is funded fully by the Robert Wood Johnson Foundation (RWJF). The evaluation is currently funded through the RWJF until 2012 (Act 1220 Evaluation Project Representative, 2010).

The ultimate goal of Act 1220 and its various intervention activities is to reduce child obesity in the state of Arkansas (Act 1220 Evaluation Project Representative, 2010). BMI data on all children in public schools in Arkansas are measured on an annual basis to assess this long-term goal (Act 1220 Evaluation Project Representative, 2010). While beyond the scope of the current project being studied, the RWJF also funds a project related to the collection and data system for student BMI, called the Arkansas BMI Database Project (Robert Wood Johnson Foundation, 2009).

Act 1220 works toward the above goals by intervening on multiple socioecological levels concurrently. Innovative ways to address childhood obesity include intervening on multiple levels concurrently, such as the individual, family, school, community, organizational, and policy levels. The goal is often to use multiple avenues to create a change in social norms, expectations and environments around nutrition and physical activity for children, their families and communities (Chapter 2 provides further detail about socioecological levels). Act 1220 and its various components <u>intervene</u> on the following socioecological levels (with activities at each level presented in parentheses): *individual* (student health report sent to parents and youth/parent education); *schools (*vending machine restrictions, cafeteria nutrition, physical activity); and *policy* and *committees* (overarching between levels, covering statewide, school and district-wide) (Act 1220 Evaluation Project Representative, 2010). Intervention activities are carried out through work by schools and committees throughout the state of Arkansas. The COPH is specifically related to the evaluation of Act 1220 only, not the intervention.

> *"The whole concept of childhood obesity, and its etiology, is very multi-factorial, it really does span practically every level of the socioecological model, from inside the individual, all the way up to federal policy. There are lots of questions to be asked and answered about how all of those things work at various levels" (Interview 3. Act 1220 Evaluation Project, October 2010).*

According the RWJF, the goals of the Act 1220 evaluation are to (Robert Wood Johnson Foundation, 2009):

- Characterize the processes for enacting and implementing all provisions of Act 1220.

- Assess changes in school environments relating to nutrition and physical activity.
- Assess changes in student and family behavior regarding nutrition and physical activity that may be associated with school environment changes.

The evaluation of Act 1220 and its various components also takes place on multiple socioecological levels concurrently. <u>Evaluation</u> takes place on the following socioecological levels (measuring areas of): *individual* (parent and youth knowledge, beliefs, attitudes and behavior around food and activity); *schools (*vending machine restrictions, cafeteria nutrition, physical activity); and *policy* and *committees* (overarching between levels, covering statewide, school and district-wide) (Act 1220 Evaluation Project Representative, 2010).

For the evaluation, data are collected on an annual basis from surveys with principals and superintendents, interviews with parents and adolescents, as well as other stakeholders involved in the implementation of Act 1220. Additionally, as part of Act 1220, BMI is measured for all students on an annual basis and reported confidentially to parents. Other components have been added to the evaluation to provide additional information about the Act's implementation and other areas of interest to show impact of the intervention. Annual evaluation reports are developed and made publicly available.

The COPH has written annual evaluation reports since the start of the project in 2004. Information about the history of Act 1220 and baseline data are described in the Year 1 report; Progress of Act 1220's implementation and evaluation findings are described in Year 2- 6 annual reports (Fay W. Boozman College of Public Health, 2010). Reports for each year are made available at: http://www.uams.edu/coph/reports/#Obesity.

The first BMI assessment found that 38% of the Arkansas' school-aged children were overweight or at risk for overweight; these rates were much higher than previous self-reported estimated had found (Phillips et al., 2004). Phillips et al. (2004) report:

> *"Baseline surveys showed parents frequently are unable to characterize accurately their child's weight status, particularly when the child is overweight. More than half (51%) of the parents of children who were overweight according to BMI-for-age percentiles incorrectly perceived the child to be of normal weight." (p.3)*

The Year Six report by Phillips et al. (2009) concluded that:

> *"This evaluation suggests that, six years after the passage of Act 1220, school environments have become healthier for students and staff. Schools continued to make changes to vending contents and other nutrition policies and to maintain changes that have been made since Year 1. Families have become more aware of the serious health problems associated with childhood obesity and generally recognize the role that schools can play in combating the problem. To date, there is little evidence showing that families have made substantial changes to improve their children's nutritional environment outside of school, yet there is growing evidence to suggest that parents have helped their children make physical activity a priority. In this past year, adolescents continued to report increased physical activity, as well as a decline in vending purchases and soda consumption. For the first time, there is preliminary evidence to*

suggest that adolescents may be reducing their consumption of fast food. It will be important to monitor this potential trend." (p.7, Executive Summary)

This 2009 evaluation report also identifies facilitators and barriers to making changes in schools, including the need for updated facilities and equipment and the need for additional funding and support to carry out Act 1220 activities (Phillips et al., 2009).

Currently, it appears that changes to healthier nutrition is occurring in public schools of Arkansas, but BMI rates have not changed (Act 1220 Evaluation Project Representative, 2010).

Benefits and Complications of Evaluating on Multiple Levels

The Act 1220 Evaluation Project was asked to describe how they benefitted by carrying out a multi-level evaluation of Act 1220, and its various components. This project was also asked how the evaluation was made more complicated by evaluating on multiple levels.

Benefits of evaluating Act 1220 with a multi-level evaluation were described by project staff as: providing a more comprehensive evaluation, capturing data to better understand changes seen or missing, and having data to address concerns around unanticipated consequences.

The project described evaluating on multiple levels as allowing for "a more comprehensive view of the impact and the process of the implementation of the Act itself in Arkansas" (Interview 2. Act 1220 Evaluation Project, October 2010). Working from a comprehensive intervention, collecting data from multiple sources, collecting data from the beginning, and capturing multiple things at once, were described as benefits of the evaluation, which also influenced its comprehensiveness (Interview 1. Act 1220 Evaluation Project, October 2010).

> *"I'm not sure this evaluation could have been done looking at a single level of measurement, looking only at schools, or only at parents, but doing it that way would have meant that we would not have been able to understand how and why things were or were not changing. We're not seeing a lot of changes around parental or family behaviors around nutrition and physical activity. Well, we could suppose why that is, but because we have information about what is happening in the schools, we can rule in some ideas and rule out some ideas, based on our knowledge about what they are exposed to. I don't know how you'd do it thoroughly without measuring at multiple levels" (Interview 3. Act 1220 Evaluation Project, October 2010).*

Multiple sources of data, and data collected at several levels, have allowed the project to understand elements that were not necessarily understood early on, but ended up being important. Unintended consequences were of particular concern, but early on, the project was not sure what these might end up being or which would be most important. Concerns around vending machine revenues, an increase in dieting or eating disorders for students, among other concerns, have been raised throughout the project. The evaluation has been

able to address many of these concerns based on data that had already been collected or was able added in for collection.

> *"…since we really didn't know what concerns would arise within the state, about the implementation of the Act, and certainly a number of concerns about adverse outcomes have emerged, a comprehensive approach allowed us to cast a fairly broad net, to have the data to better inform additional decision-making, including additional legislation and legislative intents surrounding some of the adverse outcomes" (Interview 2. Act 1220 Evaluation Project, October 2010).*

> *"We were able to get in and do a baseline before all of these policies and things started changing, which was good because we were able to get in there and show some of these things people were concerned about, unintended negative consequences, such as emotional well-being and eating disorders. We've been able to compare what happened after the initiative to what was happening before the initiate. We've been able to allay some fears" (Interview 3. Act 1220 Evaluation Project, October 2010).*

Complications of carrying out a multi-level evaluation were identified, including the lack of control and knowledge of the ultimate direction for the project and fitting multiple pieces, sources and types of data together into a coherent whole.

When asked about complications of evaluating on multiple levels, one team member described a "moving target" (Interview 1. Act 1220 Evaluation Project, October 2010).

The natural environment the project is taking place in, and lack of control, were also identified as adding to the complexity of this project's evaluation (Interview 1. Act 1220 Evaluation Project, October 2010).

> *"We know that we are not gathering all the information, there's no way we could. So, the complicatedness of what we are trying to gather and then how that might be impacted by media, or things happening in the community, or all those other factors too, makes it all very complicated. But it's important to be as comprehensive as possible and to gather as much of that information as possible, if we really want to have a clearer picture of what is happening in the schools versus what may be happening in the homes" (Interview 1. Act 1220 Evaluation Project, October 2010).*

Large amounts of data, from various participants and stakeholders, among other sources, add to the complexity of the evaluation. Beyond just the amount of data, however, is the more complicated issue of how they fit together and what they mean.

> *"Anytime you've got multiple levels of measurement going on at the same time, you've got logistical issues going on, but you also have the issue of how do you fit it together in the final understanding. How does this fit together into a big whole? We have a jig-saw puzzle going on…fitting all the pieces together and understanding how things are moving and progressing is made more complicated when you have multiple informants and multiple pieces to fit. It's*

not just a simple logistical model that gives you a set of parameter estimates. I mean, it's like gears turning, one moves and another one moves, but not necessarily in the same direction. Keeping up with all of that is conceptually a little bit more challenging" (Interview 3. Act 1220 Evaluation Project, October 2010).

"Coming up with not just the quantitative stuff, but also melding into all of that numeric understanding, a qualitative understanding of why people are giving the numerical answers the way they are, makes it more complicated, but it also makes it richer…We can compare what superintendents and principles say and think, compared to what parents and adolescents say and think. It adds richness to the evaluation, that I think is extremely valuable for decision makers, it does however, make it more complicated for the evaluators, to put it all together into a coherent whole" (Interview 3. Act 1220 Evaluation Project, October 2010).

One team member, however, was not sure that evaluating on multiple levels necessarily made the project more complicated.

"I don't know that it was really made that much more complicated. It was made more comprehensive and it required staff who was experienced and flexible…But I wouldn't say that it made it that much more complicated. It was just juggling a few more balls, if you will" (Interview 2. Act 1220 Evaluation Project, October 2010).

Contextual Elements Identified by Project Participants as Influencing Act 1220's Evaluation

Project representatives were asked to describe contextual elements influencing Act 1220's Evaluation Project. In addition to an opportunity for the project to describe other contextual elements influencing the evaluation, questions were asked about specific contextual elements. Multiple contextual elements were described during interviews as having influenced the evaluation for this project, both in the development of the evaluation and the comprehensiveness.

The following contextual elements were asked about with interview participants and are described in the following section: Sources of Information Guiding Development of the Evaluation, Resources, Evaluation Requirement by Funder, Length of Time Carrying Out Intervention/ Evaluation, Intended Use of Evaluation, Evaluator Access and Capacity, Evaluation Oversight and Consequence, and Other Contextual Factors.

Sources of Information Guiding Development of the Evaluation

The primary sources influencing the comprehensiveness of the evaluation, identified by project participants, were the project's logic model and guidance provided by the RWJF. Additional sources were influential in guiding the development of the evaluation, including the legislation, expertise of evaluation team members, access to legislators, and information from other fields.

The comprehensiveness of the evaluation was described as being "based on the logic model that we developed to really describe what we thought would impact on childhood

obesity as a result of the Act itself. And theory, to better inform development of the logic model" (Interview 2. Act 1220 Evaluation Project, October 2010). This logic model was theory-based, using social cognitive theory, and developed within their team (Interview 2. Act 1220 Evaluation Project, October 2010).

> *"We started out by developing a logic model, a conceptual model that would frame for us how we thought the elements of Act 1220, in the context in which they occurred were likely, logically, to achieve the outcomes that the framers of the legislation wanted. Ultimately they wanted to increase the percentage of students who were achieving and maintaining a healthy weight. We developed a logic model for that. That logic model gave us a framework for our evaluation. It pretty much told us what we needed to measure" (Interview 3. Act 1220 Evaluation Project, October 2010).*

External evaluation expertise from the RWJF provided support and guidance for the development of the project's evaluation, including its comprehensiveness.

> *"Fortunately we had gotten funding from RWJF and they have an excellent evaluation team there, particularly their lead evaluator, Laura Leviton. At the time this evaluation project was starting we were in constant communication with her. Having her guidance probably helped a lot at being as comprehensive as we are" (Interview 1. Act 1220 Evaluation Project, October 2010).*

The development of the evaluation was guided by multiple sources including the legislation itself, the project's logic model, expertise of evaluators on the team, and access to legislators. The logic model and expertise of team members were particular helpful to the project's evaluation in figuring out the direction the project may end up heading

> *"Obviously we used the legislation. That was our starting point. To say, okay, these are the things that may be changing, so these are the things we obviously want to be looking at. That was our jumping off point to look for related issues that we might want to gather information about. Besides the legislation, we also used some more theoretical frameworks around evaluation and community-based public health and health behavior sorts of information" (Interview 1. Act 1220 Evaluation Project, October 2010).*

The project sought to address various components of the Act's intervention, including a better understanding of perceptions by schools, parents, students, and decision-makers.

> *"Decisions were made to try to find out what parents thought about doing BMI in schools, what they thought about getting a letter at home, what they thought about nutrition in the home. Also in the school setting, trying to find out what students thought, what principals and superintendents thought and how things were implemented in the schools, how that was going" (Interview 1. Act 1220 Evaluation Project, October 2010).*

The evaluation was helped by the specificity of the legislation in some ways, such as mandates around vending machines and measuring BMI for all students. But committees to be developed and provide ongoing recommendations for future work proved harder to know how and what to measure (Interview 3. Act 1220 Evaluation Project, October 2010). "We sort of had to use the crystal ball by using what we knew about schools and about things that were likely to be recommended…And then we set about measuring those things" (Interview 3. Act 1220 Evaluation Project, October 2010).

> *"We pretty much had to evaluate what was happening at the school level, what the policies were like before and how they changed over time. We know from our interactions with schools that schools are also influenced by district-level policy decisions, so we measured things at the district level. Because we expected, as described in the logic model, that policies and practices and environments within the schools would influence the awareness of parents and students, and the public at large, about weight and around nutrition and physical activity, we needed to measure knowledge, attitudes, and beliefs that were held by parents and students and then we needed to also measure their behaviors. Because ultimately behavior change is what leads to changes in weight management" (Interview 3. Act 1220 Evaluation Project, October 2010).*

While the use of a crystal ball was described, to describe developing an evaluation for a project with an unknown future direction, the project did have two particularly valuable sources of information available. "…mostly it was the logic model and varying perspectives of the team that helped us be able to figure out how things were likely to unfold" (Interview 3. Act 1220 Evaluation Project, October 2010).

> *"Members of our project team had expertise in many of the methods, which naturally evolved out of the logic model itself…All sources of the data have contributed to better informing the project….Team members informed the methods that were developed to collect the data" (Interview 2. Act 1220 Evaluation Project, October 2010).*

In additional to general skills and expertise in evaluation, team members had specific knowledge around working with schools, evaluating community-based programs, and policy evaluation (Interview 3. Act 1220 Evaluation Project, October 2010). The team developing and carrying out the project's evaluation was also interdisciplinary, including expertise in areas such as health behavior, health policy, epidemiology, biostatistics, and psychology.

> *"Our team, as a composite, included all of those different perspectives, skills and expertise. The other thing we brought to the table was a strong understanding of the whole weight control, weight management, nutrition, physical activity, built environment, sort of state of the art recommendations around childhood obesity, so that we could anticipate, both from our own knowledge, from literature reviews and from discussions with the framers of the legislation, we were able to get a sense of what they expected, what they hoped to get out of it*

126

and the intent…we were able to anticipate where all of that was going" (Interview 3. Act 1220 Evaluation Project, October 2010).

An additional source of information described as valuable in the development of the evaluation was access to legislators involved in the development of Act 1220.

"We are very fortunate that here in Arkansas, it is a small state, you have good access…we had access to all of those individuals, to sort of pick their brains about what they were trying to do, what they expected, what they considered, but didn't make it into the bill, so we had that information" (Interview 3. Act 1220 Evaluation Project, October 2010).

Others fields contributing to the evaluation, beyond the project's team from Public Health, included: Education, in understanding schools, how they work, and collecting data in these environments; Child Development, for understanding age appropriate responses to certain things; Family Development, to better understand how decisions are made in families, who controls the menu and how eating behaviors change over the lifespan of child; and Psychology, around eating disorders and weight control (Interview 3. Act 1220 Evaluation Project, October 2010). The project also worked with the Arkansas Department of Education and Department of Health (Interview 1. Act 1220 Evaluation Project, October 2010).

Resources

Multiple resources have helped to move Act 1220's evaluation forward. Resources specifically mentioned during interviews included funding from the RWJF, relationships built with and access to schools, survey partners, legislative experts, and communication partners.

Seemingly, the most critical resource to the project was the funding made available through the RWJF. Since there was no funding put in the place, for either the intervention activities or the evaluation, and evaluation was not mandated as part of the Act itself, without this source of funding, the evaluation likely would not have happened. "Well, the funding from the foundation was certainly invaluable. The project couldn't have been done without that funding coming in and that, of course, enabled us to hire the staff that have implemented the data collection for the project itself" (Interview 2. Act 1220 Evaluation Project, October 2010).

Access to schools, resources at the Department of Education and Department of Health, and the Child Health Advisory Committee were specifically named as valuable resources of the evaluation project (Interview 1. Act 1220 Evaluation Project, October 2010). Additionally, survey research partners were described as "instrumental" to the evaluation, which uses surveys to collect data from multiple sources (Interview 3. Act 1220 Evaluation Project, October 2010).

"I don't think that we would have as much information as we have if we hadn't been able to reach the principals and superintendents on a regular basis and then give them back the

information we have gathered. That has been a really valuable partnership for us" (Interview 1. Act 1220 Evaluation Project, October 2010).

Key informants in the legislative policy arena were also considered valuable resources to the evaluation and project team members. Not only did these experts help the evaluation team better understand the legislative process and decision-maker strategies, but also how best to communicate findings to decisions-makers, such as "in what form, how often, and when during the legislative session" (Interview 3. Act 1220 Evaluation Project, October 2010). Additional communication strategies have been shared with the evaluation team by another RWJF-funded group.

> *"While we have written scientific manuscripts from the data, and continue to do that, you don't communicate with decision-makers in the policy arena in the same way that you communicate with a more scientific audience. That communications firm has been extraordinarily valuable to us in helping us communicate to that audience in a way that, I think, has been very effective" (Interview 3. Act 1220 Evaluation Project, October 2010).*

Evaluation Requirement by Funder

This evaluation was unique in that it was not mandated for inclusion by Act 1220's legislation and no funding was provided for evaluation of intervention activities. Further, funding for the evaluation was obtained by an outside source not linked to the intervention activities, intervention decision-makers, or the evaluation team. Act 1220 mandated intervention activities to take place, with no funding in place for intervention activities or an evaluation, but the evaluation took place based on funding acquired through the RWJF. As described earlier, it was the College of Public Health (COPH) at the University of Arkansas for Medical Sciences (UAMS) that received this funding and has carried out the evaluation. Requirements and oversight of the evaluation, therefore, would be with the evaluation's funder, the RWJF. This unique dynamic made the evaluation of the Act and related intervention activities external to the intervention and its decision-makers.

Evaluation team members were asked how the RWJF came to fund this evaluation project and whether specific evaluation elements were required for inclusion by the funder. In short, "No specific elements were required in the evaluation…We've always had scientific control" (Interview 3. Act 1220 Evaluation Project, October 2010).

> *"It was sort of a two-way street, we kind of met in the middle…There were people here at the college thought this had potential to be important and we really should evaluate it. And people at Robert Wood Johnson were looking at it and thought it was a cool natural experimental in childhood obesity and someone should be evaluating it. Then there was a team from another part of the institution here that was looking at the mandate for the schools to collect BMI, and saying that the schools are really not prepared to do this effectively and it is a real opportunity for us to have a nice surveillance program, but we need to be able to do it in a way that makes it consistent. The last thing you want is for schools to go out to Wal-Mart and get $9.99 scales and having students jumping up and down on them. And*

schools really had no idea how to do this. So, our colleagues in that group were talking with Robert Wood Johnson about getting funding for a database and standard protocol, done in a medically appropriate way. The college started talking with RWJF about doing an evaluation. They were also reaching out to us at the same time. It all happened about the same time… It wasn't much of a leap for us to go to them. NIH wasn't an option because of the timing, we couldn't wait nine months for funding, which would have been too late, so it really meant foundation funding" (Interview 3. Act 1220 Evaluation Project, October 2010).

