# NEW PUBLIC MANAGEMENT AND GOVERNANCE COLLIDE: FEDERAL-LEVEL PERFORMANCE MEASUREMENT IN NETWORKED PUBLIC MANAGEMENT ENVIRONMENTS

A Dissertation
Presented to
The Academic Faculty

By

Amy S. DeGroff, MPH

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Public Policy

Georgia Institute of Technology
&
Georgia State University
May 2009

NEW PUBLIC MANAGEMENT AND GOVERNANCE COLLIDE:
FEDERAL-LEVEL PERFORMANCE MEASUREMENT IN NETWORKED PUBLIC
MANAGEMENT ENVIRONMENTS

Approved by:

Dr. Theodore Poister, Advisor
Andrew Young School of Policy Studies
Georgia State University

Dr. Gordon Kingsley
School of Public Policy
Georgia Institute of Technology


Dr. Patricia Reeves
School of Social Work
University of Georgia

Dr. Judith Ottoson
Adjunct Faculty
San Francisco State University


Dr. John Thomas
Andrew Young School of Policy Studies
Georgia State University

Date Approved: May 8, 2009

To my parents, Patricia and Richard DeGroff

# ACKNOWLEDGEMENTS

I am deeply grateful to the members of my dissertation committee, Drs. Theodore Poister, Gordon Kingsley, Judith Ottoson, Patricia Reeves, and John Thomas, for their support and mentorship throughout my dissertation research. I am especially appreciative of my chair, Dr. Ted Poister, for his willingness to advise and guide me through the process, providing thoughtful insight and direction along the route and for otherwise being just an incredibly kind and generous person. I would like to thank Dr. Kingsley for instilling an interest in theory during my very first class in the doctoral program. To Dr. Ottoson, I am forever indebted for her mentorship in program evaluation, the opportunity to work with her as a doctoral student, and for her strong belief in me. For my interest in qualitative research, I owe much to Dr. Reeves. I was incredibly fortunate to discover the qualitative research program at the University of Georgia and even more fortunate to meet Dr. Reeves who has been a wonderful teacher, mentor, and advisor throughout my dissertation research. And finally, I want to thank Dr. Thomas who has encouraged my interests more broadly in public management and program implementation.

I would also like to acknowledge colleagues with the four programs at the Centers for Disease Control and Prevention who enthusiastically allowed me to include their programs in the study: Anita McLees, Craig Thomas, Dayne Collins , Betty Apt, Faye Wong, Sheila Porter, and Martha Engstrom. In addition, I am deeply grateful to all those who participated in the study and for their willingness to honestly share their experiences and perspectives. This study would not have been possible without them.

I would like to recognize my colleagues at CDC who have provided incredible support since the start of my doctoral program. Nancy Lee, Susan True, Rosemarie

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ACS | AMERICAN CANCER SOCIETY |
| APHL | ASSOCIATION OF PUBLIC HEALTH LABORATORIES |
| ASPR | ASSISTANT SECRETARY FOR PREPAREDNESS AND RESPONSE |
| ASSIST | AMERICAN STOP SMOKING INTERVENTION STUDY FOR CANCER PREVENTION |
| ASTHO | ASSOCIATION OF STATE AND TERRITORIAL HEALTH OFFICERS |
| ATMC | ADDRESSING TOBACCO IN MANAGED CARE |
| ATS | ADULT TOBACCO SURVEY |
| BRFSS | BEHAVIORAL RISK FACTOR SURVEILLANCE SYSTEM |
| CARE | RYAN WHITE COMPREHENSIVE AIDS RESOURCES EMERGENCY ACT |
| CATS | CALIFORNIA ADULT TOBACCO SURVEY |
| CDC | CENTERS FOR DISEASE CONTROL AND PREVENTION |
| CFO | CHIEF FINANCIAL OFFICERS ACT |
| CMS | CENTERS FOR MEDICARE AND MEDICAID SERVICES |
| CBO | COMMUNITY BASED ORGANIZATION |
| CCID | COORDINATING CENTER FOR INFECTIOUS DISEASES |
| COTPER | COORDINATING OFFICE FOR TERRORISM PREPAREDNESS AND EMERGENCY RESPONSE |
| CPS TUS | CURRENT POPULATION SURVEY: TOBACCO USE SUPPLEMENT |
| CSPS | COMPREHENSIVE STD PREVENTION SYSTEMS |
| CSTE | COUNCIL OF STATE AND TERRITORIAL EPIDEMIOLOGISTS |
| CTFK | CAMPAIGN FOR TOBACCO FREE KIDS |
| DCPC | DIVISION OF CANCER PREVENTION AND CONTROL |
| DDT | DIVISION OF DIABETES TRANSLATION |
| DIS | DISEASE INVESTIGATION SPECIALISTS |
| DoD | DEPARTMENT OF DEFENSE |
| DQIG | DATA QUALITY INDICATOR GUIDE |
| DHS | DEPARTMENT OF HOMELAND SECURITY |
| DSLR | DIVISION OF STATE AND LOCAL READINESS |
| DSTDP | DIVISION OF STD PREVENTION |
| FDA | FOOD AND DRUG ADMINISTRATION |
| FEMA | FEDERAL EMERGENCY MANAGEMENT AGENCY |
| FRN | FEDERAL REGISTER NOTICE |

| FY | FISCAL YEAR |
|---|---|
| GAO | U.S. GENERAL ACCOUNTING OFFICE, GOVERNMENT ACCOUNTABILITY OFFICE |
| GPP | GOVERNMENT PERFORMANCE PROJECT |
| GPRA | GOVERNMENT PERFORMANCE AND RESULTS ACT OF 1993 |
| GSU | GEORGIA STATE UNIVERSITY |
| HHS | DEPARTMENT OF HEALTH AND HUMAN SERVICES |
| HIPAA | HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT |
| HMA | HIGH MORBIDITY AREA |
| HRSA | HEALTH RESOURCES AND SERVICES ADMINISTRATION |
| HSPD | HOMELAND SECURITY PRESIDENTIAL DIRECTIVE |
| HSEEP | HOMELAND SECURITY EXERCISE AND EVALUATION PROGRAM |
| IBM | INTERNATIONAL BUSINESS MACHINES CORPORATION |
| IHS | INDIAN HEALTH SERVICES |
| IMPACT | INITIATIVES TO MOBILIZE FOR THE PREVENTION AND CONTROL OF TOBACCO USE |
| IOM | INSTITUTE OF MEDICINE |
| IPP | INFERTILITY PREVENTION PROJECT |
| JTPA | JOB TRAINING PARTNERSHIP ACT |
| KOI | KEY OUTCOME INDICATOR |
| MBO | MANAGEMENT BY OBJECTIVES |
| MDEs | MINIMUM DATA ELEMENTS |
| MFR | MANAGEMENT FOR RESULTS |
| MLS | MEDICAL AND LABORATORY SERVICES |
| MSA | MASTER SETTLEMENT AGREEMENT |
| NACCHO | NATIONAL ASSOCIATION OF CITY AND COUNTY HEALTH OFFICIALS |
| NACDD | NATIONAL ASSOCIATION OF CHRONIC DISEASE DIRECTORS |
| NBCCEDP | NATIONAL BREAST AND CERVICAL CANCER EARLY DETECTION PROGRAM |
| NCCCP | NATIONAL COMPREHENSIVE CANCER CONTROL PROGRAM |
| NCCDPHP | NATIONAL CENTER FOR CHRONIC DISEASE AND HEALTH PROMOTION |
| NCHHSTP | NATIONAL CENTER FOR HIV/AIDS, VIRAL HEPATITIS, STD, AND TB PREVENTION NCI – NATIONAL CANCER INSTITUTE |
| NCSD | NATIONAL COALITION OF STD DIRECTORS |
| NDPCP | NATIONAL DIABETES PREVENTION AND CONTROL PROGRAM |
| NIH | NATIONAL INSTITUTES OF HEALTH |
| NPM | NEW PUBLIC MANAGEMENT |
| NPR | NATIONAL PERFORMANCE REVIEW |
| NTCP | NATIONAL TOBACCO CONTROL PROGRAM |

| OMB | OFFICE OF MANAGEMENT AND BUDGET |
|---|---|
| OMEB | OUTCOME MONITORING AND EVALUATION BRANCH |
| OSH | OFFICE ON SMOKING AND HEALTH |
| PAHPA | PANDEMIC AND ALL HAZARDS PREPAREDNESS ACT OF 2007 |
| PAN FLU | PANDEMIC INFLUENZA |
| PART | PERFORMANCE ASSESSMENT AND RATING TOOL |
| PHEP | PUBLIC HEALTH EMERGENCY PREPAREDNESS |
| PL | PUBLIC LAW |
| PMA | PRESIDENTIAL MANAGEMENT AGENDA |
| PPBS | PROGRAM, PLANNING, AND BUDGETING SYSTEM |
| SNS | STRATEGIC NATIONAL STOCKPILE |
| STD | SEXUALLY TRANSMITTED DISEASES |
| U.K. | UNITED KINGDOM |
| U.S. | UNITED STATES |
| USPSTF | U.S. PREVENTIVE SERVICES TASK FORCE |
| WISEWOMAN™ | WELL INTEGRATED SCREENING AND EVALUATION FOR WOMEN ACROSS THE NATION |
| ZBB | ZERO BASED BUDGETING |

**SUMMARY**

The purpose of this research was to investigate the implications of networked public management on the design, implementation, and utilization of federal performance measurement systems. A multiple, instrumental case study of four public health programs funded by CDC and implemented nationally through vertical and horizontal network structures was conducted. Cross-case findings suggest that the networked implementation structures for the four federal-level, public health programs have important implications for the design of the performance measurement systems. Specifically, the performance measurement systems were affected by four consequences of the implementation networks: the political influence of collaborative stakeholders; network variability; dependencies on voluntary, horizontal network partners to achieve outputs and outcomes; and jointly produced outcomes that compromise assigning agency-specific attribution and accountability. While these four factors did not deter the *use* of performance measurement by any of the programs, all had important consequences for the development and subsequent design of the performance measurement systems, including limiting the choice and types of measures, level of measurement, potential uses of the measures, and resources needed to implement and support the systems.

**CHAPTER 1**

**INTRODUCTION**

The purpose of this research was to investigate the implications of networked

public management on the design, implementation, and utilization of federal performance

measurement systems through a multiple, instrumental case study of four national public

health programs at the Centers for Disease Control and Prevention (CDC). Performance

measurement has gained prominence at CDC over the past 15 years, advanced by the

influence of the New Public Management (NPM) and its accompanying reforms on

national-level public health programs. One emphasis of NPM, which advocates the

integration of corporate values in government, is increased accountability; that is, holding

government programs responsible for achieving program outcomes and demonstrating

their "value added" (Frederickson 2003). Performance measurement, a management tool

involving the ongoing monitoring of indicators of organizational and program

performance (Poister 2003), has been widely promoted and adopted as a tool to enhance

accountability while also contributing to program improvement (Kelly 2002).

Additionally, the 1993 Government Performance and Results Act (GPRA) designated

performance measurement a statutory requirement, mandating that federal agencies

develop annual performance plans and submit annual performance reports inclusive of

performance data.

Because of my responsibilities at CDC related to GPRA and, more recently, the

Office of Management and Budget's (OMB) program assessment and rating tool (PART)

process, I felt compelled to question the implications of the complex public management

context in which public health programs are implemented on the design and

implementation of federal-level, performance measurement systems. This situation

involves extensive decentralization among vertically-integrated, intergovernmental

partners at the state and local level, as well as private and non-profit agencies that operate

on more horizontal levels. In contrast to the traditional hierarchical approach to

governing, this environment is better described as involving "networks." Fundamentally,

public health problems reflect the complex intersection of social, environmental,

behavioral, and biological factors (Institute of Medicine 2003), so achieving effective

outcomes necessitates collaboration and coordination among networked partners across a

variety of sectors. In fact, CDC readily acknowledges that no one agency or program can

effectively achieve desired health outcomes and strongly endorses a collaborative

approach in public health practice[1].

    This approach to the delivery of public health services by CDC, a federal

government agency, demonstrates organizations operating in a governance framework.

Governance involves the processes of administration and management in an environment

characterized by multiple societal forces, that is, in a context where broad relationships

exist between government and its political, administrative, and social environment (Kettl

2002; Stoker 1998). A central characteristic of governance involves networks, in which

more traditional, vertical, intergovernmental arrangements exist alongside horizontal,

voluntary associations (Hill and Lynn 2005). Kettl (2002) suggests that networked

governance has emerged, in part, to facilitate more integrated and coordinated service

delivery in response to complex social problems like those in public health. In a report for

the International Business Machines Corporation (IBM) Center for The Business of

---

[1] http://www.cdc.gov/about/organization/mission.htm. (accessed 02/01/09).

Government, Kettl (2005) even highlights CDC as an example of an agency shifting to a more flexible, networked approach. Empirical studies describing public management in networked governance structures are just beginning to appear in the literature.

The practice of federal-level performance measurement has primarily been described based on the traditional view of bureaucratic hierarchy, where established lines of formal authority and, therefore, accountability, are relatively straightforward (Frederickson 2003; Frederickson and Frederickson 2006; Perrin 1998). The implications of networked governance are noteworthy, especially the severe compromising of hierarchical authority, which impose new and greater challenges on the design and implementation of performance measurement, especially at the national level. Frankly, little research has focused on what Frederickson and Frederickson (2006) suggested is "the most critical factor accounting for the effectiveness of performance measurement – third-party government" (p. 10). By better understanding the influence of networked governance on the design and implementation of federal performance measurement systems, its use as a management tool may be improved in the future. Subsequently, this research contributes to a fledgling theory building in the area of performance measurement (Jennings and Haist 2004), an undertaking encouraged by some key leaders in the field (Wholey 1999).

In this first chapter, relevant background supporting the dissertation study is provided. An important research gap is identified and the research purpose is stated. In addition, background is provided on CDC, where the research was conducted, the theoretical basis for the research is provided, and the research questions are detailed. Finally, a brief summary of the methodology is included.

## 1.1 Background

### 1.1.1 The Emergence of Networked Governance

Since the early 1990s, "governance" has emerged as a new organizing concept for public administration and management in response to various social and political influences (Agranoff and McGuire 2001a; Frederickson and Smith 2003). The term governance has been part of common parlance for decades, but more recently, the term has been used to describe a meaningful shift in how we are governing (Frederickson and Smith 2003). Indeed, governance has emerged as an organizing framework intended to capture an important turn in the practice of public administration and management, that is, the processes of governing (Heinrich, Hill, and Lynn 2004). In particular, the more recent focus on governance signals a change in the relationship between the state and society and reflects the broader relationships that exist between government and its political, administrative, and social environment (Kettl 2002). This view of governance recognizes that government operates within a vast environment of multiple societal forces rather than viewing it as a 'stand alone' institution (Stoker 1998).

Governance provides the overall context for this study for two reasons. First, a dominant feature of governance involves third parties and networks (Lynn, Heinrich, and Hill 2000). Within the governance framework, the government's work is carried out, in part, through interdependent networks rather than traditional hierarchy alone. In governance, it is networks, rather than the formal institutions of government, that dominate public policy and are increasingly responsible for delivering public services (Frederickson and Smith 2003; Milward and Provan 2006; Peters and Pierre 1998). Second, using governance as the context for this study reflects the larger backdrop within

4

which NPM, with its emphasis on results and its promotion of performance measurement, is practiced. Governance, then, provides the overall political and organizational context for this study and, as such, becomes relevant in our understanding about how networked management affects the practice of performance measurement.

Scholars have suggested that the move to networked governance is associated, in part, with the devolution of government as a direct provider of public programs to the "hollow state," a situation in which policy implementation is dominated by third parties (Boyt 2005; Milward and Provan 2000). As noted earlier, networks may also have emerged as potentially more effective organizational structures to support the collaborative relationships necessary to address complex social problems (Agranoff and McGuire 2001b; Keast et al. 2004; O'Toole 1997). Harmon and Mayer (1986) dubbed these "wicked" problems, those complex problems influenced by a variety of often interrelated factors. Such problems demand coordinated responses across sectors and levels of government in order to assure integration and efficiency in the delivery of varied programs and services (Agranoff and McGuire 2001a).

Consequently, a networked model of governance has emerged and public policy implementation now occurs in the context of actors who are mutually dependent and who are linked by a web of relationships between various organizations, sectors, and levels of government (Salamon 2002). In this environment, government represents just one of many actors and institutions involved in addressing issues and concerns of the polity (Agranoff and McGuire 2001a). This shift has led to what Kettl (2002) has termed the "fuzzy boundaries" problem wherein new challenges are posed by the loss of hierarchy's clear lines of responsibility and the state's loss of sole authority. In a network model,

interdependent relationships exist between the state and other third-party actors from the private and non-profit sectors resulting in power dependencies (Stoker 1998). These interdependencies imply that in a networked context, government loses its capacity for direct control and must rely on strategies of influence and leverage (Peters 2001). However, as Milward and Provan (2006) suggest, "networks have proven to be a very valuable public management tool . . . because they are the only organizational forms that can operate horizontally, across a range of organizations, and integrate the strengths and talents of a variety of organizations in the public, nonprofit, and for-profit sectors to effectively address critical public problems" (p. 7).

Additionally, the collaborative approach inherent in networked governance reflects the recognition that outcomes, both individual and organizational, result from the actions of varied actors addressing a multitude of factors that are contributing to the problem. For results to be achieved in this context, leaders must effectively collaborate across public, private, and non-profit sectors as well as across levels of government (Abramson, Breul, Kamensky 2006). This structure presents what has been called the "joint production problem," the challenge of coordinating program implementation across a potentially broad and decentralized service implementation network, each with its own goals and priorities (Milward and Provan 2004). As noted above, it is well recognized that a complex interplay of factors contribute to public health problems and that representatives from various sectors and levels of government must be engaged to effectively address them.

But both the fuzzy boundary and joint production problems lead to a central challenge or dilemma of governance, that of *accountability* (Kettl 2002; Peters and Pierre

1998; Stoker 1998). In fact, challenges to accountability abound in networks as program implementation is decentralized, traditional hierarchical authority is compromised, political resources are shared, and monitoring channels are diffused and become unreliable (Goldsmith and Eggers 2005; O'Toole 1997; Peters 2001; Posner 2002). Melissa Stone expressed valid concerns about accountability in these contexts asking, "How must we conceptualize accountability when the actual implementers of public policy are removed from government agencies and have their own notions of to whom and for what they are accountable?" (Cited in Boyt 2005, 537).