"We were approached by the foundation when they heard the Act was going to be implemented. We were asked to design an evaluation. I'm sure there was some back and forth discussion about what that would involve, particularly with our program officer at the time…And we got feedback from the application we submitted, but there were no requirements, per se, other than that it be comprehensive" (Interview 2. Act 1220 Evaluation Project, October 2010).

The evaluation was conceptualized as one that could be somewhat fluid, to accommodate recommendations, policies and new activities, as they unfolded (Interview 3. Act 1220 Evaluation Project, October 2010).

"…there have been things happen along the way that have became of more interest, that we added as things went on and changed in the schools, but it's all related the same basic core questions. We are trying to be somewhat adaptable, as much as we can" (Interview 1. Act 1220 Evaluation Project, October 2010).

"As the whole arena and knowledge base around childhood obesity would unfold and change, we felt strongly that we needed to be able to add questions or add components….About two years ago we added a case study component. When we realized there were some schools that were extraordinarily effective at reducing the proportion of students who are overweight. The question is why? How? What makes them different? What makes them special? Why are they able to accomplish this when others have not been able to accomplish it? What can we learn from those really successful schools? We added in-depth assessment of those schools, to get more knowledge about that" (Interview 3. Act 1220 Evaluation Project, October 2010).

The COPH feels there is much more to be evaluated and learned in this area. Other funding applications are pending to investigate several areas in more depth, including school wellness plans, adolescent snacking behaviors, and how snacking behaviors differ in rural versus urban settings (Interview 3. Act 1220 Evaluation Project, October 2010).

Length of Time Carrying Out Intervention/Evaluation

The evaluation of the project has been influenced by many things, but the length of time intervening and evaluating the project was not described by participants as being particularly influential to evaluation, as compared to other elements.

"People are amazed that we've been funded now for seven consecutive years and we expect to be funded for another two, maybe three. This makes it one of the longest running evaluations of its kind" (Interview 3. Act 1220 Evaluation Project, October 2010). When Act 1220 had been passed, but not yet enacted, the COPH began talks with RWJF about funding the evaluation. The evaluation was able to collect baseline data before schools had begun to implement intervention activities. The evaluation has been somewhat flexible, to accommodate changes in the intervention, recommendations made by the CHAC and other committees, and address concerns that have been raised by decision-makers and others. It appears that flexibility of the evaluation, more than the length of time involved with the intervention or evaluation, has influenced the evaluation in this area.

Intended Use of Evaluation

Descriptions of multiple uses and multiple users of the evaluation and its findings were described by Act 1220's Evaluation team members. Intended use was described by interview participants as having influenced the evaluation's design, methods, questions included, and modifications to the evaluation to stay responsive to stakeholders. Being able to give information to various stakeholders and decision-makers has shaped the evaluation that was developed and has been carried out by this project. Descriptions of how intended use were identified by project participants as influencing the evaluation are presented below.

> *"The intended use drove the entire design of the evaluation. There were multiple uses intended for it. It was certainly intended to better inform the scientific literature and policy makers in other states, so the use had to be fairly broad and thorough and assess the process and impact of the Act itself. But it was also intended to inform additional policy development in Arkansas. And since it was anticipated that there might be some that would come back and try to, in essence, try to undo aspects of the Act, it was intended to try to incorporate an evaluation that would address the unintended consequences that emerged over time" (Interview 2. Act 1220 Evaluation Project, October 2010).*

> *"I think ultimately the intended use of the evaluation is to be able to say these kinds of legislative or policy or regulation changes, in a system like public schools, can have an effect on childhood obesity. That's ultimately our big picture. Does this seem to have some sort of effect on the end-product? Which is a really long-term question, but along the way that becomes particular questions that the legislature has or the superintendents have about some piece of this system and whether there is a way to fine-tune it or way to change things, so that people are more comfortable with it. This legislation addressed so many different things like, PE requirements, wellness plans, Child Health Advisory Committee, BMI. There are so many pieces to it that different questions arise along the way. We've tried to give them that information as much as we can" (Interview 1. Act 1220 Evaluation Project, October 2010).*

"To a large extent, how we thought people would use the data impacted a lot what we asked and how we asked it. As other questions from decision makers have arisen, we have tried to modify the evaluation to address those questions. As the state Department of Education, school principals, superintendents, or members of the legislature have asked us questions, we've tried to analyze data we already have that shines light on the answers or in the upcoming data collection cycle we will address it, if at all possible" (Interview 3. Act 1220 Evaluation Project, October 2010).

Evaluator Access and Capacity

The project overwhelming described a lack of evaluators able to conduct complex evaluations in general, both in the access to evaluators and their capacity to carry out such evaluations. The project, however, was fortunate to have built a team with a great amount of capacity to carry out such work. Therefore, evaluator access and capacity were described by participants as an issue for the field overall, and may influence developing and carrying out these types of evaluations, but it did not directly influence this project's evaluation.

"I think we desperately need more trained evaluators, we need more people in our state health departments, in our community based settings, who are available to work with program implementation specialists, program planners, to evaluate what's going on and to help feed that information back for program improvement. I don't think we do as much of that as we should. And it might be that there aren't a lot of them out there" (Interview 3. Act 1220 Evaluation Project, October 2010).

"In my experience, people talk about evaluation a lot, but they don't really know what it is. It seems to be talked about as a required element, but people have a lot of different ideas of what that would entail and what the report or end-result or goal of it is. That's a long way around to saying I don't think there are many evaluators out there…It's probably both capacity and access. I'm not really sure who or what or where would be teaching people how to do evaluation, so I think it's probably both. I think people are trying to do it without a lot of training and not quite knowing what they are doing" (Interview 1. Act 1220 Evaluation Project, October 2010).

Evaluation Oversight and Consequence

Project members indicated that they are not accountable to the legislature since they did not fund the evaluation. "We try to keep an ear out to questions they are asking and to the extent possible, try to provide them with data to make informed decisions, but they don't have any oversight or control over us" (Interview 3. Act 1220 Evaluation Project, October 2010).

The project did describe their accountability and working relationship with their funder, the RWJF. A positive working relationship, carrying out the evaluation as intended and an internal system for monitoring the project were described. Though there were not specific elements of the evaluation required by the funder, to be held accountable for,

oversight that goes along with funding of this nature is present.

> *"We are clearly accountable to Robert Wood Johnson, as our funder. If we didn't complete something they would raise their eyebrows and want to know why not, but they aren't terribly heavy handed about it… There is oversight, definitely, we are responsible to them to be sure that the monies are spent appropriately, that the activities we said we were going to do, that they gave us money for, we have actually completed. If there are unexpended funds, we justify why, there are standard grant oversight sort of stuff, but there aren't any real consequences. At least there haven't been. For one thing, we pretty much do what we said we were going to, and don't intend to change things a lot. And if we do, our project officer understands why and approves it up front. It's a very collegial relationship" (Interview 3. Act 1220 Evaluation Project, October 2010).*

> *"Well, there were no required elements. If we had not conformed to the application that we submitted and hadn't gotten approval to depart from our plan, from the Foundation, then I'm sure they would have come back with questions. The evaluation has really been implemented pretty much as designed, with internal monitoring and feedback. Members of the investigative team have a lot of experience with multi-center clinical trials and monitoring large projects. We have a process set up to monitor the data collection and ensure that we are meeting our targets, or if we're not, to implement corrective actions early. There were no times that there were any problems" (Interview 2. Act 1220 Evaluation Project, October 2010).*

Other Contextual Factors

Most of the additional contextual factors mentioned by the project have been described previously, but were described by participants as influencing the evaluation, so they are mentioned again briefly here. These include the funding by RWJF, post hoc realization by decision-makers about the information that can be obtained through evaluation, adding to the evaluation to address anticipated concerns by various groups, and a lack of control over the contextual within which the intervention and evaluation took place.

Having funding from the RWJF was identified as having a huge influence on this evaluation, which may not have taken place, or at least not when it did, if this funding wasn't available. "Money always affects what you do. We are fortunate to be fairly well funded. On the other hand, the funding we have for telephone surveys affects sample size and a few other things" (Interview 3. Act 1220 Evaluation Project, October 2010).

> *"The opportunity for this particular part of the project, to have funding from Robert Wood Johnson Foundation, to do an evaluation, was a timely and important coincidence that everyone has benefited from…I think that's probably the biggest factor for the evaluation piece" (Interview 1. Act 1220 Evaluation Project, October 2010).*

Additionally, decision-makers have realized since passing Act 1220 with no funding for evaluation, the role of evaluation for better understanding the intervention and any impact it has had.

"…it was as if a lot of people realized post hoc, how are we going to know if this is having any effect or not, like legislators and the Department of Education, so we share our information with a lot of them, because that sort of developed after the fact" (Interview 1. Act 1220 Evaluation Project, October 2010).

As mentioned earlier, anticipated concerns in several areas were included in the evaluation, to alleviate concerns brought up by several stakeholders, decision-makers and special interest groups, such as the vending machine industry.

"…we monitored vending changes that were occurring, and asked parents, school personnel and adolescents about satisfaction with changes and beliefs about changes…These were things in the environment, special interest groups, if you will, that we anticipated might try to weigh in on the Act in future years. Elements that they were concerned about were included in the evaluation" (Interview 2. Act 1220 Evaluation Project, October 2010).

The lack of control and comparison groups was considered a contextual factor described by project participants as influencing the evaluation. The evaluation team took these constraints into consideration and developed an evaluation to measure what they could and answer as many questions as they could, given the circumstances and changing context within which the intervention and evaluation were taking place.

"Part of the issue is that this is a statewide natural experiment. We don't have a control group. Every school in the state is impacted. We haven't been able to identify another state that is similar enough to serve as a control group without other factors that would contaminate the comparison. It's not a randomized control trial, you just can't do this kind of thing this way. This is a big contextual factor that framed this evaluation. What we can look at is the natural variation in implementation of these policies. Who is out there first? Who are the early adopters? Who are the ones that go kicking and screaming? Who are the successful schools? Which policies are more likely to get changed first and which come along later? Why would that be happening? There are lots of natural variation questions that we can ask and answer. We can answer them, even though we can't do the more controlled evaluation study that you would like to do" (Interview 3. Act 1220 Evaluation Project, October 2010).

Summary for Contextual Elements

Project representatives were asked to describe contextual elements influencing the evaluation of Act 1220. Multiple contextual elements were identified by project staff as influencing Act 1220's Evaluation Project, both in the development of the evaluation and its comprehensiveness. Contextual elements identified by project staff as influencing the project's evaluation included: sources of information, resources, intended use, maintaining flexibility to address concerns, and working within the natural environment that the intervention was taking place. Not particularly influencing the evaluation: evaluation

requirements by the funder, length of time carrying out the evaluation, evaluator access and capacity, and evaluation oversight or consequence. Key sources of information for guiding the development of the evaluation included the project's logic model, expertise by evaluation team members, and guidance by the funder. Resources of most value were described by project staff as funding by the RWJF and relationships with schools and other partners. Multiple uses of the evaluation, and intended use by multiple users, were described by participants as influencing the evaluation's design, questions asked and modifications made to the evaluation. Members of the evaluation team feel that there is a lack of access and capacity to quality evaluators overall, but this did not greatly influence their specific evaluation, as they had and were able to maintain that capacity within their team. Specific elements of the evaluation were not required by the funder and a collegial working relationship with the funder, combined with following deliverables laid out, has not influenced the evaluation out of concern of oversight or consequences. Other contextual elements identified by project representatives as influencing the evaluation included: funding by RWJF, post hoc realization by decision-makers about the information gained by the evaluation, modifications to address emerging areas of concern, and the lack of control from the setting in which the intervention and evaluation activities were occurring.

Project Strengths in the Evaluation

The evaluation of Act 1220 and its various components have multiple strengths. When asked about these strengths, several were identified, including the initial logic model and framework for the evaluation, remaining flexible to accommodate changes to the evaluation and address emerging concerns, building lasting relationships and raising awareness around childhood obesity with various audiences. Additionally, the evaluation project has shared ideas, instruments and lessons learned with other projects, as described in further detail in the next section.

"I think we did a good job in thinking initially through the logic model that we anticipated, based on theory, what would drive the changes that occurred as a result of 1220 and designed an evaluation that would address it" (Interview 2. Act 1220 Evaluation Project, October 2010). In addition to the logic model, the evaluation's methods and role this has played in understanding what is going on were described during interviews as areas the project did best in its evaluation. The amount of data from various levels and sources and implications for the questions they can answer are getting noticed by others.

> *"I think having a conceptual model that guided the development of the evaluation, and has served to help us frame it over time, was a good thing…I also think having a mixed methods approach has been a really good thing. Having both the qualitative and the quantitative. Having both the interviews and the surveys has allowed us to gather and capture a lot of information fairly efficiently. And to really understand it or to gain insights into what other questions should be asked and answered" (Interview 3. Act 1220 Evaluation Project, October 2010).*

"People are fascinated by the fact that we have so much information about the different levels, about what's happening in schools and what's happening in families. We can look at all the natural variation stuff. The foundation in particular is particularly interested now that we are at the point where we can do some longitudinal work with the schools. Looking at where they started and where they are coming out. They are really pushing now for us to help answer the question "Does changing school policy affect child weight, in a very specific way?" (Interview 3. Act 1220 Evaluation Project, October 2010).

"I think we did well in being flexible enough to think about how we could better enrich the planned evaluation over time" (Interview 2. Act 1220 Evaluation Project, October 2010). This flexibility has enabled additional evaluation elements to be added, to address emerging concerns, both anticipated and not anticipated.

"One concern, that we hadn't anticipated early on, is that some people were concerned that when parents got letter about BMI category, explaining that if their child was overweight or obese that they should go to their health care provider, some people were concerned that we would overwhelm the primary care physicians, which in my opinion that probably would have been a good outcome. We added a survey to health care providers throughout the state. We were flexible enough to inform some of the concerns emerging. That was the only one that we hadn't anticipated. We've also been flexible enough to implement a separate study, but part of the overall study, to do site visits with schools that showed over time increases in body mass proportions, for overweight or obese categories, that couldn't be accounted for by demographic changes within the student populations, versus those schools with a decrease in proportion overweight and obese categories. That has provided a rich source of data from on-site review of the school environment, interview with school personnel and examination of the surrounding environment" (Interview 2. Act 1220 Evaluation Project, October 2010).

Other contributions of the project's evaluation has been building relationships and raising awareness around childhood obesity with various audiences. Further description of how the evaluation influenced these areas is provided below.

"I think building relationships. I don't know about the rest of the country, but in Arkansas, relationships are really important. Our being able to build relationships between the College of Public Health, Department of Education, Department of Health, the Education Co-ops, which are the places out there that work with these difference school districts and things. I think building those relationships will have a long term lasting impact, long after this project ends. And that, I'm really happy about" (Interview 1. Act 1220 Evaluation Project, October 2010).

"I don't know that our evaluation has directly had an educational effect around the issues of obesity, but I think it has been part of the picture, just because we do an annual report of our data every year and we send it to all the superintendents and principals, and we report it to the legislators. I think we've helped to raise the awareness around some of these issues and

maybe helped people see what they can do. I mean, I'm in Public Health. I think that's important too" (Interview 1. Act 1220 Evaluation Project, October 2010).

Act 1220 Providing Guidance to Others

Methods, lessons learned, survey questions and experience of Act 1220's evaluation project have been shared with RWJF grantees and other projects requesting such information (Interview 1. Act 1220 Evaluation Project, October 2010). The project has also received queries from other states, legislators and evaluators (Interview 3. Act 1220 Evaluation Project, October 2010).

Working with other projects funded by the RWJF, and particularly those doing similar work, has provided a nice opportunity for networking, and the sharing of information and lessons learned.

> *"We've been able to work with them to share strategies and lessons learned. The three states have shared data collection protocols, instruments, questionnaires, and questions asked. Sometimes we pick up one or two of their questions and they pick up ours. We've been able to share methods and information. Sometimes they've adopted something as-is, sometimes they've modified them to meet their own unique state circumstances. It's been a nice opportunity to share" (Interview 3. Act 1220 Evaluation Project, October 2010).*

The project's evaluation has received feedback from others. "The feedback has been generally very positive…we have won an award…and we've been recognized nationally" (Interview 2. Act 1220 Evaluation Project, October 2010). "For the most part people are very appreciative of all the information that we are able to share and then other states and other projects, who want to do something similar, are always happy to have a pattern or lessons learned to go from" (Interview 1. Act 1220 Evaluation Project, October 2010).

The project has disseminated information about the evaluation through multiple formats, and to different audiences.

> *"There are a lot of people sitting there waiting for the next report to come out. We do many presentations, and in the scientific community, we do oral presentations, abstracts and manuscripts on a fairly regular basis. Two other groups have been particularly interested: The legislature, we present annually, or at least we used to, now just when they ask for it. The legislators have been interested in what we've found out and evaluation findings as we've gone along. The second group is the Child Health Advisory Committee, the state-level committee that the law established. They continue to meet and make recommendations. They look at our reports every year as valuable sources of information about where we are and where we need to go. Where there are pockets of less activity and where are the pockets of low hanging fruit, from a policy change perspective. What kind of supports do schools need? What are parents thinking, that sort of stuff. We present to the Child Health Advisory Committee annually. Those two groups are particularly interested in the findings" (Interview 3. Act 1220 Evaluation Project, October 2010).*

Challenges Encountered When Evaluating the Project

While the evaluation of Act 1220 has many strengths and have provided guidance to other projects, several areas of challenge were also identified. These include: the scope of the evaluation and related data collection and oversight, getting baseline data and maintaining flexibility. Additionally, several areas were described by project participants as a complication of evaluating a multi-level project, but could be considered challenges as well. These include: the lack of control within this natural experiment and fitting multiple pieces and sources of information together into a coherent whole. These are described in further detail for this project in the section titled "Benefits and Complications of Evaluating on Multiple Levels." Compared to other projects studied, this project in particular commented about how challenges were not too bad or overwhelming.

The size of the evaluation project, including data collection, oversight of subcontractors, and large amounts of data, were mentioned as challenges by individuals interviewed. "Dealing with the enormous amount of data that we have is a challenge. We could probably use two or three more analysts. But, you know, you do what you can do" (Interview 3. Act 1220 Evaluation Project, October 2010). Additionally, the unknown nature of the direction of future work, including recommendations of committees and addressing emerging areas of concern, made the evaluation even more challenging.

> *"We have challenges with data collection. Because we are trying to do it statewide, because we are doing it annually, and because we're interviewing parents, children, principals, superintendents and other major players, every year, that's just a lot of collection and then it's a lot of analysis and interpretation. So, it's just a lot of big pieces to handle, but to have a good picture of what's going on, we really felt like we needed all of those pieces" (Interview 1. Act 1220 Evaluation Project, October 2010).*

> *"I don't know that we had too many problems with the implementation of the evaluation itself. I guess one lesson learned is to make sure that you have careful monitoring of data collection, with benchmarks set, and timelines laid out…If not meeting the timeline, strategize about corrective action that can occur. There were times when we weren't meeting our timeline. Mostly when dealing with the subcontractor for surveys with parents and adolescents and we had to intervene with them and tried to ensure that we gave them appropriate feedback at the time" (Interview 2. Act 1220 Evaluation Project, October 2010).*

Other challenges identified include the collection of baseline data and maintaining flexibility. "It hasn't been insurmountable. Getting good baseline was a challenge because we had to move so quickly, to get into the field and get a baseline established before the policies really started taking effect and things really started happening" (Interview 3. Act 1220 Evaluation Project, October 2010).

> *"Staying flexible has been somewhat of a challenge, while at the same time, maintaining the continuity of the evaluation protocols over time. You don't want to be so flexible that you*

change your questions every year and you can't compare year two to year one, let alone year seven to year one, but on the other hand, if something isn't working you need to be able to fix it or if something is new, you need to be able to add it. Balancing the need to be consistent with the need to be flexible and able to be spontaneous about addressing things has been a little bit of a tension and a challenge" (Interview 3. Act 1220 Evaluation Project, October 2010).

Some other challenges, of a more general nature, were described. One included a statement about evaluation in general, and the funding and timing of when evaluators are often brought into projects and the influence this can have on the evaluation. The other comment described questions about the future of the project, once evaluation funding has concluded.