### 1.1.2 NPM and the Rise of Performance Measurement

At the same time that networked governance has achieved higher levels of prominence, performance measurement, with an emphasis on outcomes, also has gained attention given the results-based management reforms of recent years associated with NPM (Behn 2003; Jennings and Haist 2004). NPM emerged in the late 1980s and early 1990s; Osborne and Gaebler's 1992 best seller, *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, is seen as a leading text for the movement. NPM advocates that government adopt market-based approaches perceived as more efficient and effective (Osborne and Gaebler 1992). Peters and Pierre (1998) suggested that in contrast to governance, which is process-oriented and based in democratic theory, NPM is more appropriately viewed as an ideological movement emphasizing responsibility and accountability through a focus on outcomes and results. They argue that NPM and governance are distinct approaches with a number of fundamental differences (Peters and Pierre 1998). The principles of NPM have been widely adopted in practice and translated into government reform that emphasizes

decentralized government, increased contracting through third parties, and an expanded

use of policy tools[2] to increase competition (Frederickson and Smith 2003). In fact, these

three aspects of NPM are likely to support a networked governance structure.

In contrast to governance, where accountability is seen as problematic, NPM

claims it as a particular strength (Osborne and Gaebler 1992; Peters and Pierre 1998).

Outcome-based performance is so central to NPM that it is often recognized as "results-

oriented government." Because of its emphasis on outcomes over processes or outputs as

a means to assess policy implementation, NPM shifts the focus of accountability from the

elected official to the administrator, holding public managers responsible for results

(Dubnick 2005). NPM's emphasis on results and accountability led to a defining moment

for the movement, the passing of the 1993 GPRA (Light 1997). GPRA was intended to

primarily addresses perceived weaknesses in government management and introduce

greater accountability for results (Frederickson 2003; GAO 2004).

The influence of NPM and GPRA has spurred what some have viewed as an

"accountability movement" characterized by the widespread implementation of

performance measurement systems (Behn 2003; Blalock and Barnow 2001; Coplin,

Merget, and Bourdeaux 2002; Grizzle 2002; Radin 2006; Schick 2001). Through the

influence of NPM, performance measurement is now recognized as a fundamental public

administrative reform (Frederickson and Frederickson 2006; Kelly 2002; Radin 2006).

While evidence of the effectiveness of performance measurement is still mixed, it has

been widely adopted and supported in practice (Barry 2000; Hatry 1997; Kelly 2002;

Turnock 2000; Wholey 2002). In addition, there has been substantial political and policy

---

[2] Policy tools or instruments reflect the activities performed by a federal agency to address public problems or achieve objectives (Frederickson 2001; Salamon 2002).

impetus for performance measurement, such as GPRA and its inclusion in former President Bush's Presidential Management Agenda (PMA) PART (GAO 2005). The overriding benefit and promise of performance measurement is its contribution toward more sound public management, including increased accountability and transparency. "What gets measured gets done," and, "You can't manage what you don't measure," are common adages reflecting the potential of performance measurement to support improved management and leadership (Osborne and Gaebler 1992). Poister (2003) suggests that the inherent logic of performance measurement engenders support for it: " … performance measurement systems provide incentives for organizations and programs to perform at higher levels, and this is the core of the logic underlying the use of monitoring systems as performance management tools" (p. 99).

## 1.2 Statement of the Problem and Purpose of the Study

In a report from the IBM Center for the Business of Government (Abramson, Breul, and Kamensky 2006), performance management and network governance are identified as two of six trends transforming government. Both represent central features of contemporary public management. Researchers have suggested, however, that performance measurement is based on assumptions rooted in a hierarchical model of bureaucracy (Frederickson 2003; Perrin 1998; Radin 2006), implying that its practice has largely been described based on a context where accountability structures are considerably more fixed than in a networked model. As previously noted, the NPM reforms emphasize results-based performance measurement as a means for assuring accountability (Jennings and Haist 2004). Unfortunately, the performance measurement literature has only begun to address the design and implementation of performance

measurement systems in networked environments (Frederickson and Frederickson 2006; Goddard and Mannion 2004; Mandell and Keast 2006; Voets, De Rynck, and Wouter 2006). And, as noted above, accountability has been identified as one of the greatest challenges facing networked governance. Radin (2006) has suggested that advocates of performance measurement make a number of assumptions that goals can be clearly defined and are the responsibility of specific actors, outcomes can be quantified and measured, data are available and accurate, and an actor will have authority to act on the results. These assumptions and their implications for the practice of performance measurement are compromised in the networked context. As reflected in the title of this dissertation, a potential conflict emerges when accountability for results meets a networked context (Milward and Provan 2006; Radin 2006).

Given such challenges, the design and implementation of federal performance measurement systems is likely more difficult in networked than hierarchical settings. And, while much has been written about the NPM reforms and the centrality of performance measurement to improved public management and accountability, the literature does not address the tensions and/or opportunities created based on its use in a framework of networked governance. Because of this gap in the literature, **the purpose of this research was to investigate the implications of networked public management on the design, implementation, and utilization of federal performance measurement systems through a multiple, instrumental case study of four national public health programs at the Centers for Disease Control and Prevention (CDC).**

The proposed research builds on earlier work by David Frederickson (2003), a qualitative grounded theory study of GPRA's implementation in five U.S. Health and

Human Services (HHS) agencies Health Resources and Services Administration (HRSA),

Centers for Medicare and Medicaid Services (CMS), National Institutes for Health

(NIH), Indian Health Services (IHS), and the Food and Drug Administration (FDA).

Although Frederickson started the study intending to explore how these five agencies

measure their performance under GPRA, the research evolved in focus to trying to better

understand how well these agencies measure third party performance (Frederickson and

Frederickson 2006). His findings were important because he identified a number of

factors influencing the implementation of federal performance measurement suggesting

that third-party governance and "network articulation[3]," were the most influential factors

(Frederickson and Frederickson 2006).

This dissertation research extends and builds on Frederickson's work. It is

intended to provide a better understanding of how federal public health programs are

grappling with the challenges posed by networked governance in developing program-

specific, performance measurement systems. Based on the research, recommendations are

made to enhance the future use of performance measurement in federal public health

programs implemented in networked environments. In addition, this research contributes

to early theory building in the area of performance measurement initiated by Jennings

and Haist (2004).

### 1.3 Study Context: CDC's National Public Health Programs

CDC, an agency of HHS, is the nation's leading public health institution. Several

of CDC's national public health programs provide rich examples of initiatives committed

---

[3] Frederickson and Frederickson (2006) define articulated as "extent to which separate organizations or institutions in a network are coupled, fit together, linked, or combined and the nature and quality of those connections," (p. 8).

11

to strengthening their use of performance measurement while being implemented through a networked governance model. Like other federal agencies, CDC has been influenced by NPM through both GPRA and PART and has worked to develop strategic plans and related performance measurement systems. Staff at CDC have developed and reported GPRA measures, programs have been reviewed by OMB's PART process, and evaluators have developed program-specific performance measurement systems for varied purposes. Goal-based strategic planning efforts are also underway at the agency level, an effort seen as generally reinforcing performance management efforts at CDC.

In addition, CDC increasingly relies on networks and partnerships to solve multifaceted public health problems. Such problems involve social, medical, economic, political, and moral dimensions which require multiple-level responses addressing individuals, communities, and populations. CDC's "Future's Initiative" is a recent leadership effort to both increase CDC's capacity to effectively collaborate with and leverage its partners while also improving accountability. Under this initiative, CDC has defined six key strategies to guide its work including two that align with the prominent issues of this study: (1) emphasizing leadership through leveraging partnerships and networks and (2) stressing accountability to sustain public trust and confidence[4]. Accountability is further emphasized as one of CDC's three core values in conjunction with respect and integrity[5]. Other federal organizations support CDC's priorities including a 2003 report from the Institute of Medicine (IOM) (IOM 2003). In that report, IOM calls for constructing an intersectoral public health system based on collaborative

---

[4] http://www.cdc.gov/osi/goals/strategicimperatives.html (accessed 02/02/09).
5 http://www.cdc.gov/about/organization/mission.htm (accessed 02/02/09).

arrangements that emphasize partnerships and networks (IOM 2003). This recommendation is not intended to disregard the importance of the traditional, intergovernmental pubic health structure that is comprised of federal, state, and local health agencies. From a constitutional and historical perspective, state and local agencies have a central responsibility in assuring the public's health (IOM 2003). In fact, many of CDC's federal public health programs are structured around the decentralized, intergovernmental relationships that exist between CDC and its state partners, who in turn, frequently fund public health agencies at more local levels.

There is, however, an increasing emphasis at CDC to employ mechanisms, often informal, that engage other sectors in addressing public health concerns. The premise underlying this broader networked approach is two-fold. First, it is expected that strategic collaboration will ensure that greater resources of all kinds (e.g., monetary, expertise, in-kind contributions) are brought to bear on the problem. And second, due to the potential of partner-generated synergy, networked approaches are more likely to have greater influence on health outcomes than what government could accomplish alone. It is noteworthy that the public health literature acknowledges the importance of partnerships and coalitions in achieving public health goals that require resources and capabilities beyond any single agency or sector (Lasker, Weiss, Miller 2001; Shortell 2002; Wandersman, Goodman, and Butterfoss 1997). While the move to a broader, networked public management approach is considered advancement for the field of public health, it also underscores the complexity of applying performance measurement to federal-level, public health programs.

As noted above, CDC typically implements national public health programs through decentralized, intergovernmental relationships. Some CDC programs support upwards of sixty or more grantees (e.g., CDC's National Breast and Cervical Cancer Early Detection Program and National Diabetes Prevention and Control Program). Most often, cooperative funding agreements, a type of grant, are used as a policy tool, a choice that provides grantees some discretion in program implementation. The "cooperative" nature of the policy tool provides program-level managers greater flexibility to structure programs in ways best suited to the needs and culture of their communities while preserving some level of control for CDC. This reinforces a long standing tradition in public health of more localized, public health service delivery.

In addition, based on CDC's institutional priorities for increased collaboration, many CDC funding announcements encourage or even mandate broad partnerships at the state and local level aimed at increasing coordinated efforts across networks. For instance, as part of CDC's National Comprehensive Cancer Control Program (NCCCP), health departments facilitate comprehensive, state-, tribe-, or territorial-wide planning for cancer prevention and control. Planning efforts typically engage state and local public health agencies, representatives of the health care and pharmaceutical industries, leading non-profit organizations such as the American Cancer Society (ACS) and Lance Armstrong Association, universities and other research institutions, cancer treatment centers, and so forth. Goals established by these planning groups often reflect the need for diverse implementation approaches on the part of coalition members. Although these networks are more informal in nature than some described in the network management

literature (e.g., Milward and Provan 1998; Milward and Provan 2006), they provide an example of public health programs implemented in a networked context.

In summary, CDC's public health programs provide the opportunity to explore the confluence of networked governance and performance measurement. As described above, in order to more effectively address the "wicked" public health problems facing our country and the globe more broadly, CDC is increasingly adopting networked public management approaches. At the same time, federal programs are struggling to develop and implement effective and useful performance measurement systems to answer calls for accountability in this more complex context.

## 1.4 Theoretical Framework

Although the literature related to performance measurement is vast, little effort has been undertaken to develop related theory (Frederickson 2003; Jennings and Haist 2002, 2004). More recently, however, Edward Jennings and Meg Haist (2004) have made an effort intended to encourage systematic empirical research. In a chapter of *The Art of Governance: Analyzing Management and Administration,* edited by Patricia Ingraham and Laurence Lynn (2004), Jennings and Haist propose twenty-five hypotheses predicting the impact of performance measurement. Their work is positioned in a larger logic or framework of governance that is proposed by Lynn, Heinrich, and Hill in the book's first chapter.

The logic of governance offered by Heinrich, Hill, and Lynn (2004) suggests that outcomes or outputs (i.e., performance) are a function of five general classes of variables environment, consumer characteristics, treatments, structures, and managerial roles or actions. That is, $P = E + C + T + S + M$. The authors clarify that they are not offering a

theory per se, "but rather an organizing device for conceptualizing and interpreting

empirical research," (Heinrich, Hill, and Lynn 2004, p. 7). Given this model, Jennings

and Haist recognize that performance management falls under the managerial efforts (i.e.,

"M") as a mediator of performance. The authors noted, "The broad features of the

governance model suggest that performance measurement or management is only one of

many factors that shape the impacts of a policy. As such, the effects are likely to be at the

margin" (Jennings and Haist 2004, p. 175).

In offering their hypotheses for performance measurement, Jennings and Haist

build on earlier work by both Gormley and Weimer (1999) and Dixit (2002). Gormley

and Weimer draw their theoretical ideas about performance measurement from literature

related to accountability, incentives, and competition. Dixit's work (2002) is primarily

drawn from literature exploring the role of incentives, that is, how a focus on results via

performance measures serves as an incentive to influence and direct performance.

Jennings and Haist's twenty-five hypotheses for future empirical study consider

incentives and accountability (nine hypotheses), organizational characteristics (eight

hypotheses), political context (three hypotheses), agency type (three hypotheses), and

leadership (two hypotheses). See appendix A for a full listing of the twenty-five proposed

hypotheses.

Given my particular research interests and the intended study context, I have used

two of three hypotheses offered by Jennings and Haist related to agency type as a

theoretical basis for this study. These two hypotheses were developed based on James Q.

Wilson's (1989) typology of four agency types: production, procedural, craft, and coping

agencies. This well-known typology incorporates two dimensions, one related to whether

outputs are observable and another to whether outcomes are observable. Feller (2002) also cites Wilson in pointing to the relevance of organizational production characteristics on performance measurement systems.

Production agencies have observable outputs and outcomes (e.g., postal service) while craft agencies have outputs that are difficult to discern but observable outcomes (e.g., Army Corp of Engineers). Outputs and outcomes are both difficult to observe in coping agencies (e.g., police departments). And in procedural agencies, outputs are evident, but agencies are challenged to attribute outcomes to their efforts. CDC is an example of a procedural agency, in that outputs can be observed but outcomes are typically difficult to observe and/or not easily attributable to the agency. For instance, one can easily count the number of cooperative agreements, contracts, or grants (i.e., outputs) that CDC awards to partner agencies. However, observing outcomes based on those outputs is more difficult. First, as noted above, it is a more complicated proposition to suggest that eventual changes in longer-term outcomes, such as morbidity or mortality, are attributable solely to CDC, especially given the nature of public health problems and the networked contexts of interest here. In reality, health outcomes are influenced by multiple factors; therefore, they are more likely the result of many different interventions and environmental factors. Next, several of CDC's prevention programs, which count incidents of disease, accidents, or suicides that are avoided, are typically problematic if not impossible to measure. Finally, many public health outcomes, such as changes in morbidity and mortality, take years to achieve, another factor moderating the ability to observe agency outcomes.

In table 1 below, Jennings and Haist revise Wilson's typology to offer the expected impact of performance measurement for each agency type (Jennings & Haist 2004, p. 185). Jennings and Haist define "observability" in two ways – the ability to *measure* outputs and outcomes and the ability to *attribute* outcomes to a particular agency.

**Table 1. Expected Impact of Performance Measures by Type of Agency[6]**

| | Outputs Observed | Outputs Unobserved |
|---|---|---|
| Outcomes Observed | Production Agency: Substantive impact is clear; performance measures likely to be used and to have impact. | Craft Agency: Because outputs are difficult to observe (and therefore to control), expectation is for symbolic impact or the adoption of outcome-focused performance incentive systems as a means of improving agency effectiveness. |
| Outcomes Unobserved | Procedural Agency: Substantive impact on procedures or outputs, but may be difficult to assess whether outcomes are impacted by performance measurement-driven changes in outputs being measured. Attention will focus on efficiency, and the relationship of inputs to outputs. | Coping Agency: Because neither outputs nor outcomes can be observed, indirect measures may be used; however, impacts will be unclear or measures will be at the activity level. |

---

Jennings and Haist's first two hypotheses organized around agency type are derived directly from the table above and are based on the premise that the use and impact of performance measurement is affected by the extent to which (1) outputs and outcomes can be observed and attributable to the agency and (2) the extent to which outcomes are controlled by the agency. However, the authors also recognize that political influences may override the actual ability of agencies to control outcomes. For instance, although strong political rhetoric demands schools' accountability for results, student outcomes are extremely difficult to attribute to a particular educational institution or program (Jennings and Haist 2004).

The first two hypotheses are as follows:

H1: "The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed" (Jennings and Haist 2004, 185).

H2: "Measurement will be more common and will have greater impact when agencies have greater control over outcomes" (Jennings and Haist 2004, 185).

The third hypothesis relates to the extent to which measures are consistent with the agency's task and goals. There are different types of performance measures, some that focus on processes and others on program outputs and outcomes. Jennings and Haist (2004) suggest that the choice of what types of measures to include in any given performance measurement system will depend, in part, on which types best reflect whether the agency is achieving its mission. Their third hypothesis is as follows:

H3: "The impact of performance measurement depends on the design of a set of measures appropriate to agency tasks and goals" (Jennings and Haist 2004, 185).

These three hypotheses provide a constructive, theoretical focus and offer valuable direction for this research. Together, the hypotheses emphasize the potential influence of organizational context and function on the design and impact of performance measurement. This seems a useful point of departure to explore the implications of networked public management on the design and implementation of federal performance measurement systems in public health. In particular, one observed limitation of Jennings and Haist's work is their lack of explicit attention to either the decentralized nature of social programs in the public sector or to the increasing emergence of networked governance. This shortcoming has been noted by others (Frederickson and Frederickson 2006). None of the twenty-five hypotheses, or in fact any of their chapter's text, explicitly addresses the reality of decentralized implementation structures characteristic of many public programs, including those at CDC. The three hypotheses highlighted above account for agency type or function, but are not particularly attendant to varied implementation structures. This weakness is not limited to the work of Jennings and Haist; neither GPRA nor much of the vast literature on performance measurement addresses the complexities imposed on performance measurement by more complex implementation structures. This limitation in the theory development work of Jennings and Haist, however, invites an opportunity for this research to contribute to their theory building by extending on their theoretical propositions.

On a more pragmatic level, the research contributes to a better understanding of the development and implementation of performance measurement in highly decentralized environments. For instance, the first hypothesis is based on the assertion that performance measurement is more likely to be used and to have a greater impact in

agencies where outputs and outcomes can be observed (and are attributable to the program or agency). Based on this proposition, Jennings and Haist (2004) suggest that production agencies (e.g., sanitation) will more frequently and effectively use performance measurement than, say, coping agencies (e.g., police department) or procedural agencies (e.g., public health) where outputs and outcomes are not easily attributable to the agency. How then, do network contexts, in which outcomes are more often the result of cooperative efforts and accountability is highly fragmented, affect the use of performance measurement, the types of measures employed, and its perceived impact?

The second hypothesis suggests that performance measures are less likely to be applied and to have less effect in organizations where control over outcomes is weak (Jennings and Haist 2004). In networked environments, the role of public management shifts from a focus on command and control to one based on negotiation and bargaining (Agranoff and McGuire 2001a; Milward and Provan 2006). Cooperative strategies become necessary to facilitate coordinated responses that are needed to more effectively address social problems. Consequently, and as noted earlier, networks involve interdependencies among agencies and partners, and these interdependencies further compromise government's authority over agencies and programs (Peters 2001). What then are the effects on performance measurement of these interdependencies and the decreased level of control over program outcomes?

Finally, the third hypothesis relates to the idea that performance measures will reflect the agency's function. Therefore, even in today's political climate where outcomes are emphasized, an agency focused on service delivery may value indicators of process

21

over outcome measures (Jennings and Haist 2004). How, then, are performance measurement systems structured to reflect the agency tasks and goals in a networked context?

## 1.5 Research Questions

Given the purpose of the research and the theoretical framework described above, three research questions were developed to guide the study. The research questions reflect the dominant feature driving the research, helping to maintain attention to the major concerns of the study and attending to the complexity and contextuality of the case (Stake 1995). The three research questions are directly derived from the first two hypotheses above. These research questions, along with the theoretical hypotheses above, guided all aspects of the research, focusing both data collection and analysis.

- How does networked public management affect the observability of CDC program outputs and outcomes?
- How does networked public management influence CDC's use of performance measurement and the types of performance measures used?
- How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement?

## 1.6 Overview of Methodology and Scope of Study

Two key characteristics of qualitative research include its aim at understanding and description as well as its focus on context and process (Denzin and Lincoln 2000; Peshkin 1993). For this study, qualitative research methods are particularly suitable given

the interest to better understand and describe the implications of network public management environments (context) for the development and implementation of performance measurement systems (process). Qualitative methods allow an in-depth exploration of the relevant processes and contextual influences reflected in Jennings and Haist's three hypotheses (Miles and Huberman 1994). These methods are also appropriate given the limited understanding of the role of networked contexts on performance measurement evident in the extant literature. The field currently lacks the depth of understanding needed to support a quantitative study exploring the influence of specific variables that could be used to make broader generalizations through statistical analysis.

Case study is most appropriate when the nature of the research addresses questions of "how" and "what" and is consistent with the research questions posed above (Creswell 2007). Recognizing the importance of context to the practice of performance measurement, Joseph Wholey (1999), a long-time leader in the field of policy evaluation, has encouraged the conduct of case study research, particularly cross-case analysis, to develop related theory and improve practice. Within the traditions of qualitative research, case study has been increasingly utilized in public administration and public health research (Brower, Abolafia, and Carr 2000; Ulin, Robinson, and Tolley 2005). Case study research is especially valuable to capture individual differences from one program setting to another or from one program experience to another (Patton 2002; Stake 2006).

Although several approaches to case study have been described (Merriam 1998; Stake 1995, 2006; Yin 2009), Stake's methodology was used given its wide use, his attention to the inclusion of multiple cases, and his specification of unique case study

approaches. Stake (1995, 2006) differentiates between intrinsic and instrumental case studies. While intrinsic case study aims to understand a particular case, instrumental case study is used to gain broader insight; that is, the case becomes "instrumental" in understanding something broader than the individual case. Multiple case study attempts to broaden instrumental understanding by including more than one case (Stake 2000). The aim of this research was to better understand the development and implementation of performance measurement in networked contexts rather than to understand aspects of a particular "case" per se. Therefore, a multiple, instrumental case study design was used. A detailed description of the methodology is included in chapter 3.

## 1.7 Summary

A growing number of programs including those in STD, preparedness, tobacco, and cancer, have shifted to emphasizing collaboration within networks as a means to achieve public health goals. This model of networked governance reflects the realities imposed by complex public health problems that require coordinated responses across agencies and sectors in order to effectively address them. Not only does CDC's leadership endorse this approach, it recognizes the importance of accountability and transparency in government. Toward that end, performance measurement has been increasingly adopted across CDC and the wider government as a performance management tool supporting accountability.

This research explored the implications of networked public management on the design and implementation of federal-level performance measurement systems. As noted earlier, performance measurement and networks have been identified as two of the six leading trends transforming government today. Given that performance measurement has

been principally described based on a hierarchical view of public management and that accountability has been cited as the greatest challenge within networked governance, a need emerged to explore the potential difficulties this intersection of context and practice presents. The literature has only begun to explore this issue. Initial theory building proposed by Jennings and Haist (2002, 2004) provides a valuable framework to help guide this investigation. Of Jennings and Haist's twenty-five proposed hypotheses, three based on agency type shaped the development of this study's research questions.

Given the fledgling state of research in the area of performance measurement as applied in networked public management environments, a multiple, instrumental case study was used to gain valuable understanding. Four "cases" of federal-funded public health programs administered by CDC were included and traditional qualitative methods were applied for data collection and analysis. By better understanding the influence of networked governance on the design and implementation of federal performance measurement systems for four public health programs administered by CDC, recommendations for future practice were developed. In addition, the research contributed to expanding the theoretical development that has been initiated by Jennings and Haist.

**CHAPTER 2**

**LITERATURE REVIEW**

This study is informed by relevant literature in performance measurement, public health, and networked management. Broader concepts, most prominently, governance and New Public Management (NPM) offer important contextual understanding helping to frame the proposed study. In the first part of this chapter, literatures related to governance, NPM, and the performance movement are all introduced. In the second part of the chapter, empirical and other literature related to performance measurement, public health, and networked management are presented.

**2.1 Governance**

The term governance has been part of common parlance for decades, but more recently, the term has been used to describe a meaningful shift in how we are governing (Frederickson and Smith 2003). Indeed, governance has emerged as an organizing framework intended to capture an important turn in the practice of public administration and management, that is, the processes of governing (Heinrich, Hill, and Lynn 2004). In particular, the more recent focus on governance signals a change in the relationship between the state and society and reflects the broader relationships that exist between government and its political, administrative, and social environment (Kettl 2002). This particular view of governance recognizes that government operates within an environment of multiple societal forces rather than viewing it as a 'stand alone' institution (Stoker 1998). This perspective also suggests that society is departing from the more traditional view of government with its "assumptions of a collective public interest,

hierarchical authority that compels compliance with mandates, and separation between politics and public administration" (Stone and Ostrower 2006, 5).

Although definitional consensus for the term governance is lacking, Heinrich and Lynn (2000) in their 2000 text, *Governance and Performance*, defined governance as "regimes of laws, administrative rules, judicial rulings, and practices that constrain, prescribe, and enable government activities, where such activity is broadly defined as the production and delivery of publicly supported goods and services" (p. 3). Given that any particular governance regime involves compromise and bargaining across the competing interests, priorities, and values of varied stakeholders, the Lynn, Heinrich, and Hill (2000) suggest that governance is an inherently political concept, a view purported earlier by Peters and Pierre (1998) and Stoker (1998). More simply, Kettl (2002) suggested, "*Government* refers to the structure and function of public institutions. *Governance* is the way government gets its job done" (p. xi).

As noted in chapter one, governance provides the overall context for this study for two reasons. First, a dominant feature of governance involves third parties and networks (Lynn, Heinrich, and Hill 2000). Within the governance framework, the government's work is carried out, in part, through interdependent networks rather than traditional hierarchy alone. In governance, it is collaborative networks, rather than the formal institutions of government, that dominate public policy and are increasingly responsible for delivering public services (Frederickson and Smith 2003; Peters and Pierre 1998). The second reason for using governance as context for this study is that it reflects the larger backdrop within which NPM, with its emphasis on results and its promotion of performance measurement, is practiced. In fact, performance or accountability for results

is often specified as the dependent variable to be explained by governance dimensions (Heinrich, Hill, and Lynn 2004; Stone and Ostrower 2007). Governance, then, provides the overall political and organizational context for this study and, as such, becomes relevant in our understanding about how networked management affects the practice of performance measurement.

Although the field recognizes that the concept of governance currently lacks a causal theory (Frederickson and Smith 2003), some academics have attempted to describe it in more explicit terms. For instance, Pierre and Peters (2005) propose five distinct modes of governance, each based on a unique model of state and society interaction, that are currently operational in existing democratic systems. Stone and Ostrower (2007) identify three features of governance. Stoker (1998) offers a list of five propositions that help describe the concept of governance. Salamon (2002) emphasizes the tools of policy implementation as he identifies five key concepts of governance. Salamon argues that the unit of analysis in the study of governance has shifted from the public agency or program to the *tools* of governance. Heinrich, Hill, and Lynn (2004) assume a multidisciplinary approach in conceptualizing governance as an expanded view of public administration. These various characterizations of governance are distinct, yet they have much in common including their emphasis of the following: the actors involved; the organizational context comprised predominantly of networks; the tools for policy implementation; the processes of governance; accountability challenges; and the recognition of the multitude of factors potentially influencing results. These characteristics overlap to some extent, but are helpful nonetheless in enhancing our understanding of governance. Each is described in greater detail below.

### 2.1.1 Actors

As mentioned previously, governance involves two principal actors, the state and society. This simple dichotomy, however, belies greater complexity, particularly for the state. While the state is commonly divided by policy areas or "stovepipes" and levels (i.e., multilevel governance comprised of federal, state, county, and city factions as in the U.S.), society is differentiated by various private and non-profit sectors each with varied goals (Pierre and Peters 2005). The important point here is that governance involves actors and institutions that are drawn well beyond government alone (Stoker 1998).

Pierre and Peters (2005) proposed five distinct modes of governance, each based on a unique model of state and society interaction, that are currently operational in democratic systems. Although the authors describe each model based on a set of characteristics, the most prominent criteria involves the balance of power between the social and public sector actors. The models cover a continuum from intensive government involvement and control to nearly no involvement. On one end of the continuum sits the "Etatiste" model in which government is assumed to be the principal actor and is able to control the involvement of other social actors, usually keeping their participation at a minimum. At the other end of the continuum resides the model titled, "Governance without Government;" scholars argue that this model is characterized by "self-steering government arrangements" controlled primarily by private actors (Pierre and Peters 2005, 12). Stoker (1998) also suggested that governance involves "autonomous self-governing networks of actors" (p. 18). This metaphor of governance without government is similar to that of the "hollow state," a characterization which implies a shift in focus from traditional public administrative practice within a

bureaucratic state to one that recognizes the loss of state power to other social actors in the policy process (Milward and Provan 2000).

In the governance without government model, the state is viewed as having lost its legitimacy, is less relevant to the policy process, and has, therefore, lost its capacity to govern (Pierre and Peters 2005). In between the etatiste and governance without government models, Pierre and Peters (2005) described three other models (i.e., liberal-democratic, state-centric, and the Dutch governance school models) that reflect shades of this balance between the public and private realms participating in contemporary governance activities.

Given the varied actors involved in governance, Stoker (1998) suggests that there has been a blurring of responsibilities and boundaries in addressing social and economic issues. This is similar to Kettl's (2002) articulation of "fuzzy boundaries" inherent in governance based on the involvement of different actors and their changing roles. The fuzzy boundaries represent the challenges posed by the loss of traditional hierarchy's clear lines of responsibility and the state's loss of sole authority. Kettl identifies a number of "fuzzy boundaries" related to the multitude of actors that are relevant for program management and suggests, "The fuzzy boundary problem confounds the central task of administration building coordinated efforts to solve complex problems" (p. 59).

**2.1.2 Organizational Context Comprised of Networks**

As mentioned earlier, a defining feature of governance involves organizational networks and their role in dominating public policy (Lynn, Heinrich, and Hill 2000; Pierre and Peters 1998; Salamon 2002; Stoker 1998). Frederickson and Smith (2003) state, "Governance is the modern theory of network management" (p. 125). Peters and

Pierre (1998) suggest that the prominence of networks in governance has occurred, in part, because of the de-legitimization of the state due to its perceived bureaucratic, slow, and seemingly unresponsive nature. Salamon (2002) argues that today's emphasis on indirect policy tools has also contributed to the increasing role of networks in governance. And, as noted earlier, others suggest that networks have evolved to more effectively tackle complex public management problems that demand a collective and coordinated response (Agranoff and McGuire 2001b; Milward and Provan 2006; O'Toole 1997).

In a governance framework, networks involve interdependent relationships between the state and other third-party actors from the private and non-profit sectors. Although the state acts as an important partner in the network, it is conceptualized as just one of the interdependent actors. As noted in the section above, Pierre and Peters (2005) describe varied models of governance differentiated, in part, by the strength (or lack thereof) of the state's power in this balance. Of relevance here, is that in a governance framework, power dependencies exists within the relationships involved in the collective action of the network (Stoker 1998).

Given the network's composition and power differentials, the straightforward command and control administrative mechanisms of the past are compromised by the fact that varied organizations now control some of the traditional tasks of government (Milward and Provan 2006). In a governance relationship, various organizations may dominate certain processes, but no one agency can easily command given that problems are no longer controlled within the boundaries of one agency (Milward and Provan 2006; Stoker 1998). Subsequently, a mutual resource dependency is created among network

partners, including those from the public sector, whereby a single organization (i.e., the state) cannot command but is more often dependent on the compliance of the other participating actors (Peters and Pierre 1998; Stoker 1998).

However, hierarchical relations with command and control structures are not entirely replaced by a configuration of networks; our governance system is rooted in a constitutional scheme that requires some amount of hierarchical relations that address political and judicial requirements for accountability (Heinrich, Hill, and Lynn 2004). Heinrich, Hill and Lynn (2004) argue that less likely are the extremes of pure hierarchy or networks (i.e., governance without government), instead, the more practical problems of governance rest "where networks are hierarchical 'tools of government'" (p. 10). Obviously, this leads to consideration of how government can most effectively manage networks (Goldsmith and Eggers 2004).

To some extent, the networks of governance are related to the issue networks described by Hugh Heclo and others in the policy literature (Heclo 1978; Parsons 1995). However, in governance networks, the network goes beyond influencing policy formation and adoption to actually "taking over the business of government" (Stoker 1998, p. 23). Whereas Heclo described the role of issue networks in policy formation, governance networks play a more significant role in long chains of policy implementation. Stoker (1998) suggests that these networks forge their capacities by blending their resources and expertise to form a long-term coalition or "regime" which represents the ultimate act of power in a system of governance.

**2.1.3 Multiple Factors Affecting Outcomes**

As noted, Heinrich, Hill, and Lynn (2004) propose an analytic framework to guide the empirical study of governance. Explicit in their "logic of governance," is the recognition that performance (i.e., individual or organizational level outcomes) are dependent on several classes of variables including environmental factors, client or consumer characteristics, treatments or interventions, structures (administrative or organizational), and aspects of management. This characteristic of governance and its implications for performance measurement is revisited later in this section.

**2.1.4 Tools of Policy Implementation**

Salamon (2002) is the strongest proponent of the expanded use of policy tools as a defining characteristic of governance. Salamon defined a policy tool or instrument of public action as "an identifiable method through which collective action is structured to address a public problem" (p.19). Stoker (1998) also recognizes that, within this new framework of governance, government can apply new tools and techniques to "steer and guide" rather than relying on the "power of government to command or use its authority" (p. 18). And Pierre and Peters (2005) acknowledge the importance of policy implementation in governance and the expanded use of less-intrusive policy instruments that often depend on private sector actors.

Salamon (2002) suggests that within governance the unit of analysis has shifted from the public agency or program to the tools of governance. Salamon (2002) argued that individual tools used in policy implementation both dictate the actors involved and represent a critical factor in structuring the implementation process. Consequently, by

focusing on the policy tool, one can predict the actors who will be involved, the roles they will play in the implementation process, the structure of that process, and its outcomes (Salamon 2002). Furthermore, as noted earlier, Salamon (2002) emphasized that the indirect nature of many of the policy tools used today have contributed to the increasing role of networks which, he contends, are largely shaped by the tool used as those tools play a large role in defining the actors and their roles.

## 2.1.5 Processes of Governance

Pierre and Peters (2005) identified a number of processes involved in governance including identifying and selecting the collective goals of society, makings decisions about how to realize those goals, mobilizing resources across the public and private sectors to reach the goals, choosing the appropriate implementation tools, and providing feedback based on evaluation to support future decision making. In all these processes of governance, the private sector may play a significant role. This is in contrast to hierarchical government, which, based on its authoritative orientation, acts primarily through mechanisms of control and regulation.

In governance, the relational interdependencies among actors involved in networks implies that the state loses the capacity for direct control and must rely, instead, on influence and leverage via a continual processes of bargaining, negotiation, persuasion, and mediating (Kettl 2002; Peters and Pierre 1998; Pierre and Peters 2005; Salamon 2002; Stoker 1998). In the context of governance, cooperation and collaboration between horizontally arranged stakeholders in the public and private sectors replace competition (Milward and Provan 2006; Salamon 2002). The recognition of the role for negotiation, compromise, and bargaining among actors coming to the table with different

agendas, goals, and priorities again emphasizes the political nature of governance (Frederickson and Smith 2003) and points to the need for a different set of enablement skills rather than management skills among public sector employees (Salamon 2002).

### 2.1.6 Accountability Challenge

The shift in administrative practice described above highlights what Stoker (1998) and others (Kettl 2002) identify as new dilemmas related to democratic accountability inherent in the governance framework. Peters and Pierre (1998) echo this concern suggesting that the issue of accountability, in particular, is unresolved in contrast to the traditional channels of answerability inherent in the bureaucratic state. Several of the other aspects of governance noted above contribute to the accountability challenges in governance. The increased devolution and resulting administrative fragmentation based on the growing involvement of the private sector in policy implementation, the challenges posed by fuzzy boundaries, the reliance on bargaining and lack of formal chains of command, the nature of the indirect policy tools used all these contribute to making accountability more difficult to assign within a governance structure.

### 2.1.7 Summary of Governance

The features outlined above help to understand how government is getting its work done in a framework defined as governance. In governance, producing and delivering public goods and services is increasingly dependent on non-governmental partners whom are often arranged and organized through networks and participate with government in processes of goal setting, decision making, resource mobilization, and feedback. The networks involved in governance do not entirely replace the hierarchical

organization of more traditional government structure, but augment those by imposing a new horizontal layer of stakeholders involved in policy formation and implementation (Heinrich, Hill, and Lynn 2004). The value of networks may be their ability to more effectively address outcomes that often reflect the influence of multiple factors. Networks are addressed in more detail later in this chapter.

Governance also involves the use of an expanded array of policy tools, often indirect in their nature. Without the strict lines of authority imposed by hierarchy, the public sector must focus on negotiation and bargaining in the governance context, in what is essentially a political process. Finally, given the extensive involvement of third parties, the network structure, use of indirect policy tools, and reliance on negotiation and bargaining, accountability becomes particularly problematic in the governance framework, an issue that will be addressed in more detail in the sections below.