"…evaluations generally aren't well funded. Programs may or may not be well funded, but almost always it is difficult to find the funding for a good, thorough evaluation. We are always doing them on a shoestring; we are always doing them after the fact. People don't think about bringing in the evaluator as they are designing the program. They tend to bring them in after the program is funded…or when the reviewer says they need an evaluator...So, evaluators are kind of caught and can't do the kind of evaluation they would do if they were included up front. I think increasingly the federal government, and to a lesser extent state government, are really beginning to pay attention to the need to evaluate, to understand what's working and how and why and what's not working and how and why. So that we can be sure that our monies are well spent, but also so that successful programs are disseminated in ways so lots of people can take advantage of them. And programs that aren't successful are either made successful or gotten rid of. But, still, I think there is a dearth of trained, theoretically trained evaluators" (Interview 3. Act 1220 Evaluation Project, October 2010).

"We don't know how it's going to end. That's one of the things about it. Even when we end, even when our funding ends, it's not going to be ended. I don't know how that will look in two years. Call me back in two years and I'll tell you what it looks like…We are funded through August 2012. And then, I don't know if anybody is going to care anymore, or if we are going to have a wrap-up. I kind of don't know what will happen at that point…The policies are there in the schools and they will probably continue fairly the same for quite some time, but the evaluation funding is for a few more years and then I don't know what will happen after that" (Interview 1. Act 1220 Evaluation Project, October 2010).

What the Project Would Have Done Differently

Considering the complexity, amount of data collected, different audiences wanting information, and length of time working on this project, interestingly, the evaluation team did not have many things that they would have done differently for the project's evaluation. Individuals interviewed felt quite satisfied with how the evaluation was developed and carried out. Some of the statements of satisfaction included: "I don't know that we'd do

anything, if much differently" (Interview 2. Act 1220 Evaluation Project, October 2010) and "I don't know that I would have done anything much differently. I wish the crystal ball had been a little less fuzzy early on, but given what we had to work with, I think we did it fairly well" (Interview 3. Act 1220 Evaluation Project, October 2010).

The only area identified, for potential change, related to data collection and rewording of questions.

> *"We would probably from the beginning have gone about data collection a little differently, just the way that we have structured how we call parents and do the telephone interviews. Some things about how our surveys are actually set up. In hindsight you think "I wish that question was worded differently." Other than that though, I really don't think there's much that we would change. I think this is one of the best projects I've ever seen" (Interview 1. Act 1220 Evaluation Project, October 2010).*

Summary for Act 1220's Evaluation

The COPH carries out a multi-level evaluation of Act 1220 and its various components implemented at multiple levels throughout the state of Arkansas. Act 1220 aims to reduce childhood obesity statewide through interventions with schools, parents, and youth. A multi-level evaluation was developed by COPH evaluation team members, using a theory-based logic model, expertise of internal team members, and some guidance provided by the funder, the RWJF. The evaluation project was funded separate from the intervention activities, though a well-timed and quickly developed partnership between the COPH and RWJF. Specific elements of the evaluation, including the design, methods, and comprehensiveness were developed by the COPH, though feedback was received from RWJF as part of the application and negotiation process.

Benefits of evaluating Act 1220 with a multi-level evaluation were described by participants as: providing a more comprehensive evaluation, capturing data to better understand changes seen or missing, and having data to address concerns around unanticipated consequences.

Complications of carrying out a multi-level evaluation were identified by project representatives, including the lack of control and knowledge of the ultimate direction for the project and fitting multiple pieces, sources and types of data together into a coherent whole.

Sources of information having the most influence on the comprehensiveness of the evaluation were described by project participants as the project's logic model and guidance provided by the RWJF. Project representatives also identified additional sources of information as influential in guiding the development of the evaluation, including: the legislation itself, expertise of evaluation team members, access to legislators, and information from other fields.

Other contextual elements described by participating project staff as influencing the comprehensiveness of the evaluation included: resources, including funding from the RWJF and relationships built, including with schools and other partners; intended use by multiple users and for multiple purposes; flexibility of the evaluation; and working within the natural experimental context the project was situated. The most influential contextual element for

the project, described in all interviews, was the funding received by the RWJF, since the evaluation may not have been conducted otherwise, or at least not from early on in the intervention. Contextual factors described by Act 1220 Evaluation participants as having less influence on the evaluation include: the required evaluation elements by the funder, length of time carrying out the evaluation, evaluator access and capacity, and oversight and consequence by the funder.

Strengths of the project's evaluation were described by project participants as: using the initially developed logic model to guide the evaluation, incorporating multiple methods and types of data in the evaluation, flexibility to address recommendations by committees and emerging concerns, building relationships throughout the state, and providing guidance to other projects.

Challenges to the evaluation of Act 1220 were identified by participants as the scope of the project and amount of data collected, getting baseline data quickly as intervention was starting, maintaining flexibility to make changes as needed without compromising the evaluation, and more generally, the timing and funding of evaluations and the future direction of this project.

Flexibility in the evaluation was described as both a challenge and strength of the project by different respondents. One person touted the flexibility of the project in addressing issues as they arose, while the other described the tension between staying consistent to allow comparison of data and being flexible enough to make changes as appropriate.

In retrospect, project participants indicated that there was little they would do differently related to the evaluation.

It should be noted that all statements about elements influencing the evaluation are based on information provided by project representatives during interviews. Therefore, they are the ones attributing influence and describing the influence on their project and the evaluation.

Summary

Four projects were selected for case study, allowing for a more detailed analysis of their projects and evaluations than would otherwise have been possible through other methods of study. A multiple case study was conducted with four cases, two cases from programs focusing on tobacco control and two cases focusing on childhood obesity.

The current chapter presented descriptive information gathered through interviews and document reviews for each case studied. Detailed description of each project, their evaluation, and contextual elements identified by project representatives as influencing the development and comprehensiveness of their evaluation were presented.

CHAPTER 5: CONTEXTUAL ELEMENTS INFLUENCING COMPREHENSIVE EVALUATIONS OF MULTI-LEVEL INTERVENTIONS: LESSONS LEARNED FROM CASES IN TOBACCO CONTROL AND CHILDHOOD OBESITY

Introduction

For this dissertation a multiple case study was conducted with four programs evaluating multi-level interventions in tobacco control and childhood obesity. These cases were selected by the author to further understand contextual elements influencing their comprehensive evaluation of the multi-level intervention.

As stated previously, the research question of focus for the multiple case study is: *What are contextual elements influencing comprehensive evaluation of multi-level interventions in tobacco control and childhood obesity?*

Data were collected through interviews and document reviews with each case selected. Multiple respondents were included for each case and a compilation of their perspectives used as a collective representation for the case. Analysis within and across cases, and across fields, were also conducted with these data.

The previous chapter presented each case individually, to provide background on that specific project and to describe contextual elements identified by project representatives as influencing the comprehensiveness of their evaluation. The current chapter provides an analytic look at case data within and across the two fields. Figure 1 depicts the analysis of case study data, as conducted, in three phases. A further description for each stage is presented below Figure 1.

Figure 1. Case Study Analyses Conducted
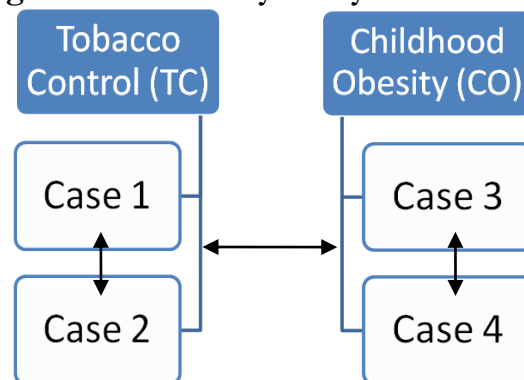
Phase 1: Analysis of data for each individual program/case. Therefore, four cases were analyzed individually, as depicted by the white boxes. These were presented individually for each case study in the previous chapter (Chapter 4).

Phase 2: Cross-case analysis within each field. Two cases were analyzed together within tobacco control and two cases within childhood obesity, as depicted by the arrows between the white boxes. These analyses are presented in this chapter.

Phase 3: Cross-field analysis comparing the fields of tobacco control and childhood obesity, as depicted by the arrow linking both sets of cases. This analysis is also presented in the current chapter.

This chapter first presents cross-case findings for tobacco control projects studied, followed by cross-case findings for childhood obesity projects studied. Next, cross-field findings are presented. Hypotheses developed early in the study are provided and compared against case data collected, to see whether hypotheses were supported or not. Cases across the two fields are compared to identify similarities and differences in strengths and challenges of their evaluations, sources of information used while developing the evaluations, and contextual elements identified by participants as influencing the comprehensiveness of the evaluations. Finally, all contextual elements from the original Framework for Comprehensive Evaluation of Multi-Level Interventions (developed in Chapter 3), are revisited and compared with data from cases studied.

All statements about elements influencing the evaluations are based on information provided by project representatives from cases during interviews. Therefore, they are the ones attributing influence and describing the impact of this influence on their project.

Cross-Case Findings for Tobacco Control Projects

This section presents a cross-case analysis of the two cases studied within the field of tobacco control, the California Tobacco Control Program (CTCP) and the New York Tobacco Control Program (NYTCP).

There were many similarities between the CTCP and NYTCP in the evaluations of their programs. Both cases had evaluations mandated as part of a legislated tobacco control program, to which the evaluation was linked. But, it was up to both programs to determine what a comprehensive evaluation of their program meant and add components as they deemed most appropriate to make their evaluation comprehensive.

Both programs have been held up as exemplar programs and evaluations, by the CDC and other groups. Their expertise has led these programs to provide guidance to other programs in tobacco control, and even other fields, including childhood obesity.

When developing their evaluations, both programs were guided by published literature and theory, experience and knowledge of external evaluation experts, and the legislation, as their primary sources of information. The NYTCP additionally used documents developed by the CDC, materials developed by other programs (including CTCP), and lessons learned from other programs to provide additional guidance for evaluation development. These materials were not yet developed when the CTCP developed its evaluation early on, and in fact, the CTCP was involved in helping to develop these materials for use by the CDC and programs such as the NYTCP. The CTCP also developed an overarching logic model, which provided guidance to the program for its evaluation. Access to, and guidance provided by, external evaluators were described by respondents from both cases as influencing the evaluation's development and comprehensiveness.

Funding for these programs was described as generous, by project representatives and in published documents, though budget cuts have had an impact on both programs in the last few years. Both CTCP and NYTCP considered funding as the most valuable resource to their programs and described funding as having influenced the comprehensiveness of their evaluations.

Both programs showed evidence of valuing, using, and facing evaluation, even when findings were negative, for the sake of improving the program. The intended use of evaluation and data were described by all informants from both cases as influencing the evaluation. Examples provided were of the intended use influencing the project's design and methods, and how evaluation findings are disseminated.

Political forces were also described by participants from both cases as influencing the program and evaluation, especially early on. The tobacco industry provided resistance and concerns were raised by politicians, and others, about tobacco taxes, smoke-free laws, and whether attempts to change social norms were a good use of funding.

The NYTCP is a much younger program than the CTCP. The NYTCP has been able to draw from resources developed and lessons learned by other programs, states and the field. In fact, specific mention was made in both program documents and interviews with the NYTCP, about the influence of the CTCP on their intervention and evaluation.

Each program has found a different way to successfully carry out their evaluations. The CTCP now conducts their evaluation internally, for the most part, while the NYTCP has theirs conducted by an external evaluation firm. Both programs did, however, rely on the expertise of external evaluation experts when developing their evaluations originally.

Both programs mentioned synergy, but neither program, even with their sophisticated and expensive evaluations, could identify the specific elements of the intervention that most impacted changes seen. It is hard methodologically to isolate the most influential elements under conditions of change. Another challenge identified by both programs was the complexity of the evaluation.

Additionally, both programs described a strength of their program's evaluation, as their ability to be flexible and adapt to changes over time in the program, the context within which the program sits, and modifying the evaluation as seems most appropriate.

Cross-Case Findings for Childhood Obesity Projects

Below is the cross-case analysis of the two cases studied within the area of childhood obesity, the Network for a Healthy California (Network) and Evaluation of the Arkansas Act 1220 of 2003 to Combat Childhood Obesity (Act 1220's Evaluation).

These two cases were less similar to each other than the tobacco control projects studied. Both the Network and Act 1220's Evaluation are conducted with internal evaluation team members, though the Network oversees implementation of their intervention activities as well. Act 1220's Evaluation is carried out by the University of Arkansas for Medical Science' College of Public Health, and is specific to only the evaluation. The legislated Act 1220 and intervention activities are carried out and overseen by others.

Neither case in this field mentioned being influenced by each other, but they were not directly asked this question either.

When developing their evaluations, both projects were guided by published literature and theory, experience and knowledge of internal evaluators, and other fields as their primary sources of information. Several Network participants described their project being influenced greatly by the California Tobacco Control Program, which is a part of their same organization, the California Department of Public Health, for additional evaluation development guidance. Act 1220's Evaluation was additionally guided by an external evaluation expert and the logic model they developed.

Funding for these projects is described as generous, both through interviews and published literature, compared to many other projects. Funding was considered one of the most valuable resources by both the Network and Act 1220's Evaluation representatives. Funding was also described as influencing the comprehensiveness of both evaluations by project participants.

Other contextual elements influencing the comprehensiveness of the evaluations for both projects were identified by participants as: experience and knowledge of internal evaluators, access to and advice from evaluation experts beyond the project, and an overarching logic model or framework. Both cases used either their funder's criteria or the legislation to guide some elements of the evaluation, but it was up to both projects to determine what a comprehensive evaluation meant and add components they deemed most appropriate to make their evaluation comprehensive. Comprehensiveness of the Network's evaluation was impacted positively by their proximity to the California Tobacco Control Program and negatively by restrictions imposed on them by their funder.

Both projects have been held up as positive examples by other groups, such as funders, and their expertise has led them to provide guidance and materials to others.

The intended use of evaluation and data were described by respondents from both programs as greatly influencing the evaluation, including the design, methods, questions asked and evaluation modifications.

Both projects identified as a challenge the scope and complexity of the evaluation and bringing together various pieces into a coherent whole. A challenge unique to the Network involves restrictions by its funder. Restrictions placed on the Network by the USDA has impacted the levels that they intervene and evaluate, evaluation comparisons that can be made, and wording used when reporting to USDA and in media campaigns, such as to not disparage any foods. Working with projects that have alternative funding has allowed some of this work, which the Network considers important, to be carried out, but it also means that the Network is unable to claim credit for some of the impact that has been made.

Both programs have described as a strength of their program's evaluation their strong internal capacity for carrying out evaluation, providing guidance to others, and their funding. Representatives from both the Network and Act 1220's Evaluation also described the importance of incorporating multiple methods and collecting different types of data in the evaluation. Act 1220's Evaluation team members described their use of a theory-based logic model as a key element to the success of their project and its evaluation. Additionally, this

project considers their ability to be flexible and make modifications as needed to address emerging issues important to their evaluation's success.

Cross-Field Findings: Comparing Tobacco Control and Childhood Obesity Cases

This section presents cross-field findings from all cases in tobacco control and childhood obesity. Hypotheses developed early in the project, before data were collected, are compared to data from the cases. Next, findings from case data are presented as three sets of tables and corresponding text to provide further description. The first section describes strengths and challenges to evaluation for each case studied. The second section covers sources of information described by project representatives as guiding the development of evaluation for each case. The third section presents contextual elements identified by project representatives as influencing the comprehensiveness of evaluations for each case. Finally, additional considerations, comparing the two fields, are presented.

Early Hypotheses Compared to Case Data

Early in the study, hypotheses were developed by the author about how certain contextual elements might influence the comprehensiveness of evaluations of multi-level interventions. A combination of early literature reviews and these hypotheses aided in the development of the Framework presented in Chapter 3. A comparison is made below with these early hypotheses and case data collected, to see whether they were supported or not.

Hypothesis: All contextual elements described in the Framework influence the comprehensiveness of evaluations for multi-level interventions.

This was not found to be supported in its entirety. All contextual elements were described as influencing the evaluations for each project, but several contextual elements were not necessarily influential to the *comprehensiveness* of the evaluation. Certain contextual elements were described by project representatives as influencing their evaluation's comprehensiveness, but as learned through the course of studying these projects and the various contextual elements, there was more nuance in the role of these elements than originally expected. Certain contextual elements were described as influencing the development of the evaluation, but distinctly different than influencing the comprehensiveness. Patterns for the role of various contextual factors were the same in both fields. Further detail about which contextual elements were described by cases as influencing comprehensiveness versus development of the evaluation is displayed in Figure 2 and described in the corresponding text discussion.

Hypothesis: Requirement of evaluation by project funder influences comprehensiveness of the evaluation, regardless of field.

Beyond looking at the influence on comprehensiveness for all contextual elements, as done through the previous hypothesis, one contextual element was proposed to have specific influence on evaluation comprehensiveness. It was thought that of the contextual elements, this one might particularly influence evaluation comprehensiveness, if for no other

reason than it was required of the projects. This was not supported, the way expected, by data collected from cases through participant interviews. Requirements by funders or criteria laid out by legislation were described for all cases as influential in the development of the evaluation, to meet necessary mandates, but that seems about as far as those requirements reach. The comprehensiveness of the evaluations were described by all cases as something the project determined themselves (or with the help of external evaluation experts) and made a reality. Requirements by funders and legislative criteria seems to have set a minimum for the evaluation, but their projects developed more complex and comprehensive evaluations beyond what was required of them. Therefore, it was not the requirement set out by the funder that determined, or even influenced, the comprehensiveness of the evaluation for any of the cases. This pattern was found for all cases and across both fields.

The hypotheses presented above were not supported, in their entirety, by case data collected. So, what did the data show? What did project representatives describe as influential in the development of their evaluations? What contextual elements were identified by participants as specifically influencing the comprehensiveness of their evaluations? How did these differ across the fields of tobacco control and childhood obesity? Further detail about findings from case data across both fields are presented below, including strengths and challenges to evaluations, sources of information described by project representatives as guiding the development of evaluations, and contextual elements identified as influencing the comprehensiveness of evaluations for each case. As previously mentioned, assertions about elements influencing evaluations are based on statements made by project representatives from cases during interviews. Therefore, it is the interviewees that are attributing influence and describing the impact of this influence on their project.

Strengths and Challenges to Evaluations
Table 1 highlights strengths and challenges to evaluating the project, as described by each case studied. These are presented together, to enable easier comparison between the four programs.

As can be seen in Table 1, representatives from the four projects identified a variety of strengths and challenges to their project's evaluation. Unique strengths and challenges were naturally reported by cases within each field as well, but commonalities across the cases and fields are the primary focus of information reported here.

Participants from cases in both fields identified challenges with the scope and complexity of the evaluation and mention was made by participants from three cases about the inability to capture synergy and determine which elements of the intervention most impacted changes seen. Respondents from each field also described challenges with bringing various pieces of the evaluation together into a coherent whole. Large amounts of data, reporting, and struggles with how to make the best use of data available and how to best communicate findings were also described by respondents from both fields. While one case's restriction by their funder influenced their evaluation and what they can "claim" as their impact, it likely does bring up a shared issue with claiming responsibility, which projects of this nature seem to face. When working on multiple levels, partnering with different local

projects, organizations and community members, the question of "who gets credit?" certainly arises, especially if there are multiple funders, decision-makers and entities involved in implementing the program (and reporting back to their funders). This may be linked to the challenge of measuring synergy mentioned above.

A strength of the evaluation identified by participants from three cases, and spanning both fields, was having the flexibility to modify the evaluation as needed, to update it, refine or broaden it, or to address emerging issues. While not directly described by the remaining case, representatives from this project shared creative ways they are able to accomplish work they feel is important, while still respecting the guidelines of their funder. This has taken flexibility and adaptation on their part. Project participants from cases in both fields also described using a logic model to guide their evaluation and decisions along the way and make the evaluation more comprehensive. Project representatives for three cases, covering both fields, described a strength of their evaluation as having strong internal capacity for carrying out the evaluation. Though participants from the fourth case did not directly identify internal capacity as a strength, description was provided about how strong their internal evaluation knowledge was, especially considering that they have evaluations conducted externally. Additionally, all cases studied were described as examples put forward by others. All projects studied have shared their expertise by provided guidance to others, and sharing materials and lessons learned.

Challenges reported by project representatives seem, to the author, to center around the scope of the projects, methodological challenges of capturing synergy and putting various pieces into a coherent whole, and communicating findings to various audiences in the most effective ways. Strengths reported by project representatives appear, to the author, to link to things these projects "had," including internal evaluation capacity, a guiding logic model, and flexibility to modify evaluations as needed.