## 2.2 New Public Management (NPM)

While governance may best be characterized as a political concept based in democratic theory, it relates closely to NPM, an influential reform movement in public management that emerged in the late 1980s and early 1990s. Peters and Pierre (1988) argue that fundamental differences exist between governance and NPM which demand that they be viewed as separate concepts. Although some similarities between the two at the operative level are acknowledged, the authors have identified a number of fundamental differences on a more theoretical level. These differences include the following: (1) governance is about process while NPM emphasizes outcomes and results; (2) governance is inter-organizational in concept while NPM has an intra-organizational focus; (3) governance is based in democratic theory whereas NPM is seen as ideological

in nature calling for the adoption of market practices for the delivery of public sector goods; (4) forms of governance maintain some public sector resources under government control in an effort to support government's ability to act, whereas NPM advocates a larger transformation of the public sector; and (5) forms of governance are seen as easier to adopt in contrast to NPM which requires more substantive cultural shifts (Frederickson and Smith 2003; Hood and Peters 2004; Peters and Pierre 1998). Hood and Peters (2004) also acknowledge the ideological nature of NPM, describing a number of paradoxes that have emerged during its tenure – chastising "one size fits all" approaches while adopting such forms in practice, promoting "evidence-based learning" while often adopting "evidence-free" approaches, and advocating for rational techniques of budget control despite the incremental budgeting experience of previously, like-minded reforms.

Peters (2001) suggests that advocates of NPM contend that government cannot do anything well so it should do as little as possible. Advocates of the NPM reform movement suggest that government is drowning in its own, overgrown bureaucracy resulting in widespread inefficiency. From this view, government is seen as out of touch with the public, economically complacent, preoccupied with due process, and indifferent to the needs of its clients or "customers" (Osborne and Gaebler 1993). The thrust of NPM is an advocacy for the adoption of market-based approaches which are believed as operating more effectively and efficiently. Proponents of NPM argue that the management and methods of the private sector are inherently superior to those applied in the traditional public sector (Peters 2001). Therefore, advocates of NPM view the public-private dichotomy as obsolete (Peters and Pierre 1998) and suggest that government should adopt fundamental corporate notions of competition to create internal markets.

Through markets, government will be enabled to do more with less by increasing efficiencies at multiple levels. Although governance models also incorporate an introduction of competition to the public sector, it is not to the same degree and instead, the focus is toward blending private and public resources through a more cooperative and collaborative approach involving inter-organizational partners (Peters and Pierre 1998).

The principles of NPM or "new managerialism" have been widely accepted in the practice of public administration although extensive empirical study supporting the approach is lacking (Frederickson and Smith 2003). In general, NPM can be described as a generic management philosophy as the approach assumes that, regardless of whether working in the private or public sector, similar management challenges prevail and should be addressed in a similar fashion (Peters and Pierre 1998). David Osborne's and Ted Gaebler's 1993 book, *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*, is viewed as a guiding text for NPM. The authors suggest that America faces a deep crisis in how government operates and called for a new kind of government, an "entrepreneurial government," to represent a paradigm shift in how government does its work (Osborne and Gaebler 1993). In their text, NPM is represented through ten principles, each summarized in an independent chapter. Their tenets of NPM include a government which (1) "steers" rather than rows, (2) empowers rather than serves, (3) encourages competition, (4) endorses a mission-based organization rather than one which is rule-driven, (5) funds outcomes rather than outputs, (6) assumes a customer-focus, (7) earns rather than spends, (8) prevents rather than cures, (9) works through decentralization characterized by participation and teamwork rather than

hierarchy, and (10) applies a market-approach to government (Osborne and Gaebler 1993).

These principles translate into a government reform emphasizing decentralized government, contracting through third parties, deregulation, increased competition, citizen or customer choice, increased discretion among managers, and outcome measurement as a means to assess effectiveness (Frederickson and Smith 2003). Similarly, Kettl (2000) identifies six core issues of NPM including productivity, marketization, service orientation, decentralization, policy, and accountability.

## 2.3 The Performance Movement: Policy Reforms

As noted above, performance is central to NPM with results so fundamental to the reform that it is often referred to as "results-oriented government." Performance, as measured through outcomes rather than outputs, is emphasized in NPM as a means to assess management and policy as well as a means of accountability (Peters 2001). As mentioned earlier, where accountability is viewed as problematic in the governance framework, it is perceived as a particular strength of NPM (Peters and Pierre 1998). In fact, a central theme in NPM is the improvement of government performance through accountability mechanisms that more directly connect providers and consumers of public services. From this view, the point of accountability is substantially shifted from the elected official to the administrator (Barberis 1998; Dubnick 2005).

Paul Light, in his respected 1997 text, *Tides of Reform: Making Government Work, 1945-1995,* inventories federal reform statutes passed during the 50 year period 1945-1995 and identifies four major public management reforms, or tides, including, "(1) scientific management with its focus on tight hierarchy, specialization, and clear chains

of command, (2) war on waste, with its emphasis on inspectors, auditors, cross checkers, and reviewers, (3) watchful eye, with its embrace of sunshine and openness, and (4) liberation management, with its cry to let the managers manage, albeit with a bit of market pressure" (p. 1).

Light (1997) suggests that the most recent reform of liberation management is not new, but instead reflects aspects of earlier reforms, particularly, scientific management. Light identifies competition through increased privatization of the public sector and deregulation as the main tools of liberation management which attempts to encourage more entrepreneurial approaches to management through public agencies emphasizing "steering" rather than rowing. The goal of liberation management, according to Light, is higher performance based on outcomes or results, which are seen as the reform's key products. Al Gore's National Performance Review (NPR) passed in 1993 is recognized as the defining moment for liberation management and the 1993 Government Performance and Results Act (GPRA) as its defining statute (Light 1997). Light noted the influence of Osborne and Gaebler on the movement as a whole, but, in particular, on Gore's NPR stating "Osborne in particular was so important to the NPR that the final report was known as REGO (for *re*inventing *go*vernment) in his honor" (p. 36). The NPR, later called the National Partnership for Reinvention Government, was headed by Vice President Gore and had strong political visibility throughout the Clinton Administration. Not surprisingly, a chief tenet of NPR involves holding public managers accountable for results (Affholter 1994).

GPRA was not necessarily a new idea. It is a statute that has roots in several earlier reforms including the Program, Planning, and Budgeting System (PPBS) from the

1960s, Management by Objectives (MBO) in the early 1970s under Presidents Ford and Nixon, and President Carter's Zero-Based Budgeting (ZBB) in the late 1970s (Light 1997; Nathan 2001; Radin 2006). In fact, the General Accounting Office (GAO[7]) suggests that GPRA improves on those earlier reforms by incorporating their best features (GAO 1997c). These reforms, including GPRA, are generally aimed at improving federal management by introducing greater rationality into government decision making, particularly the federal appropriations and budgeting process (GAO 1996).

GPRA was passed as part of a larger statutory framework to improve federal management that included the Chief Financial Officers Act (CFO), other legislation related to information technology such as the Paperwork Reduction Act of 1995, and the Clinger-Cohen Act (GAO 1997c). GPRA is seen as the centerpiece of this framework which is intended to instill a results-orientation in the federal government. It is worthwhile to note that several features distinguished GPRA from the earlier reforms (e.g., MBO, PPBS) noted above. Most importantly, GPRA was passed as federal statute and, as such, mandates roles for Congress and the legislative branch that previous reforms did not. In other words, GPRA does not go away with a changing administration like executive reform initiatives such as the Clinton and Gore's NPR.

As the federal fiscal environment tightened, GPRA was intended to address some long-standing management problems that were seen as undermining government's efficiency and effectiveness and to introduce greater accountability for results (GAO 2004). The law specifies the following purposes: (1) to improve the confidence of the

---

[7] The General Accounting Office has been renamed the Government Accountability Office.

American people in the federal government; (2) improve program effectiveness through promoting a new focus on results, customer service, service quality, and customer satisfaction; (3) help Federal managers improve service delivery; (4) improve congressional decision making; and (5) improve the internal management of the federal government (GPRA 1993). David G. Frederickson (2003), in his study of GPRA's implementation in five HHS programs, stated, "While measures of government performance and the logic of strategic management have been around since the 1930s, GPRA is certainly the most ambitious and comprehensive approach to measuring government performance and results that has ever been attempted" (p. 13).

Specifically, GPRA's requirements include the development of five-year strategic plans by all federal agencies, followed by program-specific performance plans with specific performance goals and corresponding performance indicators that allow for comparing actual results with the established performance goals (GPRA 1993). Through annual program performance reports, progress toward meeting goals based on performance information is reported to the Office of Management and Budget (OMB). The Act allowed for a seven year start-up period with the first strategic plans due to OMB in 1997 and the first performance reports supplemented with actual data due in 2000. In addition, a number of pilot demonstration projects were included as part of GPRA to assess its implementation (GPRA 1993). When passed by law in 1993, the bill received overwhelming support from Congress as well as the Clinton Administration and was endorsed by the National Academy of Public Administration (Affholter 1994). Its passage was accompanied by strident political support such as this from Senator William Roth, one of GPRA's chief authors, "For the first time, the American people will be told

what levels of service and program results they can expect from their tax dollars, followed by reports on what was actually achieved. It may sound like common sense for the government to do this – and it is – but is has not been done before by the federal government" (Cited in Light 1997, 235).

The specific focus on results or outcomes, rather than inputs or processes, is another distinguishing feature of GPRA from earlier reforms. This move to a results-orientation requires a shift in organizational culture that may necessitate new ways for administrators to conduct their work (Mihm 2000). GPRA does not aim, like some earlier reforms or executive initiatives, to reorganize government in any structural way. Rather, as noted previously, it takes government structure as a given and emphasizes improving management through goal setting and achieving results (D.G. Frederickson 2003). By attempting to replace the rule-based compliance type accountability of more traditional bureaucracy with performance and results-based accountability, GPRA's focus on results reflects the tenets of NPM (Light 1997).

As for any political reform, GPRA embodies the varied agendas of different stakeholders, some not so readily reflected in its statutory language (Radin 2006). First, some advocates of performance measurement view GPRA as a means toward better government. These stakeholders assume that government activities are measurable and measurement is the most objective means to inform public policy decision making (D.G. Frederickson 2003). This view of GPRA, and the reinventing movement in general, reflects a commitment toward a rational decision making model whereby politics are avoided and "objective" performance data will be managerially used in the policy process to improve decision making, and, in particular, decisions about budgeting (Broom, et al.

2002; Kravchuck and Schack 1996; Peters, 2001; Radin 2006). However, as Broom and her colleagues (2002) pointed out, "Virtually all systematic examinations suggest that a focus on performance does not preclude political considerations from driving decision making" (p. 2). This is especially true for the federal budgeting process, one that is intrinsically pluralistic, political, and incremental in nature (Frederickson, DG, 2003; GAO 1997c; Nathan 2001; Radin 2006; Schick 2001). There is, in fact, no evidence to suggest that GPRA has affected the federal appropriations process (Frederickson 2001; GAO 2005).

Other stakeholders, particularly professional government accountability experts, see GPRA as a means for budget cutting and government reduction through the identification of duplicative and ineffective government programs (Radin 2003). And still others view GPRA from a more apolitical stance, to address issues related to public relations or budget maximizing objectives (D.G. Frederickson 2003). Given the political nature of the policy process, the promise of GPRA, and performance measurement more generally, may lie on its potential contribution to improve decision making by policy makers, administrators, and program managers and increase government accountability to its citizens (Fredrickson 2001, 2003). As Wholey and Newcomer (1997) suggested, "The theory behind the Government Performance and Results Act is that planning and performance measurement will help agencies communicate performance expectations and results – and that the use of performance information will improve management and program effectiveness, improve policy decision making, and improve public confidence in government" (p. 93).

An evaluation of GPRA on its tenth anniversary conducted by GAO (2004) based on a randomized survey of federal managers as well as interviews, focus groups, and document reviews identified a number of positive findings. These include: (1) federal managers surveyed reported having significantly more of the types of performance measures (i.e., outcome measures) required by GPRA than at earlier times; (2) interviewees reported GPRA having positive effects on planning and reporting; and (3) reviews of plans suggested that GAO feedback was incorporated to improve plans, goals became more quantifiable over time and more results-oriented, and agencies provided increasing detail about their goals and strategies to address concerns about performance and accountability (GAO 2004). In general, research suggests that GPRA has contributed to improved program management but has not been used by legislators for influencing decision making related to appropriations (Frederickson 2001; GAO 2004).

Although GPRA has strong support, it also has its share of critics. Beryl Radin has written extensively about GPRA and the performance movement more generally (1998, 2000a, 2000b, 2001, 2003, 2006). She suggests that the GPRA legislation can best be viewed as an example of a government-wide, one-size-fits-all reform that endorses generic activities and requirements (2006). She argues that GPRA "does not fit easily into the institutional structures, functions, and political realities of the American system," rendering GPRA, like previous reforms to, "operate largely as rhetorical positions or arguments without the ability to influence substantive policy and budgetary processes" (Radin 2000a, 111).

Others have identified important limitations in the legislation itself, including GPRA's failure to recognize a number of unique attributes of public programs including:

(1) varied goals and contexts of different federal agencies, (2) the extensive degree of decentralization in program implementation structures, (3) the variety of policy instruments employed by each for implementation, and (4) the unique challenges faced by agencies in measuring the performance of public programs (D.G. Frederickson 2003). Studies of federal agencies' experience with GPRA have confirmed these issues as challenges (Frederickson 2003; GAO 1997a, 1997b, 1999, 2000; Mihm 2000). Given the uniform nature of GPRA and its requirements (i.e., Radin's claim that GPRA is another one-size-fits-all, government-wide reform), these areas of variability challenge its implementation.

Of particular note are those challenges to GPRA posed by key attributes of NPM, specifically, its emphasis on decentralization, the increased use of third parties, and expanded use of policy tools. Several researchers have pointed out that GPRA does not acknowledge government's extensive decentralization and use of third parties, including those that are arranged in networks, nor does it recognize the characteristics or purposes of those arrangements (Fossett, Gais, and Thompson 2001; Frederickson 2001, 2003; Radin 2001b). Given a decentralized context, Radin (2000b) suggests that GPRA's requirement that federal agencies define performance goals rather than the states, reflects the law's "compliance" orientation which contradicts NPM's advocacy for increased autonomy to third parties. In his qualitative dissertation study, briefly introduced in chapter one, David G. Frederickson explores the experience of five HHS agencies implementing GPRA[8]. All five agencies utilize implementation structures Frederickson

---

[8] Agencies involved in the study included Health Resources and Services Administration (HRSA), Centers for Medicare and Medicaid Services (CMS), National Institutes for

(2003) described as "articulated vertical networks of third parties." His most important

findings center on the implementation challenges to GPRA posed by third party

governance and the influence of varied policy instruments used by each federal agency[9].

The term "articulated" reflects the "extent to which separate organizations or institutions

in a network are coupled, fit together, linked, or combined and the nature and quality of

those connections," (Frederickson and Frederickson 2006, 8).

In a book written by Frederickson along with his father, H. George Frederickson,

the authors expand on the younger Frederickson's dissertation research to discuss how

these varied "articulated vertical networks" relate to government performance

(Frederickson and Frederickson 2006). Based on grounded theory methodology, the

authors identified the following explanatory factors or variables influencing the

implementation of GPRA in the five agencies (Frederickson and Frederickson 2006,

152):

1. level of third-party policy implementation

2. accountability to federal program purposes

3. nature and quality of network articulations

4. characteristics of goals

5. level of centralization of policy implementation

---

Health (NIH), Indian Health Services (IHS), and the Food and Drug Administration
(FDA).
[9] Frederickson adopted Salamon's (2002) view of policy tools. The FDA uses regulatory
policy tools; CMS uses grants for Medicare and contracted services for Medicaid; HRSA
uses grants; IHS uses grants to support direct service; NIH uses grants to support
intramural research

6.  precision and characteristics of performance measurement

7.  level and character of client and stakeholder support

8.  level of professional identity

In Frederickson and Frederickson's text (2006), each variable is presented as a heuristic continuum and the five cases placed appropriately. For instance, the continuum for the level of third-party implementation extends from completely hollowed out to that of direct government. While IHS is placed toward the latter end of the spectrum, HRSA and Medicaid reflect the hollowed out end. Frederickson and Frederickson deduce that variability in types of third parties, types of policy tools, implementation models, and arrangements in federal agency-third-party articulations and vertical networks all have significant influence on the development of useful performance measurement systems for GPRA. So although the law itself may seem a "one size fits all" type policy, programs seem to have developed varied approaches given the factors noted above.

Others address methodological issues related to GPRA arguing that the legislation makes an implicit assumption that information about performance, the heart of performance measurement, is accessible, affordable, and reliable (Perrin 1998; Radin 2006). Even more fundamentally, some point to the challenges in measuring social phenomenon arguing that it is often quite difficult, if not impossible, to develop indicators for social programs in the same way that economic indicators relate to profit are used (Perrin, 1998; Barry 2000).

Although GPRA remains in place following the close of Clinton and Gore's NPR, it has been overshadowed to some extent by the former Bush Administration's effort to link the executive branch budget recommendations to performance. While GPRA had bi-

partisan support from both Congress and the White House when it was passed as law in 1993, the Bush administration implemented its own approach to performance measurement as part of its wider President's Management Agenda (PMA) (Gilmour 2006). The effort on the part of the Bush administration to integrate performance information into budget decision making is one of the PMA's five government-wide management priorities, the Budget and Performance Integration initiative (Brass 2004). Called the Program Assessment Rating Tool or PART, the instrument was developed by OMB and includes a standard series of approximately twenty-five questions. PART is viewed as a diagnostic tool intended to make conclusions about the benefits of federal programs based on program performance and evaluation information (GAO 2005). The questions fall under four topic areas including: (1) program purpose and design; (2) strategic planning; (3) program management; and (4) programs results (GAO 2005). Each of these dimensions is weighted in the instrument's numerical scoring and then converted into an overall qualitative rating that can be compared across federal programs. PART was developed to assess different types of federal executive branch programs in order to direct funding and management decisions and brings even greater attention to outcomes and results than GPRA (Brass 2004; Frederickson and Frederickson 2006).

PART was initiated in 2002 for use in the fiscal year 2004 budget cycle and has been incrementally applied in a systematic assessment of executive branch programs that have funding associated with them (Brass 2004). GAO conducted an assessment in 2004 of the first 234 programs assessed by OMB for PART for fiscal year 2004 budget cycle (GAO 2005). Their analysis identified over 600 recommendations made by OMB as part of those reviews. GAO found that OMB judged one-half of the programs reviewed as

49

having inadequate results-based information (GAO 2005). Additionally, GAO found that half the recommendations made by OMB related to improving program assessment, in particular, to identify program outcome measures and collect or obtain improved performance data or program evaluations (GAO 2005). In 2005, GAO conducted another assessment to review whether programs had, in fact, improved their evidence base for the PART review. They found that while the PART process stimulated agencies to build their evaluation capacity, programs more typically designed evaluations to meet their own needs related to program improvement rather than broader, impact level evaluations (GAO 2005). In addition, early analysis of PART suggested that assessment results are not effectively impacting the budget recommendations of the executive branch, despite their claims to the contrary (Brass 2004).

The relationship between GPRA and PART has not been well described by OMB except to note that both are central to maintaining a focus on results (Brass 2004; GAO 2004). GAO has repeatedly recommended that OMB articulate an integrated, complimentary relationship between the two policies (GAO 2004, 2005); others have suggested that a stronger integration of the two would strengthen the executive and congressional management reforms (Brass 2004). In sum, while there is limited evidence to date of GPRA or PART's success as results-based reforms, especially in terms of their impact on federal budgeting or improving performance, they both reflect the influence of NPM and the prominence given organizational performance in government today (Cavalluzzo and Ittner 2004; Schick 2001). Significant time and resources on the part of federal agencies are spent addressing both policies (Dubnick 2005). These two "results-based" initiatives, grounded on the assumption that increased accountability through

50

performance measurement will, in fact, lead to improved efficiency and effectiveness of government, have been central to NPM and have held substantial sway in the management operations of federal agencies.

## 2.4 Performance Measurement

In her book about performance measurement Radin (2006) suggests, "It is not an exaggeration to characterize the concern about performance as ubiquitous" (p. 4). Similarly, Coplin, Merget, and Bourdeaux (2002) state, "Performance, in a word, captures the zeitgeist of the 1990s among professional researchers in public administration" (p. 699). This view is endorsed by others (Behn 2003; Blalock and Barnow 2001; Grizzle 2002; Schick 2001) and reflects the importance of performance measurement in government today. Attention to performance, long standing in the private sector, now seems infused in our culture, cutting across all levels of government and extending to the nonprofit arena as well (Hendricks 2002).

Performance in education is given particular attention by the press, but the media also addresses performance in health, the environment, welfare, transportation, foreign policy, national security and other areas of the public sector (Radin 2006). An assumption of unquestionable benefit is presumed. For instance, writing about the *U.S. News & World Reports* annual college rankings, David Leonhardt stated, "Human beings do a better job of just about anything when their performance is evaluated and they are held accountable for it. You can't manage what you don't measure, as the management adage says, and because higher education is by all accounts critical to the country's economic future, it sure seems to be deserving of rigorous measurement" (Leonhardt 2006, C1).

In this section, relevant literature on performance measurement is summarized. The literature is vast, therefore, this review is limited to publications and journal articles about government-based performance measurement, and, in particular, addresses empirically-based quantitative and qualitative studies (e.g., studies based on surveys as well as case studies by GAO and others). Literature related to the principles and practice of performance measurement is also included. Given that performance measurement is a central feature of governance and NPM's recent reforms (i.e., GPRA and PART), the practice has strong champions. Consequently, Frederickson (2003) and others (Forsythe 2001; Nathan 2001) suggest that a dominant literature for the field has a distinct normative or advocacy tone that tends to promote an idealized view of the contribution performance measurement can make to improve government. In contrast, as noted earlier, others have adopted a more critical stance. This literature, both positive and negative, might be best characterized as "commentary" and some, relevant to this study, is included here as well.

## 2.4.1 Principles and Practice of Performance Measurement

An extensive literature is available addressing the principles of performance measurement and approaches for conducting it. Lead authors in this arena include Hatry (1999), Newcomer (1997), Poister (2003), and Wholey (2002). Performance measurement is rooted in management theory and practice, in particular, in performance management. Performance management involves several independent processes including: (1) the development of agency or program mission, goals, and strategies; (2) the development of performance measurement instruments and systems; (3) the use of analytic techniques to interpret performance measures; and (4) use of performance

information for strengthening accountability, improving program effectiveness, and supporting policy-related decision making (Smith and Goddard 2002; Wholey 2002). A performance management system involves the continuous use of all four practices so that they are integrated into the core operations of an organization (Landrum and Baker 2004). Central to performance management, then, is performance measurement which Poister (2003) defines as "the process of defining, monitoring, and using objective indicators of the performance of organizations and programs on a regular basis" (p. 1). Some definitions overtly emphasize outcomes and efficiency as the object of measurement (Hatry 1999) while others are more inclusive of program inputs, outputs, and outcomes at various stages (e.g., intermediate outcomes, end or long-term outcomes) (Newcomer 1997). The indicators of performance, or performance measures, are most often quantitative, objective measures related to some aspect of agency or program performance (Poister 2003). Although some make a distinction between performance measures and performance indicators, the terms will be used interchangeably here.

The development and implementation of a performance measurement system typically involves several key steps (Hatry 2002, p. 30; Poister 2003, p. 23):

1. Assure management commitment

2. Clarify purpose and system parameters

3. Identify mission, objectives, and clients of the program

4. Identify outcomes and other performance criteria

5. Define, evaluate, and select indicators

6. Identify data sources and data collection procedures, providing for quality assurance

7. Develop an analysis plan

8. Specify the system design (reporting frequencies and channels, analytical and reporting formats, software applications, roles and responsibility for system maintenance)

9. Conduct a pilot and make indicated revisions

10. Implement (full-scale)

11. Use, evaluate, and modify system as appropriate

Because of its relevance to this study, categories of performance information used in performance measurement systems should also be defined. These categories are summarized in table 2 (Hatry 1999, 2001; Newcomer 1997; Poister 2003).

**Table 2. Categories of Performance Information Used in Performance Measurement**

**Inputs**: Inputs reflect the resources actually used to produce or deliver program outputs or outcomes. Inputs are typically described based on amount of funding, number of employees, etc.

**Outputs**: Outputs are the products and services completed or delivered based on the inputs. Outputs are the things that the program has done such as the number of trainings delivered, the number of condoms distributed, or the number of breast cancer mammograms provided.

**Outcomes:** Outcomes reflect the program's consequences and can include intermediate and long-term outcomes. Outcomes are typically influenced by external factors beyond an individual program's control.

- **Intermediate outcomes** are those expected to lead to longer-term outcomes, but are not ends in themselves. For instance, using the mammography example above, increasing the number of women re-screened annually for breast cancer is an intermediate outcome.
- **Long-term or end outcomes** are the intended results of the program. Again, using the example above, a reduction in the rate of late stage breast cancer diagnosis (morbidity) reflects a long-term outcome.

### 2.4.2 Assumptions and Orientation of Performance Measurement

Radin (2006, 19) describes a set of assumptions intrinsic to conducting performance measurement which include:

1. Goals can be defined clearly and set firmly as the basis for the performance measurement process.

2. Goals are specific and the responsibility of definable actors.

3. Outcomes can be specified independently of inputs, processes, and outputs.

4. Outcomes can be quantified and measured.

5. Outcomes are controllable and susceptible to external timing.

6. Data are available, clear, and accurate.

7. Results of the performance measurement can be delivered to an actor with authority to respond to the results.

As mentioned previously, some suggest that much of what has been advanced regarding performance measurement assumes a hierarchical context composed of two-party direct government and where accountability structures are clear (Frederickson 2003; Goddard and Mannion 2004; Perrin 1998). In a hierarchical context, performance measurement is seen as an important public management tool of command and control between principals and agents to assess accountability by ensuring results or outcomes are achieved.

### 2.4.3 Uses and Purposes of Performance Measurement

A good deal of literature advances the benefits of performance measurement including its use: (1) as a management tool for program oversight and to link goals and objectives to performance; (2) as an accountability mechanism to assess whether results

are achieved, contribute to more efficient budgeting, and improve confidence in government; (3) to support data-driven planning and decision making; (4) to increase government transparency; (5) as a monitoring tool to identify potential implementation or program theory problems and make revisions; (6) as a means to identify areas for in-depth evaluation; (7) to compare patterns of performance over time; (8) for public relations, including substantiating funding requests; (9) to inform policy makers and contribute to the policy process; (10) for responding to Congressional inquiries; and (11) for benchmarking to allow comparisons across similar programs (Barry 2000; Broom et al. 2002; Hatry 1999; Kravchuck and Schack 1996; Mark, Henry, and Julnes 2000; Mihm 2001; Poister 2003; Rossi, Freeman, and Lipsy 1999; Wholey and Newcomer 1997). Others recognize the political use of performance measurement (Feller 2002; Jennings and Haist 2004; Pollitt 1986). Feller (2002) emphasizes the influence of the political organizational context on performance measurement and stated, "Performance measurement systems and related performance indicators are political instruments used within organizational settings" (p. 439).

Behn (2003) identifies improved performance as the ultimate purpose of performance measurement, but he also described seven other distinct managerial purposes evaluate, control, budget, motivate, promote, celebrate, and learn. Behn (2003) suggests different measures are needed based on the unique purpose for the system. For example, performance indicators used to support budgeting require efficiency measures and indicators to support motivating require real-time measures (Behn 2003). The literature, then, suggests that the choice of measures may be influenced by the use and

purpose of the performance measurement system (Behn 2003), agency type (Jennings and Haist 2004), and the level of decentralization (Frederickson 2003).

### 2.4.4 Empirical Research

As discussed earlier, given prominence of performance measurement and its accepted practice in the field, a relatively small amount of empirical research has been conducted (Barry 2000; Goddard and Mannion 2004; Jennings and Haist 2002; Kelly 2002; Poister 2003; Turnock 2000; Wholey 2002). Kelly (2002) comments, "I think that performance measurement, as a tool of accountability for outcomes, is the rare situation where we understand the practice more than the theory. We know a lot about how to construct and report performance measures, but we cannot say specifically why we go to all the trouble. According to our best evidence, nothing much changes as a result of adopting performance measurement systems" (p. 375).

Based on a comprehensive literature review of impact studies related to performance measurement, Jennings and Haist (2002) note, "We have limited evidence of the impact that it [performance measurement] has on the effectiveness of government systems, the conditions under which it works or does not work, and the factors that affect the success of performance measurement" (p. 2). It was the recognition of this deficit that led Jennings and Haist to begin the theory development that provides the foundation for this research study.

Empirical research on performance measurement based on survey data represents studies at varying levels of government, but studies of public agencies at the city, county, and state levels predominate. Much of this research focuses on the level of adoption and use of performance measurement as part of the management for results (MFR)

framework. For instance, Moynihan and Ingraham (2003) assessed the adoption of MFR systems, Melkers and Willoughby (1998, 2004, 2005) conducted studies of performance budgeting, Poister and Streib studied performance management systems and their use in municipal government (1999a, 2005), and Berman and Wang (2000) examined performance measurement systems.

Based on studies conducted in 1998 and 2004, Melkers and Willoughby find widespread adoption of performance budgeting at the state-level, much of which has been either legislated or required through administrative dictates (Melkers and Willoughby 2004). Similarly, Moynihan and Ingraham (2003) examined state-level data to assess the implementation and use of MFR efforts. Based on data from 2000 collected as part of the Government Performance Project (GPP), the authors found that all fifty U.S. states were using some type of MFR system.

In a more recent study by Melkers and Willoughby (2005), the authors examine the use of performance measurement, again, primarily for budgeting purposes, based on year 2000 survey data from budgeters and administrators at the state and local levels. In this study, sixty-eight percent of respondents said that either all of their departments or at least half of their departments use performance measures. Poister and Streib found somewhat slow adoption of formal citywide strategic planning and the use of performance measurement. Based on a 1997 survey of managers representing cities with populations of 25,000 and greater, the researchers found that forty percent or less of these municipalities were using performance information in meaningful ways to support management and decision processes (Poister and Streib 1999a). In a study conducted with the same sample in 2004, Poister and Streib found that only twenty-two percent of

cities used performance measures to track progress in meeting goals and objectives detailed in strategic plans (Poister and Streib 2005). Similarly, Berman and Wang (2000) analyzed survey data from public managers working in counties with populations of 50,000 or greater and found thirty-three percent of counties using some type of performance measurement but only twenty percent demonstrating "high use."

These results suggest that although programs have often adopted performance measurements systems of various types, the meaningful use and integration of performance information has lagged. In an interesting look at factors associated with the adoption and implementation of performance measurement, Julnes and Holzer (2001) used data from a national survey of state and local government officials and found that the adoption of performance measurement is influenced by rational and technocratic factors such as goal orientation, resources, internal requirements, and information, whereas implementation was influenced by political and cultural factors.

Studies have found that performance information is perceived as useful for communications and management activities such as strategic planning, but has not been used in fiscal decision making (Melkers and Willoughby 2005). To date, little research supports the use of performance measurement for informing fiscal decision making (Melkers and Willoughby 2004). This is interesting given that budgeting is often seen as the driving force behind performance measurement (Newcomer 1997). The research by Poister and Strieb (1999a) found that cities using performance measures did so based on a desire for improved decision making and accountability rather than for simply meeting reporting requirements.

Studies examining factors affecting the use of performance information provide insights as well. Berman and Wang (2000) found that technical capacity and stakeholder support are significantly associated with increased use of performance measurement. Capacity was defined as jurisdictions' ability to: (1) relate outputs to operations; (2) collect timely data; (3) analyze data; (4) access adequate information and technology systems; (5) garner support from department heads; and (6) garner support from elected officials. They also found that widespread use of performance measurement increased satisfaction with its impact (Berman and Wang 2000). A study of federal-level managers working in the twenty-four largest executive agencies used GAO survey data to also explore factors influencing the use of performance measurement (Cavalluzzo and Ittner 2003). These results found that organizational factors such as managerial commitment to the use of performance measurement, decision making authority, and training in the techniques of performance measurement have a significant influence on its development and use.

Again at the federal level, Heinrich (1999, 2002) has conducted research based on data from the federal Job Training Partnership Act (JTPA). Findings suggest that although the use of administrative data in performance measurement results in relatively imperfect data, it still can provide useful information to inform program improvement. In addition, Heinrich's results suggest the importance of considering the effects of other factors, including organizational structure and complexity, policy choices and constraints, and service delivery practices in assessing program performance. This is consistent with the framework for governance she proposes along with Hill and Lynn discussed earlier (Heinrich, Hill, Lynn 2004).

Other than the studies of JTPA, research at the federal level has largely involved data collected by GAO to study various aspects of GPRA, and more recently PART's, implementation. Most of this research involves smaller, qualitative studies aimed at increasing understanding of GPRA's implementation, including challenges federal agencies have faced. In an evaluation of GPRA's first ten years involving both survey data and qualitative interviews, GAO found that federal managers reported having more outcome measures required by GPRA, identified positive effects of GPRA on requirements for planning and reporting, and had more quantifiable, results-based goals in place (GAO 2005).

Other GAO studies identify the many difficulties federal agencies have faced in moving toward the results-based framework inherent in GPRA. These include challenges at various stages of developing performance measurement systems such as identifying goals, developing measures, collecting data, and analyzing and using data. Programs have struggled to translate long-term goals into annual performance goals and develop related outcome measures (GAO 1997a, 1997b, 2004, Mihm 2000). This has been complicated by difficulties uncovered in coordinating cross-cutting federal programs (Mihm 2000) and the mission fragmentation that challenges program's ability to think about how their own activities contribute or are related to common outcomes (GAO 1999). Programs have faced challenges in data collection (GAO 2004) and, in regard to measurement, they have struggled to obtain valid, reliable, and useful data in a timely fashion (GAO 2004; Mihm 2000). Limited capacity (i.e., analytic and methodological expertise) and training have also been identified as impediments, no doubt contributing to the measurement challenges (GAO 2004; Mihm 2000).

A major issue identified through this research has been an inability for agencies to confidently attribute changes in outcomes to a particular program effort (GAO 1996, 1997a, 2004; Mihm 2000). By focusing on outcomes, it becomes difficult to distinguish results produced by federal programs from results caused by other, external factors. This is especially true for complicated social issues for which outcomes are undoubtedly influenced by a complex interaction of factors. Much of the methodological literature on performance measurement recognizes the important limitation of performance measurement related to attribution (Broom, et al. 2002; Hatry 2001; Poister 2003; Wholey and Hatry 1992). Blalock and Barnow (2001) stated that "the major methodological problem in most [performance] management systems is the tendency to attribute the outcomes collected in these systems to social programs or entire human investment systems – that is, to assume that these programs or systems are responsible for the outcomes collected" (p. 503). Indeed, many suggest that there remains an assumption that performance measurement assumes causality (Blalock and Barnow 2001; Frederickson and Frederickson 2006; Perrin 1998; Radin 2006). In an article describing their efforts to develop a performance management system for the Ryan White Comprehensive AIDS Resources Emergency Act (CARE) which is funded by HRSA, Kates, Marconi, and Mannle (2001) stated that the crux of performance measurement as conceived by GPRA is "measuring that impact, or, the difference between having a program and not having it" (p. 147), when, in reality, that type of scientific measurement is not possible through performance measurement systems (Broom et al. 2002; Kates, Marconi and Mannle 2001).

Similarly, Hatry (2001) noted "…the outcome data likely to be obtainable from ongoing performance measurement systems will seldom, if ever, reveal the extent to which the program has caused the outcome. Other factors – over which the program only has partial control – will inevitably be present" (p. 30). Given the influence of external factors, Hatry (2002) argues that public managers cannot be held fully accountable for results. Instead, he suggests that only limited, outcome-level accountability is present in such situations, and shared accountability may provide the best means to reach a fuller degree of accountability for outcomes (Hatry 2002). These same academics often encourage more rigorous evaluation studies to assess cause and affect relationships between the structure and processes of programs to outcomes (Blalock and Barnow 2001; Harkreader and Henry 2000; Hatry 1997; Perrin 2006).

Given this limitation, performance measurement experts have encouraged the inclusion of short-term and intermediate outcomes for which programs may have more control. Such outcomes may provide a more comprehensive view of the program and may allow for more plausible claims of attribution (Derose et al. 2002; Hatry 2001; Poister 2003). In other words, performance measurement systems need to attend to the "entire results chain" (Perrin 2006, 8).

Similarly, authors emphasize that although performance measures can help assess how well an organization is performing, these measures do not typically reveal why or how an organization achieved that level of performance (Behn 2003; Hatry 1997; Newcomer 2001). Again, other evaluation methods are needed to answer questions related to what exactly is happening and why. These issues related to attribution and the kinds of questions performance measurement can and cannot effectively address

highlight the important relationship between outcomes research and performance measurement. Perrin (2002) stated that "outcomes research, quality improvement, and performance measurement are closely related activities, each of which depends critically on one factor: the ability to establish a reasonably firm relation between the structure and processes that are within the control of the individual, institution, or system whose performance is being judged and the outcome that is considered desirable" (p. III-91).

Other research has addressed the unique challenges to performance measurement posed by decentralized implementation structures (Frederickson 2003; Frederickson and Frederickson 2006; GAO 1997a, 1997c, 1998a, 1998b). An important issue addressed by the reinvention movement is the role of the federal government. As discussed earlier, a central tenet of NPM involves devolution of federal responsibilities to third parties including state and local government, nonprofit organizations, and for-profit agencies. A shift in government has involved the reduction in the direct delivery of federal programs by federal staff and, instead, the use of varied policy tools to increase the involvement of third parties. Salamon (2002), who writes about the proliferation of policy tools, noted, "What is distinctive about many of the newer tools of public action is that they involve the sharing with third-party actors of a far more basic governmental function; the exercise of discretion over the use of public authority and the spending of public funds" (p. 2). By giving third partners greater authority and discretion in decision making, decentralized institutions are seen to offer a number of advantages including an ability to more flexibly respond to local needs and changing circumstances, a desire to implement more innovative strategies, and the ability to act more effectively than centralized institutions (Osborne and Gaebler 1993).