Table 1. Strengths and Challenges to Evaluation for Each Case

Case	Strengths	Challenges
CTCP	• **Building Internal Evaluation Capacity** • **Having Access to/ Consulting with Evaluation Experts** • **Valuing Evaluation** • **Adaptability** • **Providing Guidance and Tools to Other Projects** • **Collecting Baseline Data** • **Having Adequate Funding** • **Using a Logic Model to Frame the Evaluation**	• **Complexity of the Program** • **Linking Diverse Work of Local Programs to Whole of Program** • **Building Internal Capacity** • **Expectations Based on Funding Available** • **Lack of Value for Evaluation** • **Administrative Difficulties** • **Not Knowing the Next "Big Thing"**

Case	Strengths	Challenges
NYTCP	• **Flexibility** • **Media Evaluation** • **Active Role of Evaluation in Improving Program** • **Witnessing Norm Change** • **Adequate Funding (until recently)** • **Funding for/ Access to High Quality External Evaluators**	• **Budget Reductions** • **Bureaucratic Sensitivity/ Slowness** • **Evaluation Complexity** • **Reporting and Dissemination**
Network	• **Comprehensive Surveillance Data** • **Strong Evaluations of Interventions Before Roll-out** • **Building Internal Evaluation Capacity** • **Case Studies** • **Media Evaluation** • **Consulting with Evaluation Experts** • **Evaluating Program of Large Scope** • **Providing Guidance and Tools to Other Projects** • **Having Adequate Funding**	• **Restrictions by Funder** • **Unable to Get "Credit" for Impact Made** • **Interest in Areas/ Data Not Supported by Funder** • **New Data Requirements and Reporting by Funder** • **Scope and Complexity of the Program** • **Bringing Pieces Together** • **Diversity of Work by Local Projects** • **Measurement Issues** • **Expectations by Different Stakeholders**
Act 1220 Evaluation	• **Theory-based Logic Model to Guide Evaluation** • **Use of Mixed Methods in Data Collection** • **Flexibility to Modify Evaluation to Address Emerging Issues and Concerns** • **Lasting Relationships Built** • **Providing Guidance to Other Projects**	• **Scope of Evaluation** • **Amount of Data Collected** • **Fitting Multiple Pieces/ Sources of Data into Coherent Whole** • **Flexibility of Evaluation** • **Baseline Data with Short Timeframe** • **Lack of Control/ Natural Experiment Setting**

Table 2 presents information on the sources of information guiding development of the evaluation for each case. A check mark indicates that this source of information was described by project representatives as guiding the development of the evaluation for the case represented. Boxes left blank, or without a check mark, indicate that this source of information was not identified by respondents as guiding the development of that case's evaluation.

Table 2: Sources of Information Guiding Development of the Evaluation for Each Case

Source of Information Guiding Development of Evaluation	CTCP	NYTCP	Network	Act 1220
✓ = this element was described as <u>guiding development</u> of the evaluation for the case; left blank= element was <u>not identified as</u> <u>guiding development</u> of the evaluation for the case.				
Published Literature/ Theory	✓	✓	✓	✓
Documents (i.e.: from CDC)		✓		
Experience/ Knowledge of Internal Evaluator			✓	✓
Experience/ Knowledge of External Evaluator	✓	✓		✓
Overarching Logic Model/Framework Developed by Evaluator	✓			✓
Legislation or Funder Criteria	✓	✓	✓	✓
Other Departments within Organization			✓	
Other Programs in Same Field (TC/CO)		✓	✓	
Other Fields Beyond Own (TC/CO)	✓		✓	✓

When developing their evaluations, representatives from most cases, covering both fields, described being guided by published literature and theory, experience and knowledge of external evaluation experts, legislation/funder criteria, and other fields, as primary sources of information.

Access to, and guidance provided by, external evaluators were described by participants from cases in both fields as influencing the evaluation's development and comprehensiveness. Case representatives from the two childhood obesity projects described using the experience and knowledge of internal evaluators as a source of information when developing their evaluations. This was not identified as a source of information guiding development of the evaluation by either tobacco control project.

Project members from a case in each field described the development of their evaluation as being guided by their program's logic model. Also, the CTCP was identified as a source of information for evaluation development for the NYTCP and Network.

Beyond using the funder or legislative criteria to guide development of the evaluation, other elements described by most projects, across both fields, came from sources external to their project—published literature, external experts, and resources from other fields.

Contextual Elements Identified by Cases as Influencing Comprehensiveness of Evaluations

Table 3 presents contextual elements identified by project participants as influencing the comprehensiveness of the evaluations for all cases studied. A check mark indicates that this contextual element was described by project representatives as influencing the evaluation's comprehensiveness for that case. Boxes left blank, or without a check mark, indicate that this element was not identified by participants as influencing comprehensiveness for that case.

As seen in Table 3, many of the same contextual elements were identified by project representatives for cases in both fields as influencing the comprehensiveness of evaluations. Participants from all cases described their funding as having influence on the comprehensiveness of their project's evaluation.

While representatives from all cases had previously identified that the legislation or funder criteria was used in the development of their evaluation, no project participants described this criteria as influencing the comprehensiveness of their project's evaluation. It was up to the projects to operationalize what a comprehensive evaluation meant and guide themselves in developing this comprehensive evaluation. Participants from all projects described starting with the elements needed from the legislation or funder and adding their own elements to make the evaluation comprehensive.

Participants from at least one case in each field identified the following contextual elements as influencing the comprehensiveness of their project's evaluation: experience and knowledge of internal evaluators, access and advice from evaluation experts beyond the project, an overarching logic model or framework, and other fields beyond their own.

Only representatives from cases in tobacco control additionally described the experience and knowledge of external evaluators as influencing the comprehensiveness of their evaluation. Participants from both tobacco control projects identified this factor.

Each program has found a different way to successfully carry out their comprehensive multi-level evaluations. The CTCP, Network and Act 1220's Evaluation are all conducted internally currently; while the NYTCP continues to have theirs conducted by an external evaluation firm.

All programs have been held up as example programs and evaluations, by the CDC, their funders, and other groups. Their expertise has led these programs to provide guidance to other programs in tobacco control, childhood obesity, and even other fields.

Table 3: Contextual Elements Influencing Comprehensiveness of Evaluation for Each Case

Contextual Element	CTCP	NYTCP	Network	Act 1220
✓ = this element <u>described as influencing</u> the <u>comprehensiveness</u> of the case's evaluation; left blank= this element <u>not identified</u> as influencing the <u>comprehensiveness</u> of the case's evaluation.				
Funding Available	✓	✓	✓	✓
Experience/ Knowledge of Internal Evaluator	✓		✓	✓
Experience/ Knowledge of External Evaluator	✓	✓		
Accessibility/ Advice of Evaluation Experts Beyond the Project	✓		✓	✓
Overarching Logic Model/Framework	✓		✓	✓
Evaluation Requirements by Legislation/ Funder				
Beyond Requirements, Project Added Additional Elements to Make Evaluation Comprehensive	✓	✓	✓	✓
Project Operationalized and Guided Self in Developing a Comprehensive Evaluation	✓	✓	✓	✓
Other Fields Beyond Own (TC/CO)	✓		✓	
Length of Time Carrying Out Project	✓			
Restrictions by Funder			✓	

Contextual elements identified by project representatives and presented in Table 3 appear, to the author, to cluster around themes of external factors, knowledge factors and one other factor, which does not categorize well. External factors, including requirements by the legislation, restrictions by the funder and the project's funding, had mixed influence, as described by participants, on the comprehensiveness of their project's evaluation. Requirement of the legislation was not identified by any case as influencing evaluation

comprehensiveness and restriction by the funder was identified by only one case. Funding for the project was the only of these external factors identified by all cases as influencing their evaluation's comprehensiveness. Therefore, external constraints, such as restrictions, do not appear to have influenced the comprehensiveness for these cases overall, but having positive external factors, such as funding available, did influence the comprehensiveness. Knowledge factors identified by cases include experience and knowledge of internal and external evaluators, access to evaluation experts beyond the project, fields beyond their own, developing a logic model, and project members adding in their own components and operationalizing comprehensiveness themselves. These elements all relate to knowledge and expertise within the project, or access to experts to work with the project on their evaluation. These knowledge factors were identified by most cases, and across both fields, as influencing the comprehensiveness, with the exception of experience and knowledge of external evaluators and fields beyond their own, which were identified by two cases. The only element that did not seem to categorize easily was length of time carrying out the project, which was identified by only one project as influencing the comprehensiveness of their evaluation.

Additional Considerations

Intended use of the evaluation was described by participants from all cases as influential to the evaluation, but it is unclear whether this included the comprehensiveness as well. Project participants from all cases, and across both fields, described how the intended use of evaluation and data greatly influenced decisions made about the evaluation, its design and methods, elements to include or not, and how evaluation findings are reported and disseminated.

Flexibility is an element that was described by participants as greatly influencing the success of the evaluation for most of the cases, but had not originally been asked about. Three of the four cases, spanning both fields, described the influence of flexibility to their project's evaluation. Participants from these three cases described having the flexibility to modify the evaluation as needed, to update it, refine or broaden it, or to address emerging issues as a strength of their project and contributing to the evaluation's success. The remaining project described creative ways they are able to accomplish work they feel is important, while still respecting the guidelines of their funder. This has taken flexibility and adaptation on their part.

Oversight and accountability to funders and decision-makers were identified by participants as influencing the evaluation for cases in both fields, but not necessarily the comprehensiveness of these evaluations. Concern of potential consequences by funders or decision-makers was described as influential earlier in the program, but not very impactful at this stage, for most cases.

Other comparisons between the fields of tobacco control and childhood obesity were made while case data were being collected. Some sources of information available for one field do not seem developed or readily available for the other. One individual pointed out that materials, such as those created by the CDC, which have guided work in tobacco control are not available for projects working in childhood obesity. CDC representatives also

confirmed, at a recent American Evaluation Association conference, that resources to guide childhood obesity are not as readily available or developed, as they are for tobacco control, and that much less is known about "what works" in childhood obesity (Compton, Baizerman, & Farris, 2010; Farris, Khan, & Jernigan, 2010). It may be that these materials are being developed and will be available in the future, as childhood obesity is an area with great interest and support currently, but it has not been the focus of interest for as long as tobacco control has previously. One of the tobacco control cases studied has been actively intervening and evaluating their project for over twenty years and contributed to materials readily used by others in tobacco control and other fields, including childhood obesity. The childhood obesity projects studied have been carrying out work for a while, but still only about seven years.

Support for work being done and duplicating work by successful projects, even from other fields, can be great assets to various projects. The large majority of voters do not smoke, so policies around tobacco control may not be as controversial as something that might affect a larger proportion of the population, such as those consuming sugar-sweetened beverages. This may not have been the case in the past, but tobacco control policies seem less controversial now than just five or so years ago. Additionally, media that has worked for one project, such as the CTCP, may not be used by another, such as the Network. This is not based on the content of the field, but rather restrictions by the project's funding source. Edgy media and media attacking the tobacco industry has found great success in both tobacco control cases studied, but such media tactics are not allowed by the Network's funder. This project has strict guidelines about not disparaging any foods or attacking the food industry. A disconnect can arise when the Network's funder, the USDA, is responsible for ensuring the success of the nation's food industry, while the Network seeks to positively impact public health.

Linking Back to Contextual Elements from the Original Framework

The Framework for Comprehensive Evaluation of Multi-Level Interventions developed previously (Figure 1, Chapter 3) describes specific contextual elements proposed to influence the comprehensiveness of multi-level evaluations. Contextual elements identified in this Framework were: Resources, Evaluation Requirement by Funder, Training/ Experience of Evaluator, Length of Time Carrying Out Type of Intervention/ Evaluation, Intended Use of Evaluation, Access to Information to Guide the Evaluation, Evaluator Access, Evaluation Capacity of Project/Organization, and Agency Support. Broader contextual elements also identified in this Framework, which may also influence comprehensiveness, included: Social/ Environmental Conditions, Politics, and Competing Organizations/ Interests.
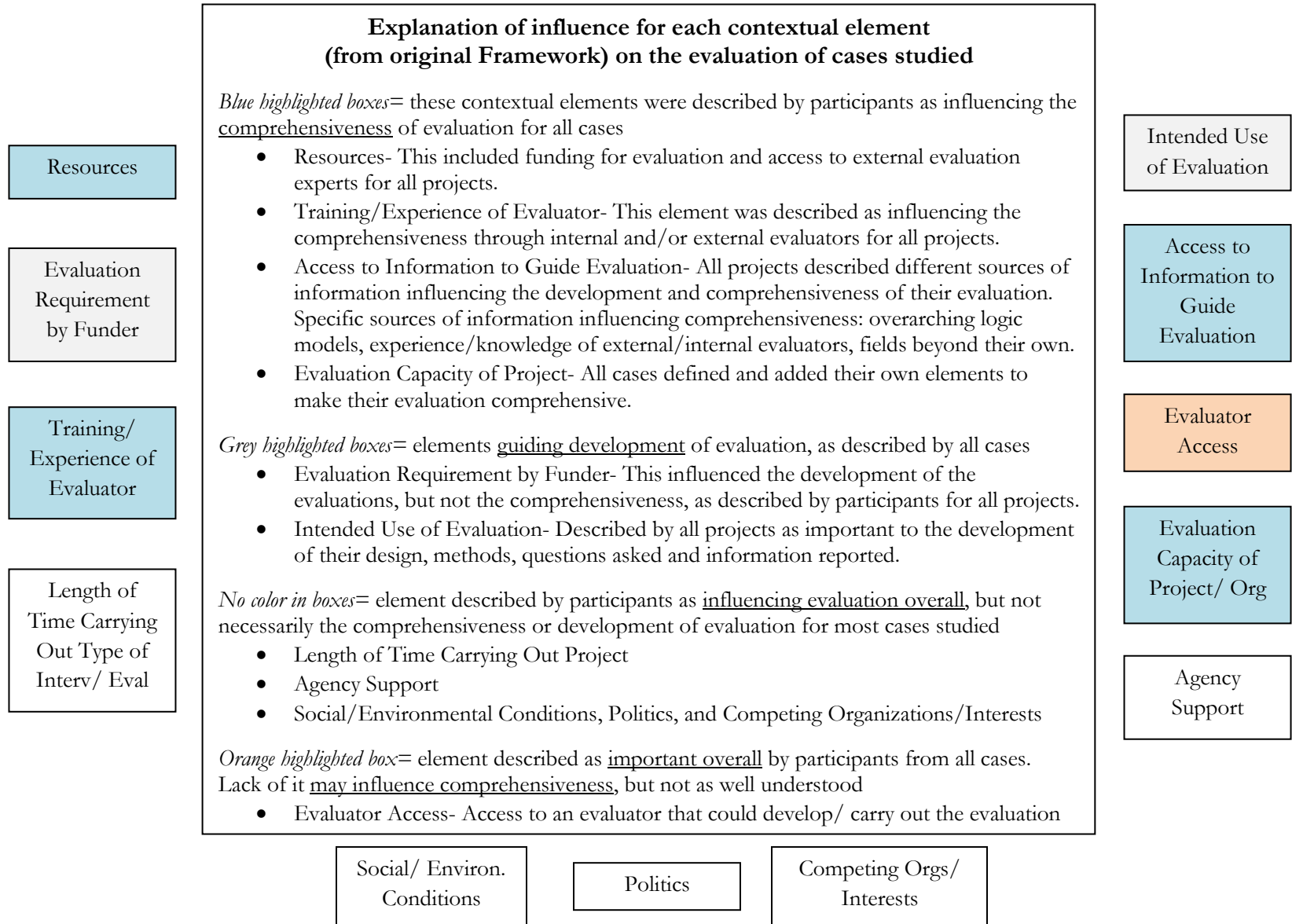
The multiple case study in this dissertation, carried out with four projects evaluating multi-level interventions, sought to provide information on how contextual elements have influenced these evaluations. Cases were studied to better understand the influence of contextual elements on the evaluation overall, and more specifically, elements influencing the

development of these evaluations, comprehensiveness of these evaluations, and other contextual elements of importance.

From case data collected, all of the contextual elements identified in the Framework were described by project participants as having influence, to at least some extent, on the evaluations of cases studied; but only some of the contextual elements were specifically identified by participants as influencing the *comprehensiveness* of evaluations for projects studied. All contextual elements are presented, compared to case data as a whole, together in Figure 2, and for each element individually in the text following the Figure. Information about how these contextual elements were identified by project representatives as influencing each case individually is described in the cross-case and cross-field analyses earlier in this chapter.

Presented next is Figure 2, which displays all contextual elements (from the original Framework for Comprehensive Evaluation of Multi-Level Interventions) and categorizes them based on findings from the cases studied. Contextual elements described by interview participants, as how they influenced the evaluation, were fairly consistent across cases and fields. When comparing case data with the Framework, several categories emerged of how contextual elements were described as influencing the evaluations of projects studied. These categories, developed by the author, include contextual elements: 1) influencing the comprehensiveness of the case evaluations, 2) influencing development of the evaluations (but not the comprehensiveness), or 3) influencing the evaluation overall, but not specifically the development or comprehensiveness. One element, evaluator access, was described by project representatives from all cases as important overall, and that a lack of evaluator access would likely influence evaluation comprehensiveness. But since none of the projects studied had a challenge with this, and its influence was not described in enough detail, actual influence was difficult to assess. Included in the center of this Figure is a brief description of the specific contextual elements within each of these categories. To be included in one of the categories above, most or all cases needed to fit the pattern. Therefore, three cases (covering both fields) or all four cases needed to have project representatives identify this contextual element as influencing the comprehensiveness, the development, or the evaluation overall.

Figure 2: Contextual Elements Influencing Evaluation of Multi-Level Interventions for Cases Studied

Resources	
Evaluation Requirement by Funder	
Training/ Experience of Evaluator	
Length of Time Carrying Out Type of Interv/ Eval	

Explanation of influence for each contextual element (from original Framework) on the evaluation of cases studied

Blue highlighted boxes= these contextual elements were described by participants as influencing the comprehensiveness of evaluation for all cases
- Resources- This included funding for evaluation and access to external evaluation experts for all projects.
- Training/Experience of Evaluator- This element was described as influencing the comprehensiveness through internal and/or external evaluators for all projects.
- Access to Information to Guide Evaluation- All projects described different sources of information influencing the development and comprehensiveness of their evaluation. Specific sources of information influencing comprehensiveness: overarching logic models, experience/knowledge of external/internal evaluators, fields beyond their own.
- Evaluation Capacity of Project- All cases defined and added their own elements to make their evaluation comprehensive.

Grey highlighted boxes= elements guiding development of evaluation, as described by all cases
- Evaluation Requirement by Funder- This influenced the development of the evaluations, but not the comprehensiveness, as described by participants for all projects.
- Intended Use of Evaluation- Described by all projects as important to the development of their design, methods, questions asked and information reported.

No color in boxes= element described by participants as influencing evaluation overall, but not necessarily the comprehensiveness or development of evaluation for most cases studied
- Length of Time Carrying Out Project
- Agency Support
- Social/Environmental Conditions, Politics, and Competing Organizations/Interests

Orange highlighted box= element described as important overall by participants from all cases. Lack of it may influence comprehensiveness, but not as well understood
- Evaluator Access- Access to an evaluator that could develop/ carry out the evaluation

Intended Use of Evaluation
Access to Information to Guide Evaluation
Evaluator Access
Evaluation Capacity of Project/ Org
Agency Support

Social/ Environ. Conditions	Politics	Competing Orgs/ Interests

155

Contextual elements from Figure 2 are presented below individually, to provide additional detail than in the Figure.

Resources

This contextual element was described by project participants as influencing evaluations overall, as well as, the comprehensiveness of evaluations. The primary resource indicated as influencing the comprehensiveness was the project's funding. Consistency of funding and length of time funded were also described by projects. Informants from all cases described being well funded compared to other projects and informants from all projects believed that their level of funding had influence on how comprehensive of an evaluation they were able to conduct.

Another resource described by several cases was access to, and advice by, external evaluation experts. This advice was described by participants as both guiding the development of the evaluation, as well as, influencing the comprehensiveness of the evaluation for most of the cases.