As addressed earlier, however, a central dilemma emerges in benefiting from third parties advantages while avoiding the creation of public programs that are so complex and unwieldy that accountability is impossible (Frederickson 2001). Some suggest that a paradox has emerged based on incompatible strategies related to NPM. On one hand, increased decentralization is meant to provide greater autonomy and discretion to third parties. On the other hand, the performance movement is based on more traditional, compliance-oriented accountability (Goddard and Mannion 2004; Hood and Peters 2004; Radin 2001, 2006). Advocates of NPM, however, counter that performance measurement in fact serves as bridge between these to aims. From their view, performance measurement is a means to avoid the traditional command and control mechanisms by substituting outcome requirements (i.e., results) for the flexibility and discretion provided third parties in implementation (Wholey 1999). Others suggest, however, that NPM has focused on performance at lower levels of analysis rather than addressing the more complex situations which involve multiple organizations, cross-organizational policy arenas, or fields involving multiple actors from varied sectors all contributing to outcomes (Bouckaert and Peters 2002). Forsythe (2001), in his text on performance in government noted, "No government-wide strategy for performance management will magically solve the difficult problems of monitoring the results of federally funded programs administered by state and local governments and by private contractors" (p. 547).

At any rate, the introduction of third party governance, including networks, imposes greater complexity for performance measurement at the federal level than at the state or local levels given the intergovernmental and intra-governmental relations

involved (Frederickson 2003; Frederickson and Frederickson 2006; Mandell and Keast 2006; Radin 2000b). The policy implementation literature has long exposed the convolution for program implementation imposed by decentralized structures (Bardach 1977; Derthick 1972; Pressman and Wildavsky 1984). In fact, implementation studies, many of which were conducted in the 1970's, contend that decentralized implementation is, in fact, the source of problems that leave public policy programs falling short of their expectations (deleon and deleon 2002; Salamon 2002; Scheirer 1981; Lester and Goggin 1998). The NPM reforms have only further complicated the implementation process with Lester and Goggins (1998) characterizing the 1990s as an "implementation era."

The recent reform movements point to the complexity of the contemporary implementation process one involving varied policy tools for multiple actors among multiple agencies each influenced by its own set of values, goals, stakeholders, and context. The effects of inter-agency interaction that occur during policy implementation result in extensive evolution in the implementation process whereby policy goals and objectives are modified by the process itself (Majone and Wildavsky 1979). As a policy is implemented across (and down) levels of government, a mutual adaptation occurs between the organizational setting and the program (Berman 1978). In other words, decentralized implementation can lead to extensive variability in implementation across sites, especially given that legislative statutes tend to fairly vague (Elmore 1978). In addition, increasing numbers of actors are involved, each with their own goals and organizational culture. It is not surprising then that new challenges emerge for performance measurement in this context. As Fossett, Gais, and Thompson (2001) contend, "Proclaiming the virtues of mission-driven federalism is, of course, one thing

and success in implementing it quite another. Intergovernmental arrangements complicate virtually all aspects of performance management – agreement on key goals, the development of indicators, the timely collection of pertinent and valid performance data, the interpretation of these data, the implementation of an incentive system (e.g., rewards for strong performers), and more" (p. 208).

A study by GAO in 1998 assessed the challenges in measuring program results in decentralized contexts. Not unexpectedly, goal conflicts often emerge as agencies at the state or local level have differing sets of priorities (GAO 1998a). In addition, programs frequently have different purposes for performance measurement (e.g., accountability, program improvement) (GAO 1998a). Adding more complexity is the variability in activities implemented at each site which is well documented in the implementation literature but also by studies of performance measurement (GAO 1998a). Frederickson (2003) suggests that GPRA makes an implicit assumption of homogeneity in regard to agency service provision activities. In reality, however, program variation can be extensive for a national program that works through state and local agencies and or community based organizations. Under this scenario, it is not uncommon for each site to modify program elements to meet the needs and fit the structure and culture of their respective environments and target populations. However, research suggests this variability challenges the ability of federal agencies to identify a common set of uniform national indicators (GAO 1998a, 1998b).

In regard to measurement, research indicates that decentralized program implementation compromises data quality and introduces variability in data collection across sites as it becomes difficult, if not impossible, to control data collected by third

parties (Frederickson 2003; GAO 1997a, 1998a). This has lead to documented problems

in aggregating data from numerous sites at a national level (GAO 1998a). And finally,

studies show that the training and capacity issues identified earlier are also exaggerated in

a decentralized context as limits in analytic resources and experience with measurement

exist on a wider scale (GAO 1998a).

Other studies have explored how performance information is being used. This

area of research is important because it is the use of performance measurement that must

justify its costs, and performance measurement is likely to influence behavior only if it is

used (Schick 2001; Wholey 1997). An evaluation of GPRA by GAO in 2004 found that

although managers had more performance data, results were mixed on whether they were

using it to improve performance (GAO 2004). In another study that examined the use of

performance measurement in sixteen federal programs, researchers found that

performance data was being used for four purposes: (1) to trigger corrective action; (2)

identify best practice; (3) motivate behavior; and (4) plan (Hatry et al. 2004). Based on

their study, these researchers also characterized a number of barriers to the use of

performance measurement including the lack of authority to make changes based on

performance information, limited understanding of how to use outcome data, and data

problems more generally (e.g., data may not be timely, data may be old by the time it is

available, data may not be disaggregated to a useful level) (Hatry et al. 2003, Hatry et al.

2004). And finally, the study found that some outcomes included in performance

measurement systems can take years to achieve which, again, compromised the system's

usefulness given that performance measurement systems are based on the premise of

providing frequent, timely information (Hatry et al. 2003, Hatry et al. 2004). The empirical research presented above is summarized in a table format in appendix B.

### 2.4.5 Theory Building

Given a deficit in theory development for performance measurement, Jennings and Haist (2002, 2004) made an effort to propose theory intended to help predict the use and consequences of performance measurement by identifying factors that influence the implementation and impact of systems of performance measurement. As noted in chapter one, using the framework for governance articulated by Heinrich, Hill, and Lynn (2004), Jennings and Haist propose twenty-five hypotheses for empirical study (Jennings and Haist 2004)[10]. The larger governance framework developed by Heinrich, Hill, and Lynn suggests that government performance is a function of structure, treatments, environment, client characteristics, and management ($P = S + T + E + C + M$) (Heinrich, Hill, and Lynn 2004). In this model, performance measurement is reflected in the "M" as one type of management tool that may affect performance outcomes in the public sector. From this perspective, the authors suggest that performance measurement is likely to have relatively marginal effects in shaping the impacts of a policy given that there may be a myriad of challenges related to administrative structures, the treatments (interventions) available, the environmental factors, and client characteristics (Jennings and Haist 2004).

The twenty-five hypotheses proposed by Jennings and Haist are based on three features of the governance system that may mediate the implementation and impact of performance measurement. These mediating factors include: (1) incentives that support managers' participation in performance measurement systems; (2) politics, that is, the

---

[10] All twenty-five hypotheses are listed in appendix A.

degree to which political interests agree on the goals of the policy and the validity of the measures; and (3) organizations, including their structure, features, culture, and leadership (Jennings and Haist 2004). As discussed in the first chapter, three of these hypotheses are proposed as a theoretical framework to guide this study.

## 2.5 Public Health and Performance Measurement

### 2.5.1 Developments in Public Health Practice

In the field of public health, attention to performance measurement has increased over the course of the past decade or more. Although public health may have lagged behind in the performance measurement movement in contrast to the health care sector (Handler, Issel, and Turnock 2001; Roper and Mays 2000), CDC has been influenced by GPRA, PART, and other aspects of NPM. The literature related to performance measurement in public health is limited, but reflects many of the same challenges highlighted in the section above. More importantly, the context of the public health system and the nature of national public health programs contribute to the existing challenges facing the effective use of performance measurement. Given that this study is based in a public health context, some background on the field is provided in this section.

In the late 1980s and early 1990s important advancements in the practice of public health occurred. Two notable works were published by the Institute of Medicine (IOM) reflecting these changes. First, in 1988, the IOM published a report titled *The Future of Public Health* which outlined three core functions of public health including assessing health status and related health needs, policy development, and assuring that needed services are provided (IOM 1988). This landmark IOM report defined public health as "what we as a society do collectively to assure the conditions in which people

can be healthy" (IOM 1988, 1). The report articulated a mission for public health and communicated a common approach for the field.

Following this important work, HHS[11] convened the Public Health Functions Working Group in 1994 to define a set of essential public health services derived from the three core functions. This working group developed the ten essential public health services (table 3 below) meant to provide a foundation for public health efforts in the United States (IOM 2003).

**Table 3. The 10 Essential Public Health Services**

Assessment
   1.  Monitor health status to identify community health problems
   2.  Diagnose and investigate health problems and health hazards in the community

Policy Development
   3.  Inform, educate, and empower people about health issues
   4.  Mobilize community partnerships to identify and solve health problems
   5.  Develop policies and plans that support individual and community health efforts

Assurance
   6.  Enforce laws and regulations that protect health and ensure safety
   7.  Link people to needed personal health services and assure the provision of health care when otherwise unavailable
   8.  Assure a competent public health and personal health care workforce
   9.  Evaluate effectiveness, accessibility, and quality of personal and population-based health services

Serving All Functions
   10. Research for new insights and innovative solutions to health problems

---

11 The Department of Health and Human Services (HHS) is comprised of following federal agencies: Administration for Children and Families, Administration on Aging, Agency for Healthcare Research and Quality, Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention, Centers for Medicare and Medicaid Services, Food and Drug Administration, Health Resources and Services Administration, Indian Health Services, National Institutes of Health, and the Substance Abuse and Mental Health Services Administration.

Other important guidance for the field of public health was provided by Healthy People 2000 and, more recently, Healthy People 2010 both of which reflect a vision for disease prevention and health promotion for the country. Healthy People 2010 details public health objectives for the Nation and provides a basis for public health planning at all levels of government[12]. The report includes twenty-eight focus areas, some of which are disease specific (e.g., cancer, sexually transmitted diseases) and others that are behavioral in nature (e.g., physical activity, tobacco use).

In 2003, IOM published a follow-up to its 1988 report titled *The Future of the Public's Health in the 21st Century* (IOM 2003). This report explicitly defines health as a public good and describes government's fundamental responsibility in promoting and protecting the public health. The report emphasizes several important developments that affect our approach to public health and also have direct relevance to performance measurement as applied to federal-level, public health agencies. First, it is recognized by scientists today that health is influenced by a multitude of interacting factors. Building from the work of Dahlgren and Whitehead (1991), a conceptual model reflecting determinants of population health was specified in the report providing a heuristic to facilitate our understanding about the multiple factors that interact in complex pathways to effect population level health (IOM 2003, 52).

The model helps to unravel the relationships between macro-level and micro-level determinants of health. At the individual level, personal traits such as sex interact with disease-related factors such as the virulence of a particular disease agent. These micro level factors "interact along complex and dynamic pathways" (IOM 2003, 53) with more

---

[12] Healthy People 2010. http://www.healthypeople.gov/About/whatis.htm (accessed 08/21/06).

proximate determinants such as individual risk behavior and social, family, and community networks as well as with mid-level determinants like living and working conditions. These in-turn, interact with the larger, up-stream, determinants at the macro level including social, economic, cultural, and environmental conditions (e.g., economic inequality, cultural values, urbanization) and policies at all levels (e.g., local, state, national, global) (IOM 2003). Consistent with lessons learned from community collaboration action (Fawcett 1999), the implication is that multiple strategies (e.g., education, healthcare, health systems, communication, policy action) are needed to more effectively impact health outcomes. The consequences for program evaluation more generally, and performance measurement specifically, are increasing challenges to disentangling individual program effects on particular health outcomes (Barry 2000; Straw 1996)

The complexity involved in assuring and improving the public's health reflected in this model points to a second major development in the field of public health that is addressed in the IOM report. More precisely, there has been a recognition that problems are too complex and resources too limited to depend on the traditional public health system (i.e., national, state, tribal, and local health departments) alone to assure the public's health. "Health is shaped by both innate factors (i.e., genes, age, and sex) and other influences from the social economic, natural, built, and political environments, ranging from the availability of shelter and food to questions of social connectedness and behavior. These multiple determinants of health, among others, constitute a reality that makes it impossible for one entity or one sector alone to bring about population health improvement," (IOM 2003, 41).

Indeed, in their report, IOM (2003) clearly calls for constructing an intersectoral public health system based on collaborative arrangements that emphasize partnerships and networks. This system depends on the contributions of varied sectors of society including the public health system, but also, the health care system, academia, communities, businesses, employers, the media, and individuals (IOM 2003). By leveraging the resources of such diverse partners, the public health system will be strengthened through increased expertise, resources, and broadened perspectives for public health action (IOM 2003). This shift in the view about the actors needed in a public health system is reflected in the fourth essential public health service listed above, "Mobilize community partnerships to identify and solve health problems." As stated in the report, "There is strong and growing evidence that 'healthy' policy must include consideration of domains that are not traditionally associated with health but whose influences have health consequences (e.g., the education, business, housing, and transportation domains)" (IOM 2003, 34).

This does not discount in any way the importance of the traditional public health structure of federal and state agencies that, from a constitutional and historical perspective, have a central responsibility in assuring the public's health (IOM 2003). This is consistent with Heinrich, Hill, and Lynn's (2004) point that hierarchical elements remain intact within a governance framework. In fact, many of our public health programs continue to be structured based on the decentralized relationship between the federal, state, and local government. For instance, at CDC, more traditional lines of authority remain intact through intra-governmental funding systems in which CDC distributes federal dollars to both state and local health departments. CDC also funds non-

governmental partners as well, including community-based organizations and national

organizations. Table 4 below summarizes customary program service delivery and

accountability structures for many of CDC's national public health programs.

**Table 4. CDC National Public Health Program Service Delivery and Accountability Structures**

| | |
|---|---|
| Federal government | CDC<br>• Distributes public funds via cooperative agreements (i.e., a form of a grant) to state and local government, non-profit community based organizations, academic institutions, etc.<br>• Provides technical assistance and support<br>• Assures appropriate fiscal stewardship<br>• Manages federal-level performance measurement systems for its national programs<br>• Reports annually to OMB on GPRA requirements<br>• Reviewed by OMB PART |
| State, Tribal, Territorial government | State, Tribal, or Territorial Health Agency<br>• Distributes public funds to local government and non-profit community based organizations<br>• Provides technical assistance and support<br>• Assures fiscal stewardship for funds distributed<br>• Delivers public health services (i.e., the 10 essential public health services)<br>• Reports to CDC on program activities and fiscal expenditures<br>• Reports performance measurement data to CDC |
| Local government or NGO | Local or Regional Government, Nongovernmental Agencies, Health Providers<br>• Delivers public health services<br>• Reports data as required |

But while these traditional approaches to public health service delivery remain

steady, there is also increasing emphasis on mechanisms, both formal and informal, to

engage other sectors in addressing public health concerns. For instance, in efforts to

address comprehensive cancer control and prevention[13] federal and state partners are

working to encourage the involvement of many sectors and agencies. These include

multiple departments within state public health agencies (e.g., cancer surveillance,

screening programs, nutrition and physical activity programs, tobacco prevention

programs, adolescent and school health), leading non-profit groups (e.g., ACS, Lance

Armstrong Foundation, Susan G. Koman Foundation), business (e.g., pharmaceutical

companies), and universities[14]. This approach to cancer prevention and control uses

strategic partnerships to leverage public outcomes more effectively than what

government could accomplish alone. This is clearly an important advancement for the

field, however, it again underscores the complexity of applying performance

measurement to federal public health programs when multiple partners are involved in

delivering varied implementation strategies to affect particular health outcomes.

Finally, a third development in public health reflects the increasing emphasis on

evidence-based practice and policies. Given the array of public health interventions

available, the complicated pathways from program implementation to health outcomes,

and limited public health resources, efforts are underway to encourage the use of

strategies with some evidence regarding their effectiveness (IOM 2003). One means to

increase accessibility to evidence-based public health practice has been the creation of the

Task Force on Community Preventive Services. This group parallels the U.S. Preventive

Services Task Force (USPSTF) which conducts scientific evidence reviews of clinical

---

[13] Comprehensive Cancer Control address the continuum of care encompassing
prevention, screening (early detection), treatment, palliation, and survivorship
[14] http://www.cdc.gov/cancer/ncccp  Accessed 8/26/06.

preventive services and publishes the *Guide to Clinical Preventive Services*[15]. The Task Force on Community Preventive Services, an HHS initiative that is organized by CDC, reviews scientific information in order to make evidence-based recommendations for public health programs that promote population health. *The Guide to Community Preventive Services* was published in 2005 and provides a review of evidence-based interventions in a number of areas including tobacco, physical activity, diabetes, and cancer (Task Force on Community Preventive Services 2005). The emphasis on identifying evidence based strategies and encouraging their implementation is promising for performance measurement as the causal links between structures, processes, and outcomes will be more fully understood. At this time, however, given the limited amount of rigorous evaluation research of public health practices, there is a significant lack of scientific evidence (Derose, et al. 2003; Perrin and Koshel 1997; Perrin, Durch, and Skillman 1999).

A fourth development in the field reflects the influence of NPM and relates to the importance of improving systems of accountability to ensure health goals are met as well as the quality and availability of public health services (IOM 2003). Of course GPRA and PART apply to federal public health programs, but there is increasing attention on the development of more specific standards for program and workplace performance that extend more broadly among stakeholders (IOM 2003).

---

[15] http://www.ahrq.gov/clinic/pocketgd.htm  Accessed 08/26/06.

These four developments in the field of public health described above are reflected in the IOM (2003, p. 33-34) report's identified six areas for change and action. These include:

1. Adopt a population health approach that builds on evidence of the multiple determinants of health.

2. Strengthen the governmental public health infrastructure – the backbone of any public health system.

3. Create a new generation of partnerships to build consensus on health priorities and support community and individual health actions.

4. Develop appropriate systems of accountability at all levels to ensure that population health goals are met.

5. Assure that action is based on evidence.

6. Acknowledge communication as the key to forging partnerships, assuring accountability, and utilizing evidence for decision making and action.