Evaluation Requirement by Funder

This element was interesting. Both tobacco control cases studied were funded through legislation, which required that evaluation be conducted. For the childhood obesity projects, one was funded through the government, the USDA, and the other through a Foundation, the RWJF. All but the project funded through the RWJF had at least some criteria they needed to fulfill to meet funder requirements, and these were used to guide some elements of the evaluation during development. The RWJF project was evaluating a legislated Act, and used criteria from the Act to guide the development of the evaluation, but it was not technically required of them through their funder. Therefore, all projects used the legislation or funder criteria as a source of information while developing the evaluation, but none of the project representatives described that these requirements had influence on the comprehensiveness of the evaluation. In other words, evaluation requirements by the legislation or funders provided a starting off point, or a minimum, but projects went well beyond these requirements to develop their comprehensive evaluations.

It was up to the projects to operationalize what a comprehensive evaluation meant and guide themselves in developing this comprehensive evaluation. Projects started with the elements required of them from the legislation and funders and added their own elements to make the evaluation comprehensive.

Training/ Experience of Evaluator

Experience and knowledge of both internal and external evaluators were described by project participants as very influential to evaluations overall, as a source of information during the development of the evaluations, and in influencing the comprehensiveness of evaluations. This element of experience and expertise of evaluators was identified by participants of all cases as influencing the evaluation and its comprehensiveness.

Projects that conduct their evaluations internally also described utilizing outside sources of evaluation expertise for feedback and guidance.

<u>Length of Time Carrying Out Type of Intervention/ Evaluation</u>

This contextual element was described by project participants to have influenced evaluations overall, somewhat, but no project participants described this element as influencing the comprehensiveness of their project's evaluation. The main description provided related to learning from mistakes made and making changes to the evaluation, though this seemed more linked to a project's adaptability than just the length of time carrying out the intervention or evaluation. All projects have been evaluating multi-level interventions for quite some time, ranging from about 7 to over 20 years. It could be argued that all of these projects had been well enough established that length of time was no longer a contextual factor influencing their evaluation. If a case had been selected within the first year or two, they may have provided different information.

Another area where length of time carrying out the project was described as seeing change involved political sensitivity and oversight for tobacco control projects. In earlier years, there was a lot of political heat and scrutiny over work being done by these projects. But, as one project described, legislators can now pass a tobacco tax without consulting them because it is not the sensitive topic it once was. It is now considered a safe thing to do and an expected norm.

<u>Intended Use of Evaluation</u>

Participants from all four cases described this element as influential to the evaluation, but it is unclear whether this included the comprehensiveness as well. Examples were provided about how the intended use of evaluation and data greatly influenced decisions made about the evaluation, its design and methods, elements to include or not, and how evaluation findings are reported and disseminated. Clearly intended use was an important factor in the evaluation for these cases.

<u>Access to Information to Guide the Evaluation</u>

The author asked case participants to describe sources of information that were influential in the development of their evaluation and the comprehensiveness of their evaluation. Representatives from all projects described access to, and use of, multiple sources of information to guide their evaluations.

Sources of information guiding the development of evaluations included: published literature and theory, document, such as those made available by the CDC, experience and knowledge of both internal and external evaluators, overarching logic models and frameworks, legislation or funder criteria, other departments within their organization, other programs within their same field, and other fields beyond their own. Most or all cases used published literature and theory, experience and knowledge of external evaluators, legislation or funder criteria, and other fields beyond their own in the development of their evaluation.

Sources of information described as influencing the comprehensiveness of evaluations included: overarching logic models and frameworks, experience and knowledge of external and internal evaluators, and other fields beyond their own.

Evaluator Access

Both access and capacity of evaluators were often described together by interview participants. Most individuals interviewed overwhelmingly described a lack of evaluators able to conduct complex evaluations in general, both in the access to evaluators and their capacity to carry out such evaluations. Examples were provided by participants of projects not being able to find evaluators, working with evaluators that did not meet the project's needs or expectations, people calling themselves evaluators, when they really should not, and people not adequately trained or experienced to conduct this work, but put in a situation to be doing it anyway.

The cases studied, however, were able to have developed a team of evaluators, through use of either internal staff or external contractors that could carry out their multi-level evaluation. Evaluators were found, with capacity to carry out such work, or capacity was built internally, when needed. Therefore, evaluator access and capacity were described by participants as an issue for the field overall, and may influence developing and carrying out these types of evaluations, but these elements were not reported as directly influencing the evaluations for cases studied.

Having access to evaluators with a high capacity for conducting multi-level evaluations has likely influenced both the evaluation and comprehensiveness of evaluations for these cases. As described above, experience and knowledge of evaluators, both internal and external, were identified by participants as guiding the development of evaluations and influencing the comprehensiveness for evaluations of all projects studied.

Evaluation Capacity of Project/Organization

As described above, evaluator access and capacity were closely linked in responses provided by projects. Access to evaluators with adequate capacity was a challenge that several of the projects dealt with in the past, and influenced how their evaluations and evaluation teams are set up currently. Poor experiences with external evaluation contractors provided more than one of the cases with the incentive needed to build their own evaluation capacity internally. All projects were described as having internal evaluation capacity, to some extent or another, even when contracting with external evaluators to carry out the bulk of evaluation work for the project. Three of the four cases studied have most or all of their evaluation work conducted internally at this time, and have built their capacity as needed to have a strong evaluation team in place currently. A key to building a comprehensive evaluation included the project taking the lead on operationalizing what a comprehensive evaluation meant and included elements to carry this out. Having evaluation capacity within the project is likely needed for this to have happened, which it did in all four cases studied.

Agency Support

Most projects appear to have the support of their agency or organization in carrying out their evaluations, but some challenges were described. Project participants from several

cases noted that they receive positive feedback about work they do, are supported financially and otherwise, and are able to proceed with their work without a lot of organizational challenge. Representatives from one case described their agency as challenging to their evaluation, in the sense that recruiting and retaining staff had been an ongoing barrier. A participant from another case indicated that changes within their organization can determine whether there is much attention being paid or criticism being directed toward their project. Some of these organizational support issues relate to the project as a whole, or the intervention, as much, if not more than the evaluation. Restrictions by the funder, potentially more than the home organization, have had negative influence on the comprehensiveness of the evaluation for at least one of the cases studied.

<u>Social/ Environmental Conditions, Politics, and Competing Organizations/ Interests</u>

Broader contextual elements which may also influence evaluation comprehensiveness were investigated, such as social and environmental conditions, politics and competing organizations and interests. These elements were not explored in as much detail as the other contextual elements above and seem to be more of a mixed bag than clearer patterns found for other contextual elements. Challenges around politics, environmental factors and competing interests were identified by participants from several of the cases. These elements were described as influencing their evaluations and sometimes even the comprehensiveness of the evaluation. Political sensitivity was brought up by participants representing both of the tobacco control projects, as important early on, but less so as time has gone by. In another example, restrictions by a project's funder do not allow work or evaluation funded on several levels that the project considers important for helping to change the environment and impact social norms. This project must work hard to develop relationships with projects funded by other sources and link partners and resources together to accomplish work they consider important. Restrictions by this funder and organizational pressures make the project work much harder and more strategically to meet the expectations of the various parties involved.

Limitations

Several limitations exist to the study of cases for this dissertation. Only four cases were studied within two different fields. The inclusion of additional cases could have provided differing information than that found with the cases included for study currently. The analysis was limited based on the small number of cases included and the intentions for the conclusions drawn. A more rigorous design, with a stronger theory-based study, would have enabled stronger conclusions to appropriately be drawn than would make sense for the current study.

Additionally, purposeful sampling was used for both the selection of cases and selection of interview participants within each case studied. Purposeful sampling of cases and interview participants reduces the opportunity to statistically generalize, but as emphasized previously, that was not the intent of this study. Including additional interview participants

or interviewing everyone associated with the projects and evaluations may have provided information beyond what was gathered and presented here.

Lastly, case study findings are based on interviews and document reviews with the four projects studied. Therefore, most information is based on self-reported information about the project, its influences and decisions made, and the evaluation. With self-report and interviews, there can be concern that information reported may have been incorrectly described, misunderstood or provided differently based on concerns about perceived response or later consequences for information shared. All informants were treated as having equal weight and expertise to speak about the project when collecting and analyzing data. Based on constraints of Committee for Protection of Human Subjects approval and agreements with the projects participating, information about interview participants is provided to add some detail about who the informants were, without providing enough detail to identify anyone individually within the project. Member checks were conducted to ensure that information presented is as intended by participants and does not report data of a sensitive nature.

Conclusions

Four cases were selected for detailed study, allowing a further understanding of the projects, their evaluations, and the contextual elements influencing the comprehensiveness of their evaluations than would otherwise have been possible through other methods of study. A multiple case study was conducted with two programs focusing on tobacco control and two programs focusing on childhood obesity. Data were collected by the author for these cases through interviews and a review of documents for each project.

Chapter 4 provided a descriptive presentation of each case individually, providing rich detail about each project and contextual elements described by participants as influencing the project's evaluation and comprehensiveness. The current chapter presented an analytic look at data collected for each case and across the fields of tobacco control and childhood obesity.

Challenges to evaluation, as reported by project representatives, seem to center around the scope of the projects, methodological challenges of capturing synergy and putting various pieces into a coherent whole, and communicating findings to various audiences in the most effective ways. A challenge, potentially impacting projects more than was described is answering the question of "who gets credit?" when multiple partners, funders, decision-makers and entities are involved in implementing such a program.

Strengths of the evaluation reported by project representatives appear to link to things these projects "had," including internal evaluation capacity, a guiding logic model and flexibility to modify evaluations as needed.

Beyond using the funder or legislative criteria to help guide development of the evaluation, other elements described by most projects, across both fields, came from sources external to their project—published literature and theory, external evaluation experts, and resources from fields beyond their own.

Participants from all cases identified their funding as having influence on the comprehensiveness of their evaluation. Interestingly, while all cases used the legislation or funder criteria in the development of their evaluation, no project participants considered these criteria as having influencing the comprehensiveness of their evaluation. All projects operationalized what a comprehensive evaluation meant and guided the development of this comprehensive evaluation themselves.

Representatives from cases across both fields identified multiple contextual elements as influencing their evaluation's comprehensiveness, including: experience and knowledge of internal evaluators, access and advice from evaluation experts beyond the project, an overarching logic model or framework, other fields beyond their own, and their project's funding. Contextual elements identified by project representatives overall appear, to the author, to cluster around themes of external factors, knowledge factors and one other factor, which does not categorize well. External factors, including requirements by the legislation, restrictions by the funder and the project's funding, had mixed influence, as described by participants, on the comprehensiveness of their project's evaluation. External constraints, such as restrictions, do not appear to have influenced the comprehensiveness for these cases overall, but having positive external factors, such as funding available, did influence the comprehensiveness. Knowledge factors identified by cases include experience and knowledge of internal and external evaluators, access to evaluation experts beyond the project, fields beyond their own, developing a logic model, and project members adding in their own components and operationalizing comprehensiveness themselves. These elements all relate to knowledge and expertise within the project, or access to experts to work with the project on their evaluation. All of these knowledge factors were identified by most cases, and across both fields, as influencing the comprehensiveness, with the exception of experience and knowledge of external evaluators and fields beyond their own, which were identified by two cases. The only element that did not seem to categorize easily was length of time carrying out the project, which was identified by only one project as influencing the comprehensiveness of their evaluation.

Each project has found a different way to successfully carry out their evaluation. Three of the projects currently conduct their evaluation internally, while the remaining project has their evaluation conducted by an external evaluation firm.

All cases studied were described through literature and during interviews as examples put forward by others. These projects have shared their expertise by provided guidance to others, sharing materials and lessons learned to other programs in tobacco control, childhood obesity, and other fields.

When comparing case study data with the original Framework for Comprehensive Evaluation of Multi-Level Interventions, all of the contextual elements from the Framework were found through data collected to have influence, to at least some extent, on the evaluations of the cases studied; but only some of the elements specifically influenced the *comprehensiveness* of the evaluations for cases studied. Based on case data, contextual elements identified by participants were further categorized by the author as: 1) influencing the comprehensiveness of the evaluation, 2) influencing development of the evaluation (but not comprehensiveness), or 3) influencing the evaluation overall, but not specifically the

development or comprehensiveness. One element (evaluator access) was described by project representatives from all cases as important overall, and participants indicated that a lack of it would likely influence evaluation comprehensiveness, but since all projects had this element, it was difficult to truly assess influence. To be included in one of the categories above, three cases (covering both fields) or all four cases needed to have project representatives identify this contextual element as influencing the comprehensiveness, the development, or the evaluation overall.

Elements from the Framework described by participants as specifically influencing the *comprehensiveness* of the evaluation for all four cases included: resources, training/ experience of evaluator, access to information to guide evaluations, and evaluation capacity of organization.

Contextual elements in the Framework were broader than those identified by case participants as influencing the comprehensiveness of their evaluations. For example, cases were asked about project resources and the influence on the evaluation's comprehensiveness, but what were described were specifically having funding and access to, and advice by, external evaluation experts.

While differences exist between these fields and the evaluations carried out by projects in these fields, as has been described, many similarities were found between cases within the same field and across the fields of tobacco control and childhood obesity. Strengths, challenges, sources of information used while developing the evaluations, and contextual elements described as influencing the comprehensiveness of the evaluations had many overlapping areas between the cases and both fields.

CHAPTER 6: IMPROVING EVALUATIONS OF MUTLI-LEVEL INTERVENTIONS IN TOBACCO CONTROL AND CHILDHOOD OBESITY: WHAT HAVE WE LEARNED AND WHERE DO WE GO FROM HERE?

Introduction

A multiple case study was conducted with four projects evaluating multi-level interventions in tobacco control and childhood obesity. These cases were selected by the author to further understand contextual elements influencing their comprehensive evaluations of multi-level interventions. Data were collected through interviews and document reviews with each case selected. Analysis within and across cases, and across fields, were conducted with these data.

Descriptive presentation of each case individually provided rich detail about each project and contextual elements described by case representatives as influencing their evaluation and its comprehensiveness. Comparing cases within each field and across tobacco control and childhood obesity provided information on similarities and differences in strengths and challenges to their evaluations, sources of information guiding development of their evaluations, and contextual elements identified as influencing the comprehensiveness of their evaluations. Further, comparison was made of case data collected with the Framework developed early in the study, to determine the influence of contextual elements on the projects studied.

This chapter addresses the following question: *What lessons learned, promising practices, and recommendations can be derived from the study of cases and assessment of contextual elements, in order to improve evaluations of multi-level interventions?*

This chapter describes lessons learned through cases studied in tobacco control and childhood obesity, including data compared across both fields and against the Framework developed earlier in this dissertation. Promising practices and recommendations are also presented, with the intent of translating data gathered from the cases into information that can be used by others in the fields of evaluation and public health. This information can be valuable to other projects and may help improve the quality and comprehensiveness of evaluations conducted by others in the future. Individuals from different perspectives, such as, project staff, internal and external evaluators, funders, and consumers of evaluation, can find value in the lessons that have been learned and suggestions provided for future work.

Lessons Learned from Cases in Tobacco Control and Childhood Obesity

Lessons learned were developed by comparing case study data with the original Framework (Figure 1, Chapter 3) and through descriptions by project representatives about strengths and challenges to their evaluations, sources of information guiding development of their evaluations, and contextual elements influencing the comprehensiveness of their evaluations.

Hypotheses developed early in the study were not supported by case data, in that areas explored were related to the evaluations, but not fully in the ways expected. One hypothesis looked at a specific contextual element, evaluation requirements by a project's funder, which was hypothesized to have particular influence on the comprehensiveness of multi-level evaluations. Project staff and evaluators for all cases described during interviews that this element did influence the development of their evaluations, but not the comprehensiveness. Additionally, it was hypothesized that all contextual elements presented in the original Framework would influence evaluation comprehensiveness. Again, findings were mixed. All contextual elements were described by participants from each project as influencing their evaluation, to some extent, but only certain elements were indicated to have influenced the *comprehensiveness* specifically. Some elements were described as influencing the development of the evaluation or influencing the evaluation overall, but not necessarily the comprehensiveness. Further detail about hypotheses developed and comparison to case data were presented in Chapter 5.

Though these hypotheses were not supported as written by case data, they were not completely off-base. And the original Framework, compared to data collected from respondents, aided in the identification of elements contributing in different ways to the evaluations studied. Contextual elements described by interview participants, as how they influenced the evaluation, were fairly consistent across cases and fields. When comparing case data with the Framework, several categories emerged of how contextual elements were described as influencing the evaluations of projects studied. These categories, developed by the author, include contextual elements: 1) influencing the <u>comprehensiveness</u> of the case evaluations, 2) influencing <u>development</u> of the evaluations (but not the comprehensiveness), or 3) influencing the evaluation overall, but <u>not</u> specifically the development or comprehensiveness. One element, evaluator access, was described by project representatives from all cases as important overall, and that a lack of evaluator access would likely influence evaluation comprehensiveness. But since none of the projects studied had a challenge with this, and its influence was not described in enough detail, actual influence was difficult to assess. Contextual elements in each of the categories described above (and displayed in Figure 2, Chapter 5) are presented briefly below. To be included in one of the categories above, most or all cases needed to fit the pattern. Therefore, three cases (covering both fields) or all four cases needed to have project representatives identify this contextual element as influencing the comprehensiveness, the development, or the evaluation overall.

Contextual elements described by participants from all cases as influencing the <u>comprehensiveness</u> of their evaluation are listed below.

- *Resources-* This included funding for evaluation and access to external evaluation experts for all projects.
- *Training/Experience of Evaluator-* This element was described as influencing the comprehensiveness through internal and/or external evaluators for all projects.
- *Access to Information to Guide Evaluation-* Sources of information influencing evaluation comprehensiveness for all projects were: use of overarching logic model/ framework, experience/ knowledge of evaluators, and fields beyond their own.

- *Evaluation Capacity of Project-* All cases defined and added their own elements to make their evaluation comprehensive.

Contextual elements identified by representatives from all cases as <u>guiding the development</u> of their evaluation are listed below.
- *Evaluation Requirement by Funder-* This influenced the development of the evaluations, but not the comprehensiveness, for all projects.
- *Intended Use of Evaluation-* Described by all projects as important to the development of design, methods, questions asked and information reported.

Contextual elements described by participants from most projects as <u>influencing evaluations overall</u>, but not specifically the comprehensiveness or development of the evaluation are listed below.
- *Length of Time Carrying Out Project*
- *Agency Support*
- *Social/Environmental Conditions, Politics, and Competing Organizations/Interests*

A contextual element described by representatives from all projects as <u>important overall</u>, and the lack of it would likely influencing evaluation <u>comprehensiveness</u> was evaluator access.
- *Evaluator Access-* Access to an evaluator that could develop/ carry out the multi-level evaluation.

In general, contextual elements from the Framework were broader than those identified by the case representatives as influencing the comprehensiveness of their evaluation. For example, cases were asked about project resources and the influence on the evaluation's comprehensiveness, but the projects described more specifically having the funding to do evaluation and access to, and advice by, external evaluation experts as key resources influencing the comprehensiveness of their project's evaluation.

Main findings from case data centered on strengths and challenges to the evaluation, sources of information guiding development of the evaluation, and contextual elements influencing the comprehensiveness of the evaluation, as described by representatives from each case and across both fields.

Strengths of the evaluations across both fields are listed below.
- Having the flexibility to modify the evaluation as needed;
- Using a logic model to guide their evaluation; and
- Having strong internal capacity for carrying out their evaluation.

Challenges of the evaluation described across both fields are listed below.
- Scope and complexity of the evaluation;
- Bringing various pieces of the evaluation together into a coherent whole;

- Capturing synergy;
- Having large amounts of data and struggles with making the best use of data; and
- Reporting and how best to communicate findings, especially to different audiences.

Strengths reported by project representatives appear, to the author, to link to things these projects "had," including internal evaluation capacity, a guiding logic model, and flexibility to modify evaluations as needed. Challenges reported by project representatives seem instead to center around the scope of the projects, methodological challenges of capturing synergy and putting various pieces into a coherent whole, and communicating findings to various audiences in the most effective ways.

Another challenge involved answering the question of "who gets credit?" when multiple partners, funders, decision-makers and entities are involved in implementing such a program. Considering that these interventions and evaluations span multiple socioecological levels, often using multiple strategies within each of these levels, and are often quite costly, it would make sense that it would be extremely difficult, if not impossible, to untangle the impact of these different components. Even comprehensive and expensive evaluations, with experienced evaluators and sophisticated designs have struggled with capturing the synergy and untangling various components and impact taking place within their project.

Presented below are main findings for sources of information identified by project representatives as guiding development of their evaluations and contextual elements influencing comprehensiveness of evaluations for cases studied. These findings are presented when all cases identified the same element, as well as, when the same element was found across both fields. To be considered present across both fields, as described below, representatives from at least one tobacco control case and one childhood obesity case had to identify this element.

Sources of Information Identified as Guiding Development of Evaluation for <u>All Cases</u> are listed below.
- Published literature/ theory;
- Expertise of internal and/or external evaluators; and
- Criteria from the legislation or funder.

Sources of Information Identified as Guiding Development of Evaluation across <u>Both Fields</u> are listed below.
- Published literature/ theory;
- Expertise of external evaluators;
- Overarching evaluation logic model or framework;
- Criteria from the legislation or funder;
- Other programs in the same field; and
- Other fields beyond their own.

Beyond using the funder or legislative criteria to guide development of the evaluation, other elements described by most projects, across both fields, came from sources external to their project—published literature, external experts, and resources from other fields.

Contextual Elements Described as Influencing Comprehensiveness of Evaluation for <u>All Cases</u> are listed below.
- Funding available;
- Experience and knowledge of internal and/or external evaluators; and
- Project operationalizing evaluation comprehensiveness and adding elements to their evaluation to make it comprehensive.

Contextual Elements Described as Influencing Comprehensiveness of Evaluation across <u>Both Fields</u> are listed below.
- Funding available;
- Experience and knowledge of internal evaluators;
- Accessibility and advice of evaluation experts beyond the project;
- Overarching evaluation logic model/ framework; and
- Project operationalizing evaluation comprehensiveness and adding elements to their evaluation to make it comprehensive.

Contextual elements identified by project representatives appear, to the author, to cluster around themes of external factors, knowledge factors and one other factor, which does not categorize well. These are presented below in further detail.

External factors, including requirements by the legislation, restrictions by the funder and the project's funding, had mixed influence, as described by participants, on the comprehensiveness of their project's evaluation. Interestingly, while all cases used the legislation or funder criteria as a source of information to guide the development of their evaluation, no projects considered these criteria as having influenced the comprehensiveness of their evaluation. Restriction by the funder was identified by only one case, but it was quite influential to their project. Funding for the project was the only of these external factors identified by all cases as influencing evaluation comprehensiveness. Therefore, external constraints, such as restrictions, do not appear to have influenced the comprehensiveness for these cases overall, but having positive external factors, such as funding available, did influence the comprehensiveness.

Knowledge factors identified by cases include experience and knowledge of internal and external evaluators, access to evaluation experts beyond the project, fields beyond their own, developing a logic model, and project members adding in their own components and operationalizing comprehensiveness themselves. These elements all relate to knowledge and expertise within the project, or access to experts to work with the project on their evaluation. All of these knowledge factors were identified by most cases, and across both fields, as influencing the comprehensiveness, with the exception of experience and knowledge of external evaluators and fields beyond their own, which were identified by two cases. Use of

internal and external evaluators differed for evaluation development and comprehensiveness across the fields. The cases in tobacco control described utilizing only the expertise of external evaluators for the development of the evaluation, but both internal and external evaluators were said to have influenced the comprehensiveness of the evaluation. The opposite was found for cases in childhood obesity, where they utilized both internal and external evaluation experts during evaluation development, but considered only internal evaluators as influencing the comprehensiveness of their evaluations. Additional findings from case data collected involved the role of access to and capacity of evaluators to carry out these evaluations and how intended use influenced the evaluations for the cases. All cases agreed that access to evaluators with capacity to carry out multi-level evaluations was problematic in the field generally. This was noted for statewide evaluations, such as theirs, as well as evaluations done at local and regional levels. Concern was raised that people were not being trained to conduct such evaluations or were calling themselves evaluators when they should not, if they were found at all. These cases, however, have found evaluators with appropriate capacity to carry out their complex evaluations, and while they differ from each other, they have each found a system of using internal or external evaluators for their evaluation that works for them. Intended use influenced the evaluation for all projects, but in different ways; this included influencing evaluation design, methods, questions asked, modifications made to evaluations, and how data were reported and disseminated.

The only element that did not seem to categorize easily was length of time carrying out the project, which was identified by only one project as influencing the comprehensiveness of their evaluation.

Promising Practices by Cases

Evaluation successes and contextual elements identified as influencing evaluation comprehensiveness were merged, by the author, to determine promising practices from these cases. Five factors for evaluation success and comprehensiveness emerged, including: Resources, Expertise of Evaluators, Evaluation Capacity of the Project, Access to Information to Guide Evaluation, and Flexibility. These factors were described by representatives from cases studied and are described below and presented in Figure 1.

As can be seen, several elements within these factors overlap, which is why Figure 1 presents these factors as a Venn diagram. The combination of having strong internal capacity for evaluation, access to external evaluation expertise, adequate and consistent funding, a logic model to guide the evaluation, and flexibility to modify the evaluation as needed, appear to be a golden combination for having the tools to successfully develop and carry out a comprehensive multi-level evaluation.

It is not just about these individual elements, but rather how they overlap and work with each other. For example, a primary resource described as influential for case evaluations was the funding available, but this funding not only provided for the project and evaluation to be carried out, but in the case of several projects studied, enabled consulting with external evaluation experts and the building and potentially maintaining of internal evaluation capacity. Similarly, expertise of evaluators can bring not only their knowledge, but also

sources of information that helped guide evaluations. These evaluators brought to the project access to published literature and theory, methods and design to evaluate their projects, and often worked with these projects to develop an overarching logic model to guide their evaluation. But, the project had to have access to and likely funding available to work with external evaluation experts. In that example, expertise of evaluators was linked also to elements of access to information to guide and resources. As a last example, the flexibility to modify evaluations as needed is a function of having the capacity to realize that modifications should be made and how that should happen, as well as, having the decision-making power to make such adjustments.

Figure 1. Five Factors for Evaluation Success and Comprehensiveness

Resources

Flexibility

Expertise of Evaluators

Access to Information to Guide

Evaluation Capacity of Project
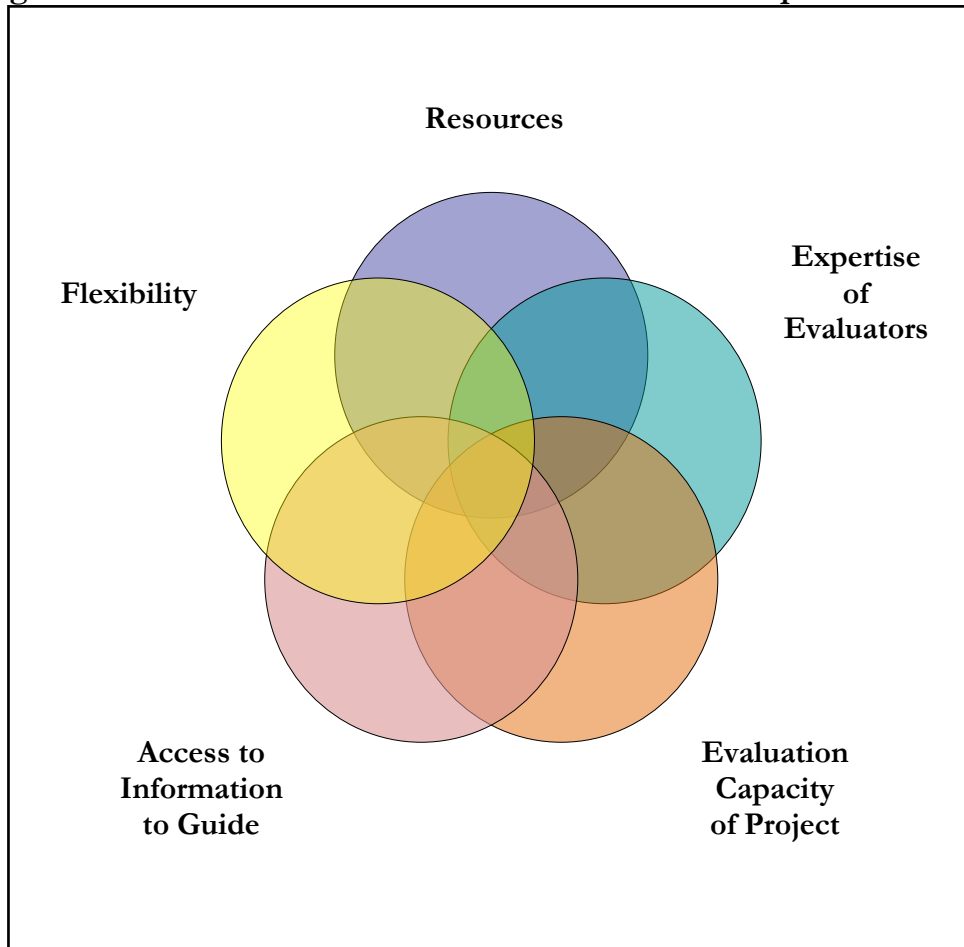
<u>Resources</u>

Funding for their evaluation and access to external evaluation experts were described by participants from all projects as the most valuable resources influencing the comprehensiveness of their evaluations. Representatives from all cases, and published documents, indicate that all projects were well funded. Additionally, all cases had been funded consistently for many years, ranging from seven to over 20 years. The amount and

consistency of the funding may have enabled the projects to have other elements of success, such as building internal capacity and contribution to the project by external evaluators.

Expertise of Evaluators

Experience, knowledge and expertise of internal and external evaluators were described by project participants as influencing both the development and comprehensiveness of their evaluations.

The role of external evaluators and advise by external evaluation experts, whether a long-term solution for a project or not, clearly plays a vital role in guiding the development and comprehensiveness of evaluations. External evaluation experts were described by participants from all projects as influential, whether they currently have their evaluation conducted by an internal or external evaluation team.

Both internal and external evaluators brought to these projects knowledge and access to published literature, theory, logic models, and methodologies to aid in developing and carrying out multi-level evaluations.

Evaluation Capacity of the Project

Internal capacity for carrying out evaluation was specifically identified as an element of success for cases in both fields. Having access to an external evaluator is not enough. Building internal capacity for evaluation, whether to actually carry out evaluation activities, oversee contracts, or modify the evaluation as necessary, also appears quite important to the success of a program in consistently implementing their evaluation. Three of the four projects studied now conduct the majority of their evaluation activities internally, though this was not always the case. Learning from past experiences, some of these projects made a conscious decision to build internal capacity to improve their evaluation and avoid challenges experienced in the past.

Additionally, project members were identified as having added elements to the evaluation, to make it comprehensive, beyond the minimum guidelines set by funders or the legislation. It was also the projects themselves that operationalized what a comprehensive evaluation meant and developed a plan to carry it out. All four cases studied described this process and its influence on the comprehensiveness of their evaluation.

Access to Information to Guide Evaluation

All projects studied used different sources of information to guide the development and influence the comprehensiveness of their evaluations. Specific sources of information described by participants as influencing the evaluation's comprehensiveness included using an overarching logic model or framework, experience and expertise of internal and external evaluators, and resources from fields beyond their own. Using a logic model to guide the evaluation was described by representatives from several projects as an element of their project's success.

Flexibility

        Having the flexibility to modify the evaluation as needed, whether to adapt to changes in the project, address emerging issues or make tweaks to improve the evaluation, were described as an element of success by most cases and across both fields. Additionally, several projects described changing how their evaluation was conducted, including moving from external to internal evaluators or adapting to restrictions of the funder over the course of their project. Having the flexibility, capacity, and decision-making authority to make changes to the evaluation were described as improving the evaluations for these projects.

        Lessons learned and promising practices may be used to improve evaluations of projects beyond those studied. This information may be used by different people and for different stages of a project's evaluation, therefore, having multiple strategies for improvement can be beneficial. Examples of how lessons learned from projects studied can benefit multiple audiences differently are presented next.

        Several key elements were described by representatives from the cases studied as influential to their evaluation, both in its development and comprehensiveness. This information can provide guidance to people in different roles and at different stages of a project's evaluation. Funders may consider some ideas more promising than a project staff member, who might find something else of greater interest. Likewise, an internal evaluator may find lessons of value differently than an external evaluator. While funding was identified as critical for the evaluation of these projects, so was expertise by external evaluators, and the knowledge and flexibility of project team members in developing and carrying out complex multi-level evaluations. There is a role for various project participants in moving evaluations forward and making improvements.

        Likewise, at different stages of a project's evaluation, different elements may provide more value than at other times. For example, specific sources of information were described by representatives from all cases as guiding the development of their evaluation, but different elements were indicated, again, for all cases, as influential in making evaluations comprehensive. Depending on the purpose, this differing information can provide guidance to other projects, whether wondering where to look for help when developing their own evaluation or providing guidance related specifically to increasing comprehensiveness. Published literature, theory and criteria from the legislation or funder were described uniquely for guiding the development of the evaluation by all projects. Whereas, funding available and the project's ability to operationalize comprehensiveness and add their own elements to the evaluation were described specifically by participants from all projects as what influenced the comprehensiveness of their evaluations.

Recommendations for Improved Evaluations and Future Work

        Recommendations in this section include a combination of addressing challenges identified, promoting promising practices identified, and providing other areas of consideration for future work. Information presented can be used by multiple audiences to

improve evaluations of multi-level interventions in the future. Different levers may be useful to different people and at different times. Multi-level projects, such as those studied, are not carried out quickly or simply, so having multiple suggestions for improvement, to help with the development, comprehensiveness and other areas of the evaluation, can be beneficial.

A challenge described by representatives from three of the cases studied included measuring the synergy of their work being carried out on multiple levels concurrently. Even with sophisticated and expensive evaluations, these projects could not identify specific elements of the intervention that most impacted changes seen. A recently published document by the Institute of Medicine (2010) recommends thinking about obesity from a systems perspective, in order to better understand the complexity of the issue at hand. A presentation at the American Evaluation Association conference about this Institute of Medicine (IOM) report elaborated further, describing that a systems approach moves beyond a socioecological model to account for the dynamic nature of what is going on in obesity (Economos & Chatterji, 2010). A just-released publication by Williams and Hummelbrunner (2010), titled "Systems Concepts in Action: A Practitioner's Toolkit" may provide additional contribution to better understanding the idea of systems approach when working in these fields.

- Recommendation: Future work should seek to develop methods and advance analysis techniques to support projects in their attempts to measure synergy.
- Recommendation: Further study of a systems approach and its applicability to capture the dynamic nature of work being done and impact made in obesity and other public health areas, such as tobacco control.

A related and ongoing question of "who gets credit?" is another area to address, to move the field forward. When multiple partners, funders, decision-makers, and entities are involved in implementing and evaluating a program that spans multiple socioecological levels, there is naturally a question of who gets the glory and who gets the blame, depending on the results. Because these programs are complex and it is difficult to untangle the impact of different components, it is challenging to know whose funding and effort had the most influence. This is an ongoing source of challenge for multi-level projects.

- Recommendation: Strategies should be developed to increase the options to accurately measure this element. Additionally, a shift in expectations by funders and partners may reduce the need to separate out by different components and focus more on the efforts and impacts of the whole.

Funding was described by participants from all projects as the most valuable resource for their evaluation and they linked this resource to the comprehensiveness of their evaluations. Project respondents also described the consistency of funding and long duration of the project as influencing their evaluation. Funding for projects such as these can be complicated, especially when multiple funding sources are combined, often with their own reporting requirements and priorities. Various constraints on funding exist. Examples include working with categorical funding or when funding can only be used for some work

and not other work, and this can be a source of frustration to projects trying to fund their multi-level work. Projects with multiple sources of funding need to balance requirements for each of these funding sources, administration of different funding, and often reporting of different information on different timelines.

- Recommendation: Having a sole source of funding, with the flexibility to use it as appropriate to implement the evaluation, can reduce the amount of effort spent on bureaucratic/administrative issues, navigating the system, separating out funding and activities to report to individual funders, and provide more cohesion in the program. Another option is increased flexibility by funders around reporting and other requirements, to enable the merging of funding sources and decrease the burden on the project implementing the intervention and/or evaluation.
- Recommendation: Funding projects for longer periods of time or tying into a consistent stream of funding can also improve evaluations by maintaining consistency in staff, capacity, and institutional knowledge of the project.

Also described by participants as a challenge for the field, is access to evaluators with training, knowledge and capacity to develop and carry out complex evaluations. Almost all participants interviewed described this as a concern. Several participants described not knowing who was training people to do this work and considered that a problem. Professional training and academic programs do exist, but they are not training enough people to meet the need out in the field. Graduate programs focused on evaluation do exist, but they are not plentiful.

- Recommendation: More work should be done to recruit people into the field of evaluation, provide training and resources to improve capacity, and improve networking to link experienced evaluators with programs in need.
- Recommendation: Increase knowledge, capacity and evaluation training by attending evaluation conferences and professional development programs. Examples of such opportunities include: American Evaluation Association's Annual Conference, American Evaluation Association and CDC's joint Summer Institute, American Evaluation Association's Professional Development Webinars, The Evaluator's Institute Professional Development Courses, Claremont Graduate University's Professional Development Workshops, and the University of Michigan's Institute of Social Research Summer Institute. Some of these organizations offer certificate programs. For more information on these professional development programs, see references listed (American Evaluation Association, 2010; American Evaluation Association and CDC, 2010; Claremont Graduate University; The Evaluator's Institute, 2010; University of Michigan).
- Recommendation: Increase the number of graduate programs offering concentrated study in the field of evaluation.

Another area for improvement relates to reporting and communication of findings. Several projects studied described struggling with how best to report their data, especially when multiple stakeholders were involved and wanted to see information in different ways.

Learning from previous experiences can, however, be quite powerful. One case described another project (in the current cases studied) as providing guidance, through tools, sources of information disseminated by the other project, and enabling them to learn from what worked and what did not. They were able to fast-track developing an evaluation of their own that was comprehensive and effective for them. This project had their own political climate and contextual factors to work with, so an exact translation did not work, as is most often the case, but within a few years they were also considered a leader within their field. Determining the project's audience is another key element to this process, so that information can be tailored to that audience and ultimately utilized by them. One project described the process of disseminating information back to the community where it was collected, to improve the feedback loop, but constraints of timing and a system not supporting such work made it too challenging to do.

- <u>Recommendation</u>: Developing and disseminating information in an easy-to-understand format and for multiple audiences can share lessons learned and promising practices, in a way that can actually be used.
- <u>Recommendation</u>: Reporting and dissemination of information, including providing information back to communities involved, when appropriate, should be included in the process of the evaluation, so that it is not left out. It may require support by funders, the project's organization, and project leaders to make this an expectation and part of the evaluation, rather than something tacked on to the end of the project, if there is time or budget remaining.

Beyond areas described by projects studied as areas for improvement, a general call to the field has been made to improve evidence, interventions and evaluations in multiple chronic disease areas, including tobacco control and childhood obesity. At a recent American Evaluation Association conference, several sessions focused on these two fields and work being carried out to advance evaluations and interventions in these areas. Much of the discussion focused around childhood obesity as a field of great growth, but with a limited evidence base. A presentation about an IOM framework described that there is an evidence gap, particularly for high quality evidence that goes beyond simple interventions, but that action needed to take place to curb the obesity epidemic, so the field is working with what is has (Economos & Chatterji, 2010). Representatives from the CDC indicated recently that resources to guide childhood obesity are not as readily available or developed, as they are for tobacco control, and that much less is known about "what works" in childhood obesity (Compton et al., 2010; Farris et al., 2010). Nine hundred million dollars are being spent over the next two years by the CDC to work with states to develop and evaluate policies in tobacco control and nutrition, physical activity and obesity (Compton et al., 2010). The CDC is also looking to use evaluation to apply lessons learned in one program across other chronic disease prevention programs (Compton et al., 2010).

- <u>Recommendation:</u> Expand evidence base, especially for high quality evidence of multi-level interventions and evaluations. Particular emphasis should be paid to childhood obesity, which has less information readily available currently to guide interventions and evaluations.

Requirement by funder was proposed as a contextual element influencing evaluation comprehensiveness, which was not supported by case data. These requirements did, however, influence the development of the evaluations for all cases. Therefore, projects did include what was asked of them from their funder or the guiding legislation. It was the projects themselves that made their evaluations comprehensive, but this required having the capacity, budget and access to tools to accomplish. If requirements were more stringent, rather than minimal requirements, additional projects in the field may have more comprehensive evaluations than is currently found. At the same time, it should be noted that flexibility from the funder and for the project was an important element of success. Balancing the tension between these two areas should be given consideration.