## 2.5.2 Empirical Research on Performance Measurement in Public Health

As mentioned earlier, the literature on performance measurement in public health, especially empirical studies, is small. Two reports published in the 1990s and commissioned by the National Research Council (NRC) at the request of HHS address performance measurement in the context of a proposed performance partnerships grants program (Perrin and Koshel 1997; Perrin, Durch, and Skillman 1999). These reports summarize some of the key challenges to performance measurement as applied in public health and make some recommendations for improving its practice. Many of the challenges addressed in these reports reflect those described earlier including: (1) issues related to attribution (i.e., the lack of firm causal links between inputs, processes, and outcomes; program resources often represent a small part of the total resources

contributing to the outcome; public health outcomes are influenced by multiple factors);

(2) issues related to data (i.e., limits of available data sources; quality and comparability

of available data; difficulties in specifying numerators and denominators for specific

indicators; costs of data collection and reporting); and (3) issues related to achieving

outcomes in public health (e.g., the length of time needed to achieve health outcomes

such as lowering the incidence of cancer is often extensive) (Perrin and Koshel 1997;

Perrin, Durch, and Skillman 1999).

Given that this study explores issues for performance measurement imposed by

the increasing involvement of networks in our Nation's public health infrastructure, the

final section of this chapter provides a more detailed summary of the emerging literature

in networked public management.

## 2.6 Networked Public Management

### 2.6.1 Defining Networked Public Management

As discussed earlier[16], a defining feature of governance involves organizational

networks (Lynn, Heinrich, and Hill 2000; Pierre and Peters 1998; Salamon 2002; Stoker

1998). Milward and Provan (2000) use the term "hollow state" as a metaphor to describe

this phenomenon in which the government increasingly funds but does not directly

provide public services. As noted in the section above, given the complexity of public

health problems as well as the influence of NPM, networks are increasingly being used to

more effectively and synergistically implement public policy and achieve public health

goals.

---

[16] Section 2.1.2 addressed networks as a characteristic of a governance framework

O'Toole (1997) defines networks as "structures of interdependence involving multiple organizations or parts thereof, where one unit is not merely the formal subordinate of the others in some larger hierarchical arrangement. Networks exhibit some structural stability but extend beyond formally established linkages and policy-legitimated ties" (p. 45). O'Toole's definition has been widely cited and goes on to state, "The institutional glue congealing networked ties may include authority bonds, exchange relations, and coalitions based on common interest, all within a single multi-unit structure" (O'Toole 1997, 45).

Milward and Provan cite a definition by McGuire that more explicitly recognizes the linkages that are the foundations of networks, "networks are 'structures involving multiple nodes – agencies and organizations – with multiple linkages. A public management network thus includes agencies involved in a public policy making and administrative structure through which public goods and services are planned, designed, produced, and delivered (and any or all of the activities). Such network structures can be formal or informal, and they are typically intersectoral, intergovernmental, and based functionally in a specific policy or policy area" (McGuire cited in Milward and Provan 2006, 9). The networks of interest for this study are also collaborative in character. Collaboration can be defined as "a concept that describes the process of facilitating and operating in multi-organizational arrangements for solving problems that cannot be achieved, or achieved easily, by single organizations" (McGuire 2006, 678).

As mentioned in chapter one, networks impose horizontal relationships aimed at improving service integration, often with nongovernmental partners, on top of vertical, or hierarchical ones reflecting the intergovernmental relationships of traditional federalism

(Heinrich, Hill, and Lynn 2004; Milward and Provan 2004). As noted earlier, Heinrich,

Hill, and Lynn (2004) contend that networks will continue to include the hierarchical

relations with command and control structures given that our governance system is

fundamentally rooted in a constitutional scheme that includes political and judicial

requirements for accountability. In fact, intergovernmental relationships reflect the most

typical form of network connections in carrying out public programs (O'Toole and Meier

2004).

Within a network structure, policy processes become more complex because

collective decision making and problem solving are required among mutually

independent actors, each representing various organizations, sectors, and levels of

government (Agranoff and McGuire 2001a; Salamon 2002). As mentioned previously,

these interdependent relationships are often established and maintained based on the

exchange of resources, expertise, information, and technology (Agranoff and McGuire

2001a). Salamon (2002) identifies key attributes characterizing network settings. These

include the following:

1. Pluriformity – Networks exhibit pluriformity by involving a diverse range of agencies and organizational types which often have limited experience collaborating as well as a limited understanding of each other's styles of operation.

2. Self-referential – Networks are self-referential in that each organization participating in the network has independent interests and each approaches the relationship with a unique set of perspectives and incentives.

3. Asymmetric interdependencies – Networks involve asymmetric interdependencies in which all network actors are interdependent, but in asymmetric ways.

4. Dynamic – The features of networks constantly change in response to various conditions requiring network membership to shift, leadership to change, and goals and strategies to evolve.

### 2.6.2 Increasing Role of Networks in Public Management

Networks are not necessarily a new phenomenon in government; in fact, Kettl (2003) notes that since World War II there has been an increase in government partnering with nongovernmental partners as a means to practically deal with the increased complexity of public programs. However, the prominence of networks has grown in recent years for several reasons. First, where the bureaucratic form of governance dominated the industrial age, networks are seen as more appropriate in an information and global age where the world is more complex and diverse (Agranoff and McGuire 2001b).

Second, NPM's emphasis on increased decentralization and reliance on third parties for policy implementation has contributed to the network model (Milward and Provan 2004). Related, Salamon (2002) suggests that the rapid expansion of indirect policy tools used in decentralized government has also played a role in contributing to network management. Agranoff and Meier (2001b) suggest that particular policy tools encourage the utilization of network models.

Third, the emergence of more wicked social problems which require the involvement of many actors in order to effectively address them has also contributed to the increase of networks in public management (O'Toole 1997; Agranoff and McGuire 2001b). As noted above, this is certainly the case in public health where many problems reflect the collision of economic, social, physical, and environmental factors. Fourth, in a government funding atmosphere characterized by "silo" grant-making and contracting, network management may be the best structure to achieve goals that require more holistic solutions.

And finally, networks may be emerging simply based on funders' requirements for collaboration and other political imperatives that encourage networking (O'Toole 1997; O'Toole and Meier 2004). Again, in public health, there has been an increase in the number of these partnerships and coalitions over the past decade as more people recognize achieving public health goals requires the resources and capabilities of more than any single agency or sector (Lasker and Weiss 2003). Many federal agencies include collaboration such as state-wide coalitions as a funding requirement (Wandersman, Goodman, and Butterfoss 1997).

### 2.6.3 Network Types and Structures

Networks can vary in structure, size, and complexity (O'Toole and Meier 2004) and are referred to as "partnerships, coalitions, alliances, strategic alliances, consortiums, and networks" (Milward and Provan 2004, 8). Based on a study of twelve networks in the Midwest, Agranoff (2003) identified four types of networks including informal networks, developmental networks, outreach networks, and action networks. In a more recent report, Milward and Provan (2006) also describe four types of networks service implementation networks, information diffusion networks, problem solving networks, and community capacity building networks. Each is briefly summarized in table 5 below (Milward and Provan 2006, 11).

**Table 5. Public Management Networks – Types and Characteristics**

| Network Type | Key Characteristics |
|---|---|
| Service Implementation Networks | • Government funds the service under contract but doesn't directly provide it (frequently health and human services).<br>• Services are jointly produced by two or more agencies.<br>• Collaboration is often between programs of larger organizations.<br>• Horizontal management of service providers is a key task.<br>• A fiscal agent acts as the sole buyer of services.<br>• Key management tasks include encouraging cooperation, negotiating contracts, planning network expansion, etc. |
| Information Diffusion Networks | • Horizontal and vertical ties between interdependent government agencies.<br>• Primary focus is sharing information across departmental boundaries.<br>• Commonly used for disaster preparedness and other "high uncertainty" problems. |
| Problem Solving Networks | • Primary purpose is to help organizational managers set the agenda for policy related to a critical national or regional problem.<br>• Focus is on solving existing complex problems rather than building relationships for future problems.<br>• Often emerges from information diffusion networks.<br>• Relationships may be temporary, to address a specific problem, and then become dormant after the problem is resolved.<br>• May be either designed or emergent. |
| Community Capacity Building Networks | • Primary goal is to build social capital in community-based settings.<br>• Network purpose is both current and future oriented.<br>• May be created by participants (bottom-up) or by private and government funders (top-down).<br>• Often involves a wide range of agencies with many emergent sub-networks to address different community needs that may arise. |

## 2.6.4 Challenges of Managing in a Network Public Management Context

Network management is understood as different from hierarchical public management. Some specific challenges to managing in a network context are summarized below.

## 2.6.4.1 Joint production problem

Intrinsic to networks is what has been termed the "joint production problem," the challenge to coordinate program implementation across a potentially broad service implementation network (Milward and Provan 2004). In networks, leaders must effectively collaborate with representatives from the public, private, and non-profit sectors as well as across levels of government (Abramson, Breul, Kamensky 2006). As previously discussed, the complex issues addressed by networks require the actions of varied actors to address the multitude of factors contributing to them. The joint production problem points to a central management challenge in network settings, that is, achieving program success in an environment that often involves voluntary collaboration with a variety of actors over whom public managers have little formal authority (O'Toole and Meier 2004). However, a qualitative study of twelve networks in the Midwest conducted by Agranoff (2003) found that government actors were often the conveners of networks. And although trust was found to generally replace more formal authority in this context, government retained status as a core actor and key partner given their legitimacy to address social problems (Agranoff 2003).

## 2.6.4.2 Fuzzy Boundary Problem

As noted above, in a networked model, government represents only one of many actors and institutions involved in addressing social problems (Agranoff and McGuire 2001a). This has led to what Kettl (2002) has termed the "fuzzy boundaries" problem, that is, new management challenges created by the loss of hierarchy's clear lines of responsibility and the state's loss of sole authority. As Kettl (2003) suggests, "Managing

85

government programs effectively thus increasingly depends on bridging the fuzzy boundaries that separate those who make policy from those in the complex interdependent chain of those who share responsibility for implementing it" (p. 60).

These interdependent relationships that exist between government and other third-party actors from the private and non-profit sectors result in power dependencies (Salamon 2002; Stoker 1998). Responsibility for management in networks is shared, therefore, traditional hierarchical authority is compromised (O'Toole 1997). The central authority of government hierarchy is replaced by voluntary cooperation and strategies of influence and leverage (Frederickson and Smith 2003; Peters 2001). For public managers, more time is spent managing the interdependencies between their organization and the others involved in the network (Agranoff and McGuire 2001a). In part, this stems from the multiplicity of interests and values represented in a network; the participants each come with their own set of values, interests, and goals that must somehow be reconciled.

2.6.4.3 Management Capacity

Agranoff (2003) says of network management, "Network management is considered to be a different type of non-hierarchical management, where information and expertise is substituted for authority structure, through a self-organizing process, held together by mutual obligation that develops over time, by reaching consensus-based decisions, and by blending knowledge bases from different organizational arenas into innovative technologies that can become the 'DNA' of networks" (p. 6). To effectively manage in networks, different capacities are needed than those applied in hierarchical arrangements (Agranoff and McGuire 2001b). But, at this time, horizontal management across organizations is poorly understood and government has limited capacity to

86

effectively manage networks (Milward and Provan 2004, 2006). Kettl (2002) suggests a "management gap" exists, that is, government has not adapted its public management systems to administer effectively in a network environment. Similarly, Milward and Provan (2004) suggest that government lacks effective coordinating structures for managing in network settings.

O'Toole (1997) identified four factors differentiating management in networks – a lack of direct managerial supervision over those for whom their performance is judged, diffuse monitoring channels that are often unreliable, the absence of a shared organizational culture, and a need to integrate potentially diverse organizational needs into action. A study of intergovernmental networks in public school districts differentiated between a "structural network," which the authors defined as the more stable, intergovernmental network, and a "behavioral network," the larger network of partners with whom managers collaborate (O'Toole and Meier 2004). Results suggest that public management, measured by stability in personnel within the structural network and behavioral networking, was positively related to improved performance (O'Toole and Meier 2004). Using the same data set and building on the idea of behavioral networking, Goerdel (2006) found that proactive management on the part of public managers supported organizational success. That is, public managers who initiated contact and collaboration with network actors more frequently achieved organizational success (Goerdel 2006).

Kettl (2002) and others (Mandell and Keast 2006; Agranoff and McGuire 2001a) recognize managing networks as primarily a "people problem" and emphasize skill building in the areas such as communication, negotiation, facilitation, and persuasion

skills that will support the bridge-building and "boundary spanning" necessary to effectively navigate this context as well as ability to effect network arrangements in way that further coordination.

<u>2.6.4.4 Accountability</u>

Accountability represents another major management dilemma in network environments. As noted earlier in the section addressing governance, accountability is seen as particularly problematic given the network structure (Page 2004; Peters and Pierre 1998; Stoker 1998). Both the fuzzy boundaries and joint production problems contribute to making it more difficult to assign accountability in a network structure. In addition, monitoring channels may be more diffuse and unreliable (O'Toole 1997).

Posner (2002) offers three reasons for accountability challenges in networks. First, in network settings, authority and political resources are shared and third parties have significant power. The goals and objectives of those third parties may differ significantly from those of government (Agranoff and McGuire 2003). Stakeholders may represent competing values and be held responsible for different outcomes (e.g., ones related to efficiency vs. effectiveness) (Milward and Provan 2004). Second, as discussed earlier, in networks, information asymmetries exist whereby third parties often have more information and insider-knowledge about what is occurring in the field. Third parties can also hide information about performance, and government often has not supported more costly monitoring to track and assess performance, especially when decentralized programs are involved (Posner 2002). And finally, implementation can involve several layers of actors, including levels of government, non-profit organizations, and for-profit

sectors (Posner 2002). The layers of agencies and actors further fragments and challenges accountability structures.

In summary, given a network structure where program implementation is decentralized, authority is compromised, political resources are shared, and monitoring channels are diffuse and unreliable, demands for accountability are particularly challenged (Goldsmith and Eggers 2005; O'Toole 1997; Peters 2001; Posner 2002). This situation is exacerbated by the current lack of capacity and tools of governance to effectively navigate in a network environment (Kettl 2002; Milward and Provan 2004, 2006). Accountability may need to be conceptualized differently for networks than for single organizations, considering ideas of responsibility and responsiveness (Agranoff and McGuire 2001b).

### 2.6.5 Network Public Management and Theory

Although public administrators have been working in a network environment for many years, related empirical analysis and theory building is more recent and has been relatively slow to develop (Kettl 2003; O'Toole and Meier 2004). Kettl (2003) acknowledges a lack of consensus in the field on whether networks reflect an approach, theory, or loose construct; he suggests that to date, the analysis of networks best represents a framework that may offer some early steps toward the development of tools for better managing within a network model.

The literature clearly reflects, however, increased attention to issues related to managing public agencies which operate in environments characterized by networks. There seems little question that networks have emerged as an important form of governance (Agranoff 2005; Agranoff and McGuire 2003; Frederickson and Smith 2003;

Kettl 2002; O'Toole and Meier 2004). In fact, their importance has led some to call for the development of a paradigm for network management similar to the hierarchical, bureaucratic model that proceeded it (Frederickson and Smith 2003; Agranoff and McGuire 2001b).

In a forum held in 2005 and sponsored by The IBM Center for The Business of Government, public management experts identified using networks to organize in response to "non-routine" or wicked problems as one of the three most important challenges facing the U.S. government (Kettl 2005). This group concluded that, although hierarchical management structures continue to work well in addressing more routine problems (i.e., Social Security, garbage collection), networked approaches are needed to more effectively tackle complex problems (e.g., social service problems, terrorism) where responsibility for solving them is more diffuse (Kettl 2005). The group also concluded that problems such as those related to public health and homeland security must rely on distributed organizations, that is, efforts that integrate hierarchical and networked approaches (Kettl 2005).

The report included a case study example of how CDC Director, Dr. Julie Gerberding, has recently reorganized CDC in order to more effectively respond to public health emergencies (e.g., bioterror episodes, bird flu outbreaks). Central to this reorganization was facilitating CDC's ability to more effectively collaborate through networks to improve information sharing within and outside the institution and to leverage needed expertise of CDC's partners (Kettl 2005). Kettl (2005) argues that the twenty-first century problems facing CDC called for its transformation from a more

traditional hierarchical organization to one that reflects new goals and an emphasis on performance, increased collaboration, and improved service delivery.

### 2.6.6 Performance Measurement and Network Governance

Only a few sources were identified in the literature explicitly investigating the relationship of performance measurement and network governance and most of it comes from Europe and Australia. The study of GPRA implementation by Frederickson and Frederickson (2006) previously described is relevant here as well. Based on a review of performance systems in seven different public sectors within the United Kingdom (U.K.), Goddard and Mannion (2004) explored whether performance measurement systems can also support horizontal relationships more characteristic of a network approach to governance. They differentiated vertical with horizontal performance measurement systems based on aspects of measurement, analysis, and action (use). These characteristics are summarized in table 6 below (Goddard and Mannion 2004, 82).

**Table 6. Characteristics of Vertical and Horizontal Performance Measurement Systems**

| | Vertical, Hierarchical Approach | Horizontal, Network Approach |
|---|---|---|
| Measurement: Scope | • National (center) perspective<br>• Focus on dimensions of performance important to center<br>• Quantitative orientation for indicators<br>• Lack of consultation on system design | • Focus on issues of concern to customers, clients, and provider staff<br>• Participative process of system design<br>• Include qualitative elements of performance<br>• Focus on dimensions of performance important to local partners |
| Measurement: Purpose | • Accountability (financial and political) and external control<br>• Focus on attaining minimum standards/uniform standards<br>• Use performance data to mediate relationships between different parts of the system | • Emphasize use by organizations for their own purposes<br>• Focus on improvement across all organizations<br>• Use of data to support informal systems and channels of communication and dissemination<br>• Use of peer review process |
| Analysis | • Center focuses on providing data definitions and ensuring consistency in data collection, analysis, and presentation<br>• Analysis undertaken by center and fed back down to organizations<br>• Organizations collect data only for the purpose of meeting central priorities<br>• Analysis highlights best and worst performers | • Some use of anonymized information collected by and shared between, participating organizations<br>• Organizations use the data they collect for local purposes<br>• Analysis focused on continuous quality improvement across all organizations<br>• Use of inspection to cover "softer" areas of performance not captured by quantitative indicators |

|  | Vertical, Hierarchical Approach | Horizontal, Network Approach |
|---|---|---|
| Action: Dissemination Incentives | • Focus on achievement of indiscriminate targets by all regardless of starting point<br>• Publication of results in league table format<br>• Focus on meeting needs of external stakeholders (e.g., public)<br>• Incentives targeted at poor performers<br>• Incentives targeted at institutional level<br>• Incentives directed largely at extrinsic motivation | • Encouragement for organizations to access and use the data for their own quality improvement purposes<br>• Focus on helping organizations to make sense of the data and share best practice<br>• Incentives exist for encouraging performance improvement regardless of the starting point of the organization<br>• Incentives directed at the individuals whose behavior affects performance<br>• Incentives directed at intrinsic motivation |

The authors found both vertical and horizontal features of performance measurement systems in the seven different public sectors reviewed in the U.K. That is, they found that vertical and horizontal approaches seem to exist on a continuum with agencies having adopted elements of both in their performance measurement systems. This is consistent with Heinrich, Hill, and Lynn's (2004) contention that in the governance framework a continuum emerges related to the horizontal and vertical dimensions of administrative structures rather than simply one or the other alone. Like Radin (2006), Goddard and Mannion argue against a "one-size-fits-all" approach to a performance measurement system and suggest a more nuanced strategy to integrate top-down approaches with ones more horizontal in nature. Their findings promote more of a contingency perspective, suggesting that vertical approaches may be more effective when policy dictates a small number of important targets by focusing the actions of agents.