- Recommendation: Increasing evaluation requirements by funders may improve the comprehensiveness of evaluations in the field generally. But this should only be done with other supports and flexibility included, to increase the chance of the project's success at carrying out their evaluation. Supports include adequate and consistent funding, evaluation capacity, access to evaluation expertise and tools to develop and carry out a comprehensive evaluation.

Promising practices for further promotion include the use of a logic model or framework, accessing external evaluation expertise, building internal evaluation capacity, acquiring adequate funding for a comprehensive evaluation, and maintaining flexibility to modify the evaluation as needed.

Bringing multiple pieces together into a coherent whole was described as a challenge by several projects. The complexity, scope, and mass amounts of data can further overwhelm projects carrying out work on a large scale. Several projects described the use of a logic model or framework and working with experienced evaluators as not only influencing the comprehensiveness of their evaluation but also a great strength of their evaluation.

- Recommendation: Developing a logic model or framework to guide the project and its evaluation can be helpful, as well as, working with an evaluator that has experience working on projects of this scale and level of complexity.

External evaluation experts were described repeatedly as playing a key role in guiding the development and the comprehensiveness of evaluations. Having access to external evaluators or evaluation experts outside of the project can improve the development and comprehensiveness of evaluations regardless of whether evaluations are ultimately carried out internally or externally.

- Recommendation: Projects should consider consulting with external evaluation experts by bringing them in during the planning phase of the evaluation and throughout the project to gather new ideas and gain a fresh perspective about their evaluation.

Building internal capacity may be a way to improve the quality of evaluations, so that they could stay consistent for a longer period of time. Though, this would likely depend on the turnover rate within the organization considering this option.

- <u>Recommendation</u>: Projects should prioritize building internal capacity. Having a readily available team of trained evaluators can address both access and capacity issues for these organizations.

Flexibility to modify the evaluation as needed was identified by several projects as important to their success. Sometimes this flexibility came from vague legislation language, limited oversight by funders, or the project having enough autonomy to make their own decisions and carry out their work how they saw most appropriate. Another component of this, beyond having the option of flexibility, is the project having team members that are knowledgeable enough to adapt the evaluation as needed, whether the project or its context had changed, or for many other number of reasons. This factor appears to include a combination of projects having both flexibility and the capacity to make use of it as needed.

- <u>Recommendation</u>: Developing a recommendation for how a project becomes more flexible is more difficult than some of the other factors described previously. Flexibility may come from a combination of strong leadership and competency in the program, as well as, organizational and funder support. This may be something that funders and organizations need to develop expectations around, beyond the project working on their own to gain more flexibility. Building internal capacity to know how to adapt the evaluation as needed may be increased as internal evaluation capacity is built or increased.

In addition to the recommendations described above, future work beyond this dissertation can go in several directions, including further study of contextual elements influencing evaluation comprehensiveness, further study of other elements within the Framework, and study with additional projects and in fields beyond tobacco control and childhood obesity. The Framework developed early in this study provides a reasonably accurate list of factors described by multiple projects as influencing their evaluations of multi-level interventions, but only certain factors were described as specifically influencing the evaluation's comprehensiveness. Future studies could investigate the influence of these and additional contextual factors on other evaluations, their comprehensiveness, and other areas of consideration. Broader contextual elements of Social/Environmental Conditions, Politics, and Competing Organizations/Interests were investigated in a limited capacity in the current study. Further study on the influence of broader contextual factors could provide more detail about the role they play within evaluations and evaluation comprehensiveness. Areas of the Framework beyond contextual elements were not explored with these cases. Future work could investigate the middle section of the Framework, whether using the same or other cases. Additionally, future studies could look at other cases within these fields, or fields beyond tobacco control and childhood obesity, that are carrying out multi-level interventions, to see if similar patterns are found.

Conclusion

In-depth study of several projects evaluating multi-level interventions in tobacco control and childhood obesity was carried out through this dissertation to fill a current gap in the fields of evaluation and public health. This study expands knowledge by having assessed current multi-level projects and their context, in order to better understand contextual elements described as influencing the comprehensiveness of their evaluations.

Lessons learned through the study of these cases informed the determination of promising practices and recommendations, which were described in this chapter.

Promising practices included a combination of having strong internal capacity for evaluation, access to external evaluation expertise, adequate and consistent funding for the evaluation, a logic model and other sources of information to guide the evaluation, and flexibility to modify the evaluation as needed, were found as key factors for the success and comprehensiveness of projects studied. This pattern was found across cases and across both fields. Beyond these individual elements, discussion was provided about how these elements overlap and work with each other. For example, a primary resource described as influential for case evaluations was the funding available, but this funding not only provided for the project and evaluation to be carried out, but in the case of several projects studied, enabled consulting with external evaluation experts and the building and potentially maintaining of internal evaluation capacity. Similarly, expertise of evaluators can bring not only their knowledge, but also sources of information that helped guide evaluations. These evaluators brought to the project access to published literature and theory, methods and design to evaluate their projects, and often worked with these projects to develop an overarching logic model to guide their evaluation. But, the project had to have access to and likely funding available to work with external evaluation experts. In that example, expertise of evaluators was linked also to elements of access to information to guide and resources. As a last example, the flexibility to modify evaluations as needed is a function of having the capacity to realize that modifications should be made and how that should happen, as well as, having the decision-making power to make such adjustments.

Recommendations were developed to address challenges identified, promote promising practices, and provide considerations for future work. Recommendations were developed to address areas including: capturing synergy, determining "who gets credit", funding issues, access to evaluators and improved training of evaluators, reporting and communication, building an evidence base, use of logic models, inclusion of external evaluation experts, building internal capacity, and flexibility to modify evaluations as needed. The ultimate goal of this dissertation is to provide the fields of evaluation and public health with recommendations to increase comprehensiveness and quality of evaluations for multi-level interventions in the future. Suggestions for future work were also provided.

CHAPTER 7: LIMITATIONS AND CONCLUSIONS

Limitations

Limitations to this dissertation do exist. Background information and development of the Framework relied heavily on literature reviews and publically available sources of information; therefore, there are several limitations for this reason alone. Only information published in peer-reviewed journals or published online in some other format were included in the literature reviews. There are likely large amounts of information in working documents or internally used information that did not make it to a publically available source, which may have further informed this dissertation. Additionally, limitations exist as to how searches are conducted and the possibility that other sources were available publically or in peer review journals, but not captured in the current study. The necessary step of limiting search terms and search engines used may have resulted in some missed information. The use of publically available information limited the amount of information known about recent evaluations conducted. If information was not in articles or reports reviewed early in the study, and not found elsewhere, it was left unknown about these projects. Such missing information may have resulted in a project not being selected for further study through the multiple case study.

Several limitations exist to the study of cases for this dissertation. Only four cases were studied within two different fields. The inclusion of additional cases could have provided differing information than that found with the cases included for study currently. The analysis was limited based on the small number of cases included and the intentions for the conclusions drawn. A more rigorous design, with a stronger theory-based study, would have enabled stronger conclusions to appropriately be drawn than would make sense for the current study.

Additionally, purposeful sampling was used for both the selection of cases and selection of interview participants within each case studied. The opportunity to statistically generalize was not the intent of this study, and is often not a goal of most case studies or qualitative research. Including additional interview participants or interviewing everyone associated with the projects and evaluations may have provided information beyond what was gathered and presented here. Additionally, including more cases would have allowed more opportunities for attempted cross-case replication of findings and greater confidence in the validity of findings.

Case study findings are based on interviews and document reviews with the four projects studied. Therefore, most information is based on self-reported information about the project, its influences and decisions made, and the evaluation. With self-report and interviews, there can be concern that information reported may have been incorrectly described, misunderstood or provided differently based on concerns about perceived response or later consequences for information shared. All informants were treated as having equal weight and expertise to speak about the project when collecting and analyzing case data. Based on constraints of Committee for Protection of Human Subjects approval and agreements with the projects participating, information about interview participants is

provided to add some description about informants, without providing enough detail to identify anyone individually within the project. Member checks were conducted to ensure that information presented is as intended by participants and does not report sensitive data.

Based on the necessity to keep the scope of this dissertation manageable, only contextual elements from the original Framework were explored in detail during through the four cases studied. Study of the middle section of the Framework would provide additional information about these projects and may have further informed influences around the comprehensiveness of the evaluations. Also, broader contextual elements, such as social/environmental conditions, politics, and competing organizations/interests were not studied in the same level of detail as other contextual elements for the cases studied. Further study on the influence of broader contextual factors could provide more detail about the role they play within evaluations and evaluation comprehensiveness.

Conclusions

There are large gaps in information currently about the way evaluations of multi-level interventions are conceptualized and implemented in public health practice. Existing evaluation literature includes few critiques of evaluation practice with respect to multi-level interventions and provides little guidance to evaluators as they attempt to design and implement comprehensive evaluations of such interventions. Multi-level interventions that attempt to create change or impact at two or more levels are being developed and implemented with increasing frequency in several fields, including public health. However, these interventions are often not comprehensively evaluated. If evaluation is conducted, it is often piece-meal, and important elements such as data collection on all levels intervened are often lacking. The reasons behind this disconnect between the comprehensiveness of the intervention and evaluation is not clear.

This dissertation was conducted to fill a current gap in the fields of evaluation and public health. This study expands knowledge by having assessed current multi-level projects and their context, in order to better understand contextual elements described as influencing the comprehensiveness of their evaluations.

The author reviewed literature and current evaluation practice around multi-level evaluations and developed a Framework for Comprehensive Evaluation of Multi-Level Interventions (Framework). Four projects evaluating multi-level interventions in tobacco control and childhood obesity were studied through a multiple case study to understand contextual elements influencing their evaluation. Analysis of case data took place individually for each project, across projects within the same field, and across the fields of tobacco control and childhood obesity. Case data were also compared with the Framework developed early in the study, to assess how contextual elements influenced case evaluations.

Cases studied provided information on strengths and challenges with their evaluations, sources of information used in the development of their evaluations, and contextual elements that influenced their evaluations. While the original Framework and hypotheses developed early in the study were not fully supported, case data provided

valuable information on nuance around contextual elements, such as how different elements influenced evaluation development versus evaluation comprehensiveness.

Strengths of the evaluation reported by project representatives appear, to the author, to link to things these projects "had," including internal evaluation capacity, a guiding logic model, and flexibility to modify evaluations as needed. Evaluation challenges reported by project representatives seem to center around the scope of the projects, methodological challenges of capturing synergy and putting various pieces into a coherent whole, and communicating findings to various audiences in the most effective ways.

Beyond using the funder or legislative criteria to guide development of the evaluation, other elements described by most projects, across both fields, came from sources external to their project—published literature, external experts, and resources from other fields.

Contextual elements identified by project representatives appear, to the author, to cluster around themes of external factors, and knowledge factors. These are presented below in further detail.

External factors, including requirements by the legislation, restrictions by the funder and the project's funding, had mixed influence, as described by participants, on the comprehensiveness of their project's evaluation. Interestingly, while all cases used the legislation or funder criteria as a source of information to guide the development of their evaluation, no projects considered these criteria as having influenced the comprehensiveness of their evaluation. Funding for the project was the only of these external factors identified by all cases as influencing evaluation comprehensiveness. Therefore, external constraints, such as restrictions, do not appear to have influenced the comprehensiveness for these cases overall, but having positive external factors, such as funding available, did influence the comprehensiveness.

Knowledge factors identified by cases include experience and knowledge of internal and external evaluators, access to evaluation experts beyond the project, fields beyond their own, developing a logic model, and project members adding in their own components and operationalizing comprehensiveness themselves. These elements all relate to knowledge and expertise within the project, or access to experts to work with the project on their evaluation. Use of internal and external evaluators differed for evaluation development and comprehensiveness across the fields. The cases in tobacco control described utilizing only the expertise of external evaluators for the development of the evaluation, but both internal and external evaluators were said to have influenced the comprehensiveness of the evaluation. The opposite was found for cases in childhood obesity, where they utilized both internal and external evaluation experts during evaluation development, but considered only internal evaluators as influencing the comprehensiveness of their evaluations.

Additional findings from case data collected involved the role of access to and capacity of evaluators to carry out these evaluations and how intended use influenced the evaluations for the cases. All cases agreed that access to evaluators with capacity to carry out multi-level evaluations was problematic in the field generally. This was noted for statewide evaluations, such as theirs, as well as evaluations done at local and regional levels. Concern was raised that people were not being trained to conduct such evaluations or were calling themselves evaluators when they should not, if they were found at all. The cases studied,

however, have found evaluators with appropriate capacity to carry out their complex evaluations, and while they differ from each other, they have each found a system of using internal or external evaluators for their evaluation that works for them.

Ultimately, the combination of having strong internal capacity for evaluation, access to external evaluation expertise, adequate and consistent funding for the evaluation, a logic model and other sources of information to guide the evaluation, and flexibility to modify the evaluation as needed, were found as key factors for the success and comprehensiveness of projects studied. This pattern was found across cases and across both fields.

Great overlap was found between cases and across the fields studied. These projects developed their evaluations using similar sources of information and expertise. Also, many similar elements were described as influencing the comprehensiveness of their evaluations. These projects seemed to describe a similar pattern of promising practices and strategies, more than something unique to their project or field, for developing and implementing a comprehensive evaluation. Projects in childhood obesity were somewhat newer than those in tobacco control, and able to draw from resources and lessons learned from earlier projects. Also, projects in tobacco control specifically described providing guidance to childhood obesity programs, to help build their capacity around multi-level interventions, evaluation development and evaluation comprehensiveness. But, projects studied in childhood obesity described providing guidance to others as well (though not to tobacco control programs). Projects studied may be providing guidance to others based on their successes around evaluation comprehensiveness, expertise, and amount of time they have been funded and evaluating their programs, rather than specific to their field. And it may be that childhood obesity as a newer field, and with large amounts of new funding, is looking to others, especially other programs with many years of multi-level intervention and evaluation experience, for more guidance at this time. Therefore, the field itself may not be a particularly distinguishing factor for the cases studied in evaluation development, comprehensiveness, or contextual elements influencing these areas. This may mean that lessons learned from the study of these cases have greater opportunity for transferability to other fields.

Recommendations were developed to address challenges identified, promote promising strategies, and provide considerations for future work. Areas to address and related recommendations are presented briefly below and in further detail in Chapter 6.

Challenge: Measuring the synergy of work being carried out on multiple levels concurrently. Even with sophisticated and expensive evaluations, projects studied could not identify specific elements of the intervention that most impacted changes seen.

- Recommendation: Future work should seek to develop methods and advance analysis techniques to support projects in their attempts to measure synergy.
- Recommendation: Further study of a systems approach and its applicability to capture the dynamic nature of work being done and impact made in obesity and other public health areas, such as tobacco control.

Challenge: Answering the question of "who gets credit?" when multiple partners, funders, decision-makers, and entities are involved in implementing and evaluating a

program that spans multiple socioecological levels. Who gets the glory? Who gets the blame? How to untangle the impact of the different components of these complex programs?

- Recommendation: Strategies should be developed to increase the options to accurately measure this element. This may be linked in with the issue of measuring synergy as described above. Additionally, a shift in expectations by funders and partners may reduce the need to separate out by different components and focus more on the efforts and impacts of the whole.

Challenge: Consistent and adequate funding were described as by projects as influencing their evaluation success and comprehensiveness. Funding for projects such as these can be complicated, especially when multiple funding sources are combined, often with their own reporting requirements and priorities. Various constraints on funding exist, including working with funding that can only be used for some work and not other work. Projects with multiple sources of funding need to balance requirements for each of these funding sources, administration of different funding, and often reporting of different information on different timelines.

- Recommendation: Having a sole source of funding, with the flexibility to use it as appropriate to implement the evaluation, can reduce the amount of effort spent on administrative issues, navigating the system, separating out funding and activities to report to individual funders, and provide more cohesion in the program. Another option is increased flexibility by funders around reporting and other requirements, to enable the merging of funding sources and decrease the burden on the project implementing the intervention and/or evaluation.
- Recommendation: Funding projects for longer periods of time or tying into a consistent stream of funding can also improve evaluations by maintaining consistency in staff, capacity, and institutional knowledge of the project.

Challenge: Access to evaluators with training, knowledge and capacity to develop and carry out complex evaluations. Almost all participants interviewed described this as a concern. Several participants described not knowing who was training people to do this work and considered that a problem. Professional training and academic programs do exist, but they are not training enough people to meet the need out in the field. Graduate programs focused on evaluation do exist, but they are not plentiful.

- Recommendation: More work should be done to recruit people into the field of evaluation, provide training and resources to improve capacity, and improve networking to link experienced evaluators with programs in need.
- Recommendation: Increase knowledge, capacity and evaluation training through evaluation conferences and professional development programs. Examples of such opportunities include: American Evaluation Association's Annual Conference, American Evaluation Association and CDC's joint Summer Institute, American Evaluation Association's Professional Development Webinars, The Evaluator's Institute Professional Development Courses, Claremont Graduate University's Professional Development Workshops, and the University of Michigan's Institute of

Social Research Summer Institute. Some of these organizations offer certificate programs. For more information on these professional development programs, see references listed in Chapter 6.

- <u>Recommendation:</u> Increase the number of graduate programs offering concentrated study in the field of evaluation.

Challenge: Reporting and communication of findings. Several projects studied described struggling with how best to report their data, especially when multiple stakeholders were involved and wanted to see information in different ways. Determining the project's audience is another key element to this process, so that information can be tailored to that audience and ultimately utilized by them.

- <u>Recommendation:</u> Developing and disseminating information in an easy-to-understand format and for multiple audiences can share lessons learned and promising practices, in a way that can actually be used.
- <u>Recommendation:</u> Reporting and dissemination of information, including providing information back to communities involved, when appropriate, should be included in the process of the evaluation, so that it is not left out. It may require support by funders, the project's organization, and project leaders to make this an expectation and part of the evaluation, rather than something tacked on to the end of the project, if there is time or budget remaining.

Challenge: Improve evidence for interventions and evaluations in multiple chronic disease areas, including tobacco control and childhood obesity. At a recent American Evaluation Association conference, several sessions focused on these two fields and work being carried out to advance evaluations and interventions in these areas. Much of the discussion focused around childhood obesity as a field of great growth, but with a limited evidence base.

- <u>Recommendation:</u> Expand evidence base, especially for high quality evidence of multi-level interventions and evaluations. Particular emphasis should be paid to childhood obesity, which has less information readily available currently to guide interventions and evaluations.

Challenge: Requirement by funder was found to influence development of evaluations, but not evaluation comprehensiveness, for projects studied. Projects did include what was asked of them from their funder or the guiding legislation, but these requirements were minimal. It was the projects themselves that made their evaluations comprehensive, but this required having the capacity, budget, and access to tools to accomplish. If requirements were more stringent, rather than minimal requirements, additional projects in the field may have more comprehensive evaluations than is currently found. At the same time, it should be noted that flexibility from the funder and for the project is important as well. Balancing the tension between requirements and flexibility should be given consideration.

- <u>Recommendation:</u> Increasing evaluation requirements by funders may improve the comprehensiveness of evaluations in the field generally. But this should only be done

183

with other supports and flexibility included, to increase the chance of the project's success at carrying out their evaluation. Supports include adequate and consistent funding, evaluation capacity, access to evaluation expertise and tools to develop and carry out a comprehensive evaluation.

Challenge: Bringing multiple pieces together into a coherent whole was described as a challenge by several projects studied. The complexity, scope, and mass amounts of data can further overwhelm projects carrying out work on a large scale.

- Recommendation: Developing a logic model or framework to guide the project and its evaluation can be helpful, as well as, working with an evaluator that has experience working on projects of this scale and level of complexity. These were described by projects studied as influencing their evaluation's success and comprehensiveness.

Promising Strategy: External evaluation experts were described repeatedly as playing a key role in guiding the development and the comprehensiveness of evaluations. Having access to external evaluators or evaluation experts outside of the project can improve the development and comprehensiveness of evaluations regardless of whether evaluations are ultimately carried out internally or externally.

- Recommendation: Projects should consider consulting with external evaluation experts by bringing them in during the planning phase of the evaluation and throughout the project to gather new ideas and gain a fresh perspective about their evaluation. This can be done even if evaluations are being carried out internally.

Promising Strategy: Building internal capacity may be a way to improve the quality of evaluations and keep them consistent for a longer period of time. Though, this would likely depend on the turnover rate within the organization considering this option.

- Recommendation: Projects should prioritize building internal capacity. Having a readily available team of trained evaluators can address both access and capacity issues for these organizations.

Promising Strategy: Flexibility to modify the evaluation as needed was identified by several projects as important to their success. Beyond having the option of flexibility, is the project having team members that are knowledgeable enough to adapt the evaluation as needed, whether the project or its context had changed, or for many other number of reasons. This factor appears to include a combination of projects having both flexibility and the capacity to make use of it as needed.

- Recommendation: Developing a recommendation for how a project becomes more flexible or leverages flexibility is more difficult than some of the other factors described previously. Flexibility may come from a combination of strong leadership and competency in the program, as well as, organizational and funder support. This may be something that funders and organizations need to develop expectations around, beyond the project working on their own to gain more flexibility. Building

internal capacity to know how to adapt the evaluation as needed may be increased as internal evaluation capacity is built or increased.

Information gathered through this dissertation can be used by multiple audiences to improve evaluations of multi-level interventions in the future. Multi-level projects, such as those studied, are not carried out quickly or simply, so having multiple suggestions for improvement, to help with the development, comprehensiveness and other areas of the evaluation, can be quite helpful. The ultimate goal of this dissertation was to provide the fields of evaluation and public health with information to increase comprehensiveness and quality of evaluations for multi-level interventions in the future.

Future work beyond this dissertation can go in several directions, including advancing recommendations provided, further study of elements within the Framework, or studying contextual elements beyond those in the Framework, which may also influence evaluation comprehensiveness. Additionally, future studies could look at other cases within these fields, or fields beyond tobacco control and childhood obesity, that are carrying out multi-level interventions, to see if similar patterns are found.

REFERENCES

Act 1220 Evaluation Project Representative. (2010). Personal Communication, September 21, 2010.

Albuquerque, M., Starr, G., Schooley, M., Pechacek, T., & Henson, R. (2003). *Advancing Tobacco Control Through Evidence-Based Programs.*

American Evaluation Association. Professional Development Webinars.   Retrieved November 10, 2010, from http://comm.eval.org/EVAL/coffee_break_webinars/Home/Default.aspx

American Evaluation Association. (2010). Evaluation 2010, AEA Annual Conference. Retrieved November 10, 2010, from http://www.eval.org/eval2010/default.asp

American Evaluation Association and CDC. (2010). Summer Evaluation Institute. Retrieved November 10, 2010, from http://www.eval.org/SummerInstitute10/default.asp

Ariza, C., Nebot, M., Tomas, Z., Gimenez, E., Valmayor, S., Tarilonte, V., et al. (2008). Longitudinal effects of the European smoking prevention framework approach (ESFA) project in Spanish adolescents. *European Journal of Public Health, 18*(5), 491-497.

Austin, W. D., Crankshaw, E., Donoghue, S., Farrelly, M. C., Holden, D., Loomis, B., et al. (2003). *Comprehensive Evaluation Plan for the New York Tobacco Control Program- Final Report*: RTI.

Beery, B., Ottoson, J., Green, L. W., Senter, S., Cahill, C., Greenwald, H., et al. (2007). *Final Report: Group Health Community Foundation's Evaluation of the Robert Wood Johnson Foundation Active Living Research Program.*

Benn, J., Burnett, S., Parand, A., Pinto, A., Iskander, S., & Vincent, C. (2009). Studying large-scale programmes to improve patient safety in whole care systems: Challenges for research. *Social Science & Medicine, 69*(12), 1767-1776.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge: Harvard University Press.

Brownson, R. C., Fielding, J. E., & Maylahn, C. A. (2009). Evidence-Based Public Health: A Fundamental Concept for Public Health Practice. *Annual Review of Public Health, 30*, 175-201.

California Department of Public Health. (2009). California Tobacco Control Update 2009: 20 Years of Tobacco Control in California. Sacramento, CA: California Tobacco Control Program.

California Tobacco Control Program Representative. (2010). Personal Communication, September 30, 2010.

Centers for Disease Control and Prevention. (1999a). Best Practices for Comprehensive Tobacco Control Programs. Atlanta, GA.

Centers for Disease Control and Prevention. (1999b). Framework for Program Evaluation in Public Health. *Morbidity and Mortality Weekly Report, 48*(RR11), 1-40.

Centers for Disease Control and Prevention. (2002). Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Economic Costs--United States, 1995-1999. *Morbidity and Mortality Weekly Report, 51*(14), 300-303.

Centers for Disease Control and Prevention. (2005). Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses--United States, 1997-2001. *Morbidity and Mortality Weekly Report, 54*(25), 625-628.

Centers for Disease Control and Prevention. (2009). Childhood Overweight and Obesity. Retrieved May 1, 2009, from http://www.cdc.gov/nccdphp/dnpa/obesity/childhood/index.htm

Centers for Disease Control and Prevention. National Center for Health Statistics. (2006). NHANES data on the Prevalence of Overweight Among Children and Adolescents: United States, 2003–2006. Retrieved May 1, 2009, from http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overweight/overwght_child_03.htm

Chappell, N., Funk, L., Carson, A., MacKenzie, P., & Stanwick, R. (2006). Multilevel community health promotion: How can we make it work? *Community Dev J, 41*(3), 352-366.

Claremont Graduate University. Professional Development Workshops. Retrieved November 14, 2010, from http://www.cgu.edu/pages/465.asp

Compton, D., Baizerman, M., & Farris, R. (2010). *A National Strategy to Enhance Quality of State Evaluations: The Case of the Division of Nutrition, Physical Activity and Obesity.* Paper presented at the American Evaluation Association, San Antonio, Texas.

Conference Report from Institute for Global Tobacco Control at Johns Hopkins Bloomberg School of Public Health. (2002). Evaluating comprehensive tobacco control interventions: challenges and recommendations for future action. *Tobacco Control, 11*(2), 140-145.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ, 337*, a1655.

Davison, K. K., & Birch, L. L. (2001). Childhood overweight: a contextual model and recommendations for future research. *Obesity Reviews, 2*, 159-171.

Department of Health Services. (2006). *California Obesity Prevention Plan: A Vision for Tomorrow, Strategic Actions for Today.* Sacramento, CA.

Doll, L., Bartenfeld, T., & Binder, S. (2003). Evaluation of Interventions Designed to Prevent and Control Injuries. *Epidemiol Rev, 25*(1), 51-59.

Economos, C., & Chatterji, M. (2010). *Bridging the Evidence Gap in Obesity Prevention- A Framework to Inform Decision Making.* Paper presented at the American Evaluation Association, San Antonio, Texas.

Farris, R., Khan, L., & Jernigan, J. (2010). *Development and Implementation of Community Strategies and Measurements to Prevent Obesity.* Paper presented at the American Evaluation Association, San Antonio, Texas.

Fay W. Boozman College of Public Health. (2010). Reports, Toolkits and Downloadable Materials. Retrieved September 13, 2010, from http://www.uams.edu/coph/reports/#Obesity

Foerster, S. B., Silver, L. D., Kohatsu, N. D., Frieden, T. R., Bassett, M. T., & Horton, M. B. (2007). Childhood Obesity on the Front Lines. *American Journal of Preventive Medicine, 33*(4S), S175-S177.

Fourney, A., Linares, A., Mitchell, P., & Sugerman, S. (2009). Final Report of the FFY 2008 Impact Evaluation. Sacramento, CA: California Department of Public Health, Cancer Control Branch.

Gentile, D. A., Welk, G., Eisenmann, J. C., Reimer, R. A., Walsh, D. A., Russell, D. W., et al. (2009). Evaluation of a multiple ecological level child obesity prevention program: Switch what you Do, View and Chew. *BMC Medicine, 7*, 49-60.

Green, L. W., Richard, L., & Potvin, L. (1996). Ecological foundations of health promotion. *American Journal of Health Promotion, 10*(4), 270-281.

Gregson, J., Foerster, S. B., Orr, R., Jones, L., Benedict, J., Clarke, B., et al. (2001). System, Environmental, and Policy Changes: Using the Social-Ecological Model as a Framework for Evaluating Nutrition Education and Social Marketing Programs with Low-Income Audiences. *Journal of Nutrition Education, 33*, S4-S15.

Griffith, J. R. (2009). Assessing Childhood Obesity Programs in Low-Socioeconomic and Diverse Communities. *Journal of the National Medical Association, 101*, 421-429.

Haw, S. (2007, September 10-11). *Scotland's Smokefree Legislation: Results from a comprehensive evaluation.* Paper presented at the Towards a Smokefree Society Conference, Edinburgh, Scotland. Available at: http://www.smokefreeconference07.com/programme.php.

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising Interventions as Events in Systems. *Am J Community Psychol, 43*, 267-276.

Hyland, A., Hassan, L. M., Higbee, C., Boudreau, C., Fong, G. T., Borland, R., et al. (2009). The impact of smokefree legislation in Scotland: results from the Scottish ITC Scotland/UK longitudinal surveys. *Eur J Public Health, 19*(2), 198-205.

Independent Evaluation Consortium. (2002). Final Report. Independent Evaluation of the California Tobacco Control Prevention and Education Program: Waves 1, 2, and 3 (1996-2000). Rockville: Maryland: The Gallup Organization.

Institute of Medicine. (2004). Preventing childhood obesity: health in the balance. Washington DC: National Academies Press.

Institute of Medicine. (2010). Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision Making. Washington DC: The National Academies Press.

Interview 1. Act 1220 Evaluation Project. (October 2010).

Interview 1. California Tobacco Control Program. (September 2010).

Interview 1. Network for a Healthy California. (August 2010).

Interview 1. New York Tobacco Control Program. (August 2010).

Interview 2. Act 1220 Evaluation Project. (October 2010).

Interview 2. California Tobacco Control Program. (September 2010).

Interview 2. Network for a Healthy California. (September 2010).

Interview 2. New York Tobacco Control Program. (August 2010).

Interview 3. Act 1220 Evaluation Project. (October 2010).

Interview 3. California Tobacco Control Program. (September 2010).

Interview 3. Network for a Healthy California. (September 2010).

Interview 4. Network for a Healthy California. (September 2010).

Jurg, M. E., Kremers, S. P. J., Candel, M., Van der Wal, M. F., & Meij, J. (2006). A controlled trial of a school-based environmental intervention to improve physical activity in Dutch children: JUMP-in, kids in motion. *Health Promotion International, 21*(4), 320-330.

Kothari, A., Edwards, N., Yanicki, S., Hansen-Ketchum, P., & Kennedy, M. A. (2007). Socioecological Models: Strengthening Intervention Research in Tobacco Control. *Drogues, santé et société, 6*(1), 1-24.

Levy, D. T., Hyland, A., Higbee, C., Remer, L., & Compton, C. (2007). The role of public policies in reducing smoking prevalence in California: Results from the California Tobacco Policy Simulation Model. *Health Policy, 82*(2), 167-185.

Louisiana Public Health Institute. (2009). *The Louisiana Campaign for Tobacco-Free Living (TFL) 2007-2008 Annual Report*.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco, CA: Jossey-Bass.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An Ecological Perspective on Health Promotion Programs. *Health Education Quarterly, 15*, 351-377.

Merzel, C., & D'Afflitti, J. (2003). Reconsidering Community-Based Health Promotion: Promise, Performance, and Potential. *Am J Public Health, 93*(4), 557-574.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Modayil, M. V., Cowling, D. W., Tang, H., & Roeseler, A. An evaluation of the California community intervention. *Tobacco Control, 19*(Suppl 1), i30-i36.

Mueller, N. B., Luke, D. A., Herbers, S. H., & Montgomery, T. P. (2006). The Best Practices: Use of the Guidelines by Ten State Tobacco Control Programs. *American Journal of Preventive Medicine, 31*(4), 300-306.

Nastasi, B. K., & Hitchcock, J. (2009). Challenges of Evaluating Multilevel Interventions., *American Journal of Community Psychology* (April 8, 2009 ed.).

Network Program Representative. (2010a). Personal Communication, August 5, 2010.

Network Program Representative. (2010b). Personal Communication, November 9, 2010.

Network Program Representative. (2010c). Personal Communication, October 27, 2010.

New York Tobacco Control Program Representative. (2010a). Personal Communication, July 15, 2010.

New York Tobacco Control Program Representative. (2010b). Personal Communication, November 1, 2010.

Novotny, T. E., & Siegel, M. B. (1996). California's Tobacco Control Saga. *Health Affairs, 15*(1), 58-72.

Nutbeam, D. (1998). Evaluating health promotion - progress, problems and solutions. *Health Promot. Int., 13*(1), 27-44.

O'Sullivan, R. G., Jay, M., & Powers, E. (2004). *Are focus group transcriptions necessary?* Paper presented at the American Evaluation Association Annual Conference, Atlanta, GA.

Ogden, C., Carroll, M., & Flegal, K. (2008). High Body Mass Index for Age Among US Children and Adolescents, 2003-2006. *JAMA, 299*, 2401-2405.

Patton, M. Q. (1997). *Utilization-Focused Evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Phillips, M. M., Raczynski, J. M., & Walker, J. F., for the Act 1220 Evaluation Team. (2004). *Establishing a Baseline to Evaluate Act 1220 of 2003: An Act of the Arkansas General Assembly to Combat Childhood Obesity.*

Phillips, M. M., Raczynski, J. M., & Walker, J. F., for the Act 1220 Evaluation Team. (2005). *Year Two Evaluation: Arkansas Act 1220 of 2003 to Combat Childhood Obesity.*

Phillips, M. M., Raczynski, J. M., & Walker, J. F., for the Act 1220 Evaluation Team. (2009). *Year Six Evaluation: Arkansas Act 1220 of 2003 to Combat Childhood Obesity.*

Ramanathan, S., Allison, K. R., Faulkner, G., & Dwyer, J. J. M. (2008). Challenges in assessing the implementation and effectiveness of physical activity and nutrition policy interventions as natural experiments. *Health Promot. Int., 23*(3), 290-297.

Richard, L., Lehoux, P., Breton, É., Denis, J.-L., Labrie, L., & Léonard, C. (2004). Implementing the ecological approach in tobacco control programs: results of a case study. *Evaluation and Program Planning, 27*(4), 409-421.

Robert Wood Johnson Foundation. Tobacco Policy Change Program. from http://www.tobaccopolicychange.org

Robert Wood Johnson Foundation. (2009). Arkansas Act 1220 Evaluation: A Project of Information for Action: School Policies to Prevent Childhood Obesity. Retrieved October 1, 2010, from http://www.rwjf.org/pr/product.jsp?id=38549

Roeseler, A., & Burns, D. (2010). The quarter that changed the world. *Tobacco Control, 19*(Suppl 1), i3-i15.

Rohrbach, L. A., Howard-Pitney, B., Unger, J. B., Dent, C. W., Howard, K. A., Cruz, T. B., et al. (2002). Independent Evaluation of the CA TCP: Relationships between Program Exposure and Outcomes 1996-1998. *American Journal of Public Health, 92*, 975-983.

RTI International. (2004). *First Annual Independent Evaluation of New York's Tobacco Control Program.*

RTI International. (2005). *Second Annual Independent Evaluation of New York's Tobacco Control Program.*

RTI International. (2006). *Third Annual Independent Evaluation of New York's Tobacco Control Program.*

RTI International. (2007). *Fourth Annual Independent Evaluation of New York's Tobacco Control Program.*

RTI International. (2009a). *2009 Independent Evaluation Report for the New York Tobacco Control Program*.

RTI International. (2009b). *Fifth Annual Independent Evaluation of New York's Tobacco Control Program*.

Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health, 56*(2), 119-127.

Saksvik, P. O., Nytro, K., Dahl-Jorgensen, C., & Mikkelsen, A. (2002). A process evaluation of individual and organizational occupational stress and health interventions. *Work and Stress, 16*(1), 37-57.

Sallis, J., Bauman, A., & Pratt, M. (1998). Environmental and Policy Interventions to Improve Physical Activity. *American Journal of Preventive Medicine, 15*(4), 379-397.

Samuels & Associates. UC Berkeley Center for Weight and Health. UCLA Center to Eliminate Health Disparities. Abundantia Consulting. Field Research Corporation. (2009). *Healthy Eating, Active Communities: Phase I Evaluation Findings, 2005-2008*.

Shenkin, J. D., & Jacobson, W. F. (2010). Using the Food Stamp Program and Other Methods to Promote Healthy Diets for Low-Income Consumers. *American Journal of Public Health, 100*(9), e1-3.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ, 336*(7656), 1281-1283.

Silva-Sanigorski, A., Prosser, L., Carpenter, L., Honisett, S., Gibbs, L., Moodie, M., et al. (2010). Evaluation of the childhood obesity prevention program Kids - 'Go for your life'. *BMC Public Health, 10:288*.

Smedley, B., & Syme, S. (2000). *Promoting health: Intervention strategies from social and behavioral research*. Institute of Medicine. Washington DC: National Academy Press.

Spoth, R. L., & Greenberg, M. T. (2005). Toward a Comprehensive Strategy for Effective Practitioner-Scientist Partnerships and Larger-Scale Community Health and Well-Being. *American Journal of Community Psychology, 35*(3-4), 107-126.

Stake, R. E. (2006). *Multiple Case Study Analysis*. New York, NY: The Guilford Press.

Starr, G., Rogers, T., Schooley, M., Porter, S., Wiesen, E., & Jamison, N. (2005). Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs. Atlanta, GA: Centers for Disease Control and Prevention.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.

Sugerman, S. (2008). Successful Evaluation Strategies: California. Sacramento, CA: California Department of Public Health, Cancer Prevention and Nutrition Section.

Tang, H., Cowling, D. W., Koumjian, K., Roeseler, A., Lloyd, J., & Rogers, T. (2002). Building Local Program Evaluation Capacity Toward a Comprehensive Evaluation. *New Directions for Evaluation, 95*, 39-56.

The Evaluator's Institute. TEI Certificate Program.   Retrieved November 14, 2010, from http://tei.gwu.edu/certificate.htm

The Evaluator's Institute. (2010). TEI Courses in San Francisco 2011.   Retrieved November 14, 2010, from http://tei.gwu.edu/courses_sanfrancisco.htm

Trickett, E. J. (2009). Multilevel Community-Based Culturally Situated Interventions and Community Impact: An Ecological Perspective, *American Journal of Community Psychology*.

U.S. Department of Health and Human Services. (2001). Introduction to program evaluation for comprehensive tobacco control programs. Atlanta, GA: Centers for Disease Control and Prevention. Office on Smoking and Health.

U.S. Department of Health and Human Services. (2002). Physical Activity Evaluation Handbook. . Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

U.S. Department of Health and Human Services. (2004). Health, United States, 2004, with Chartbook on Trends in the Health of Americans. Hyattsville, MD: U.S. Department of Health and Human Services. Centers for Disease Control and Prevention.

U.S. Department of Health and Human Services. (2005). Introduction to program evaluation for public health programs: A self-study guide. Atlanta, GA: Centers for Disease Control and Prevention.

U.S. Department of Health and Human Services. (2008). Introduction to Process Evaluation in Tobacco Use Prevention and Control. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

U.S. Department of Health and Human Services. Office of Disease Prevention and Health Promotion. (2005). Healthy People 2010 Fact Sheet.   Retrieved May 12, 2009, from http://www.healthypeople.gov/About/hpfact.htm

University of Michigan. Institute for Social Research Summer Institute.   Retrieved November 14, 2010, from http://si.isr.umich.edu/

Warner, K. E. (2009). Tobacco Research Methodology: First Things First. *Cancer Epidemiology Biomarkers & Prevention, 18*(12), 3140-3142.

Weber, M. D., Bagwell, D. A., Fielding, J. E., & Glantz, S. A. (2003). Long-term Compliance with California smoke-free workplace law. *Tobacco Control, 12*, 269-273.

Williams, B., & Hummelbrunner, R. (2010). *Systems Concepts in Action: A Practitioner's Toolkit.* Palo Alto, CA: Stanford University Press.

Yang, H., Shen, J., Cao, H., & Warfield, C. (2004). Multilevel Evaluation Alignment: An Explication of a Four-Step Model. *American Journal of Evaluation, 25*(4), 493-507.

Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th ed.). Thousand Oaks, CA: Sage.