Their review also identifies unintended consequences related to gaming and misconduct that can emerge under these circumstances. In contrast, they find a key benefit of horizontal approaches to performance measurement in the creation of a more open environment that supports freer exchange of information leading to continuous learning and program improvement (Goddard and Mannion 2004).

In a recent paper by Mandell and Keast (2006), both from Australia, the authors suggest that newer, "non-traditional," performance measures are needed to evaluate performance in collaborating networks. They argue that traditional performance measures focusing on tasks completed and clients served may be appropriate for single agency efforts but are inadequate for network arrangements. Rather, the authors suggest that these types of performance measures must be augmented by ones assessing the performance of the network itself. Outputs and outcomes of importance to networks include those related to the processes that occur throughout the development of a network as well as relationship building, trust, and more tangible elements such as revised policies and procedures (Mandell and Keast 2006). Mandell and Keast (2006) suggest that the effectiveness of collaborating networks is not based on the effectiveness of any single agency, but rather "by the ability of all organizations in the network to act as a cohesive whole" (p. 4). Consequently, the authors argue that performance measures must be developed with input from all network members rather than imposed by a government agency (Mandell and Keast 2006). In addition, they acknowledge that network structures will compromise the ability to hold individual government agencies accountable for individual outcomes (Mandell and Keast 2006). In fact, within a network, Page (2004)

suggests that the attempt to hold individual agencies responsible for broad outcomes may actually risk setting them up for failure.

When operating through networks, then, the issue emerges of including common measures of collaborative performance. That is, to better manage networks, should an effort be made to develop measures that capture the performance of the network itself as suggested by Mandell and Keast? In an earlier paper by Keast et al. (2004), the authors noted, "The difficulty is that the types of results that occur through network structures do not have to do with generating programs or numbers (although that is part of the secondary results), but have to do more with changing relationships and perceptions, which are much more intangible" (p. 367). The authors advocate for the recognition of outcome measures that reflect the benefits of network operations – "systemic change, relationship building, innovative operating procedures, and community inclusion" (Keast et al. 2004, 370).

Similarly, Voets, De Rynck, and Van Dooren (2006), all from Belgium, consider the challenge of measuring collaborative performance in networks given that it may offer an important tool when managing in these contexts. They ask, "Do we need a different conceptualization of performance, if precisely these less tangible and measurable dimensions are at the heart of collaborative networks?" (Voets, De Rynck, and Van Dooren 2006, 21).

In the U.S., Robert Agranoff (2005) has given this idea some attention as well, considering how to develop measures that capture the "value-added" of the collaborative network. Agranoff (2005) proposed developing measures relating to the individual level (e.g., individual gains based on network participation), the organizational level (e.g.,

access to new information or expertise, increased integration), and collaborative levels (e.g., process related to relationship building, etc.).

## 2.7 Chapter Summary

This chapter summarized several literatures relevant to the proposed study including those related to governance, NPM, policy reforms related to NPM and performance measurement, performance measurement, public health, and networked public management. These literatures all contribute to framing the study in the current discourse. Governance provides an important overall context given the centrality of networks in the framework. New Public Management, with its policy reforms (i.e., GPRA, PART) and emphasis on performance measurement, also provides invaluable context for the study.

As the dissertation title suggests, the literature recognizes potential challenges for developing and implementing performance measurement in networked public management contexts. Although performance measurement has been widely adopted in practice, network contexts pose unique challenges given their characteristics outlined above. Studies by GAO and Frederickson addressing GPRA's implementation imposed by third-party government and decentralized implementation structures provide an important foundation from which to build the proposed study. Theoretical work by Jennings and Haist also provide an essential basis for the research.

The literature suggests that government is struggling to adapt and transform its public management practices to fit new structures that support more agile and coordinated responses to the complex policy problems facing the country today.

Performance management is one strategy that has been advocated as a means to improve

the results of public programs in these contexts (Kettl 2005). However, a better

understanding of how programs are addressing the challenges posed by network

governance in developing and implementing performance measurement is needed in

order to enhance its practice. Therefore, more research is needed. The next chapter

describes the methodology used for this study that addresses this gap in the research.

# CHAPTER 3

# METHODOLOGY

Fundamentally, the selection of research methods is dictated by the researcher's philosophical view, as well as the study's purpose, the nature of the research questions, and the available resources (Creswell 1998; Patton 2002). Qualitative methods are most suitable when a subject is simply insufficiently understood and researchers are exploring the "how" and "what" (Creswell 2007; Ulin, Robinson, and Tolley 2005). Both are relevant here. First, as already noted, the extant literature addressing the convergence of performance measurement and networked governance is scarce. Second, as reflected in the research questions, the proposed study is exploratory in nature, aimed at better understanding the implications of networked public management on the design and implementation of federal-level performance measurement. Consequently, a qualitative methodology is applied because it enables a more detailed description and explanation of federal performance measurement applied in these contexts.

The field of qualitative research has grown substantially over the past two decades with its methods increasingly utilized by researchers in the applied social sciences, including among others, those in public policy, public management, public health, and evaluation (Brower, Abolafia, and Carr 2000; Miles and Huberman 1994; Rossman and Rallis 2003; Ulin, Robinson, and Tolley 2005). Notably, some classic texts in the field of public policy have been based on qualitative research including Pressman and Wildavsky's (1973) account of policy implementation and Allison's (1971) analysis of the Cuban missile crisis. Agranoff and Radin (1991) argue the importance of systematic and rigorous case study approaches to explore complex issues in public administration

that cannot be effectively addressed using other research methods. Furthermore, qualitative research can make an important contribution toward building a bridge between theory and practice that will benefit public managers (Agranoff and Radin 1991; Brower, Abolafia, and Carr 2000).

This chapter summarizes the study's research methods. First, some key characteristics of qualitative research are described to introduce the broad research approach adopted for this study. This is followed by a justification of the qualitative approach chosen for the study, case study methodology. In particular, features of a multiple, instrumental case study, the specific type of case study applied here, are detailed. The chapter then presents a thorough overview of the sampling, data collection processes, data analysis procedures, and the presentation of findings. Next, the chapter addresses issues related to the study's rigor, particularly those of reliability and validity. The chapter closes by attending to ethical considerations, study limitations, and researcher assumptions.

As introduced in chapter one, the three research questions guiding the study are as follows:

1. How does networked public management affect the observability of program outputs and outcomes?

2. How does networked public management influence the use of performance measurement and the types of performance measures used?

3. How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement?

### 3.1 Characteristics and Strengths of Qualitative Research

The characteristics of qualitative research reflect many of its strengths and contributions as a broad research method. Qualitative research can be characterized by its 1) naturalistic approach, 2) aim toward understanding, 3) focus on interpretation, 4) inductive strategy, 5) use of multiple methods, 6) emergent and flexible design, 7) simultaneous conduct of data collection and analysis, and 8) the researcher's role as the primary instrument of data collection. First, qualitative research involves data collection and methods that are applied in real-world settings that unfold "naturally" (Lincoln and Guba 1985; Rossman and Rallis 2003). The naturalistic approach emphasizes the importance of understanding the *context* in which the phenomenon or person operates based on the belief that the setting significantly influences human behavior (Bogdan and Biklen 2007).

Second, in contrast to quantitative research, with its aim at explanation, prediction, and broader generalization, qualitative research is intent on generating deeper understanding of the whole as well as of the specific and particular (Miles and Huberman 1994; Patton 2002; Stake 1995). By simultaneously applying data collection and analytic methods that are open-ended yet rigorous, qualitative research uncovers nuance and complexity related to the object of study (Janesick 2000). This allows researchers to uncover a richness that is often conveyed in "thick" or "rich" description a narrative providing a deep level of understanding intended to facilitate theory building, explicate relationships, or provide important contextual details not possible through quantitative methods (Denzin and Lincoln 2000).

Third, in qualitative study, the researcher strives to understand the multiple perspectives of those closest to the phenomenon of interest and achieve an "emic" or insider perspective of the view of those studied (Stake 2000). In other words, qualitative research aims to make meaning of phenomena based on individuals' perceptions of them (Creswell 2007; Denzin and Lincoln 2000). Stake (1995) suggests that interpretation is the single most distinctive characteristic of qualitative research.

Next, interpretations are largely drawn inductively in qualitative research rather than deductively, as in the quantitative arena. Categories, themes, and patterns are derived directly from the data (Janesick 2000). It would be misleading to suggest, however, that qualitative work is solely inductive; theoretical frameworks and research questions provide essential structure for research (Miles and Huberman 1994). That is, qualitative research often begins deductively with a conceptual framework and a set of analytic categories, and then it proceeds inductively to derive meaning from the research experience and collected data (Miles and Huberman 1994). But fundamentally, qualitative research involves an effort to inductively derive meaning from the multiple perspectives of those included in the study (Ezzy 2002).

Fifth, qualitative research utilizes multiple methods to collect varied empirical data. Most frequently, data collection involves interviews, observations, and document review (Patton 2002). The use of these varied approaches facilitates a more holistic view of the phenomenon under study and allows for the triangulation of data sources, improving the study's reliability and validity.

Sixth, the methods selected in qualitative research are used in a flexible manner, that is, the approach involves emergent research design (Rossman and Rallis 2003). Miles

and Huberman (1994) recognize the need to "bend the methodology to the peculiarities of the setting" (p.5). This requires that investigators understand the theoretical or policy issues under study as they must make analytic judgments throughout data collection (Yin 2009). This suggests a seventh, closely related characteristic – that "analysis begins in the field" as researchers are attentive to emerging themes and patterns, consistently reflecting on data to help shape the study design (Glesne and Peshkin 1992). Consequently, data collection and analysis occur concurrently in qualitative research necessitating the flexibility noted above (Creswell 2007). For instance, an interviewee may identify another important person to interview; an observation may reveal new issues for study; and a specific document examined may suggest another for analysis (Janesick 2000).

Lastly, in qualitative work, the researcher becomes the instrument for data collection (Creswell 2007; Miles and Huberman 1994). There is a purposeful researcher-subject interaction intended to facilitate deeper understanding. Within this relationship, the researcher assumes a neutral or nonjudgmental approach emphasizing empathy, respect, sensitivity, and responsiveness (Patton 2002). This is in contrast to quantitative research where the researcher assumes the role of objective observer (Stake 1995). Given the role of the researcher in qualitative researcher, attention to voice and reflexivity is essential. Patton noted, "The qualitative researcher owns and is reflective about her or his own voice and perspective; … the researcher's focus becomes balance understanding and depicting the world authentically in all its complexity while being self-analytical, politically aware, and reflexive in consciousness," (Patton 2002, 41).

**3.2 Methodological Approach: Multiple, Instrumental Case Study**

Case study, one tradition of qualitative research, is an empirical method most appropriate when the nature of the research addresses "how" and "what" (Creswell 2007). Case study research aims to deepen understanding by closely studying the complexities of an individual case in order to produce rich description (Rossman and Rallis 2003). The case can focus on individuals, programs, communities, or states, but regardless of the unit of analysis used, a qualitative case study "seeks to describe that unit in depth and detail, in context, and holistically" (Patton 2002, 54).

Case study has been identified as a valuable method for the study of public policy and public administration. Stake suggested, "The utility of case research to practitioners and policy makers is in its extension of experience," (Stake 2000, 449). For this study, case study serves as a useful qualitative approach given the research questions and the unit of analysis (i.e., public health programs nationally administered by CDC). The "case" is defined more precisely in section 3.3 below.

Stake (1995, 2006), Yin (2009), and Merriam (2009) are all widely cited for their approaches to case study research. For this study, Robert Stake's methodology has been adopted because of its wide use, his attention to the inclusion of multiple cases, and his specification of unique case study approaches. Stake (1995, 2006) differentiated between intrinsic and instrumental case studies. While intrinsic case study aims to understand a particular case, instrumental case study is used to gain broader insight; that is, the case becomes "instrumental" in understanding something broader than the individual case. Instrumental case study may involve a single case or multiple cases. Multiple case study attempts to broaden instrumental understanding by including more than one case (Stake

2000). Consequently, multiple case study research starts with the phenomenon of interest and studies individual cases for what they can tell us about that phenomenon (Stake 2006).

This study incorporates a multiple, instrumental case study design. Multiple, instrumental case study allows the researcher to study the phenomenon of interest, performance measurement, in different networked management contexts. The researcher examines the individual cases' similarities and differences in order to better understand the object of study (Stake 2006). Although each case is explored in depth, it is always with a view toward better understanding the phenomenon of interest (Stake 2006). Stake (2006) noted, "It is supposed that the complex meanings of the Quintain[17] are understood differently and better because of the particular activity and contexts of each Case" (p. 40). Therefore, by studying the development and implementation of federal-level performance measurement in different contexts, each involving networked public management, a more nuanced understanding of the phenomenon may be unveiled. Both the unique situational analysis of each case and the similarities and differences across cases contribute to a greater understanding of performance measurement designed and implemented in networked environments.

Miles and Huberman (1994) and Yin (2009) recognized that the inclusion of multiple cases can enhance the confidence of findings. By examining multiple cases that may be both similar and contrasting, the validity and stability of findings are strengthened (Miles and Huberman 1994; Yin 2009). Finally, while multiple, instrumental case studies often involve a research "team," Stake (2006) emphasized that the approach is also

---

[17] Stake introduces the term "Quintain" in his most recent text on multiple case study to represent the broader phenomenon under study in multiple, instrumental case study.

appropriate for dissertation research although the student occupies all roles (e.g., director, data gatherer, and analyst). This issue will be revisited later in the chapter.

### 3.3 Defining the Case

In conducting a case study, the first obligation is to define the "case" (Yin 2009). Stake (2006) defined the case as a "bounded system," reflecting the fact that the case is an object rather than a process. A case is purposive, an integrated system; the case represents a context and experience (Stake 2006). In sum, the case reflects the unit of analysis and defines the scope for which findings and conclusions can be made (Patton 2002)[18].

For this study, the "case" is defined as "a federally-funded, public health program nationally administered by CDC and implemented through a decentralized network that has developed or is in the process of developing a national-level performance measurement system." This definition provides a "boundedness, context, and experience" that usefully defines the case (Stake 2006). The case is bounded at the federal level, specifically at CDC. While the national program involves a decentralized, network implementation structure, the primary interest is to describe the federal-level experience in developing and implementing performance measurement systems. This choice is largely driven by pragmatic considerations related to both scale and regulations imposed by OMB[19].

---

[18] Case study has come to mean more than the unit of analysis, however. The term also reflects a methodological approach as well as the product of analysis (Stake 1995; Patton 2002).
[19] The Office of Management and Budget (OMB) requires Health and Human Services (HHS) approval for research involving nine or more persons as part of the Paperwork

The multiple, instrumental case study provides an in-depth study and analysis of each of four selected public health programs administered by CDC and allows for valuable cross-site analysis. Stake (2006) suggested that the inclusion of at least four programs provides more useful variation and is preferable to two or three. At the same time, Stake (2006) emphasized that the first objective is always to understand the individual case and, in particular, its functioning and activities. For this study, the researcher aimed to interpret patterns across each case, examining it in terms of its own situation, prior to conducting any cross-case analysis (Stake 2006).

### 3.4 Sample Selection: The Cases

Well chosen cases are essential to assuring that the phenomenon of interest is understood (Stake 2006). In contrast to quantitative research, which often uses random sampling techniques derived from statistical probability theory in order to assure generalizability, qualitative research involves purposeful or judgment sampling, with the principle aim of selecting information-rich cases leading to the greatest understanding (Patton 2002; Stake 1995; Yin 2003). Purposeful sampling allows researchers to "tailor" the sample to the study (Stake 2006). Cases are selected based on conceptual rather than a representative basis (Miles and Huberman 1994).

Stake (2006) suggested broad criteria for sample selection in a multiple, instrumental case study that included assuring that the cases: (1) are relevant to the phenomenon of interest; (2) provide diversity across contexts; and (3) provide opportunities to learn about both complexity and contexts. However, the most important

---

Reduction Act of 1995. At CDC, this approval process takes approximately 12-15 months. OMB approval is **not** required for research involving federal employees.

factor for case selection is that the case provides a meaningful opportunity to learn (Stake 2006). Of course, case selection is also influenced by more practical issues. For instance, issues of access and willingness to participate must be addressed (Rossman and Rallis 2003). For this study, the researcher's association at CDC facilitated entrée with relevant stakeholders in the proposed program areas. Based on preliminary discussions with CDC staff from each national program, a willingness to participate was assured prior to the dissertation proposal defense; formal introduction, approvals (i.e., Division Directors support to participate), and permissions (e.g., CDC human subjects review) were secured prior to initiating field work.

As noted above, for this study, a case is defined as a federally-funded, public health program nationally administered by CDC and implemented through a decentralized network that has developed or is in the process of developing a national-level performance measurement system. Consideration was given to whether cases could be included that would be differentiated based on a relevant variable. For instance, in his dissertation research addressing GPRA, Frederickson (2003) included five cases from HHS[20], each of which used different policy tools (e.g., regulation, grants, contracts, direct service). This was consistent with his research questions to explore the influence of policy tools on programs' choice of GPRA measures.

For this research, a differentiating variable relevant to the proposed research purpose and questions involves the network structure itself. It is presumed that while all of the potential cases (i.e., federally-funded, public health program nationally

---

[20] Frederickson included the following five federal agencies in his study of GPRA implementation: National Institutes of Health (NIH), Food and Drug Administration (FDA), Indian Health Services (HIS), Centers of Medicare and Medicaid Services (CMS), and Heath Resources and Services Administration (HRSA).

administered by CDC) are vertically integrated in their implementation structure (e.g.,

funding may move from CDC to a state health department to a county health

department), some have more extensive horizontal network characteristics than others.

By including cases that represented both types, there is an opportunity to make some

comparisons in addressing the research questions. Consequently, a stratified, purposeful

sampling strategy is used. Stratified sampling is one of a number of approaches to

purposeful sampling described by Patton (2002) in qualitative research and involves

breaking the sample into categories. For this study, the sample is stratified based on

characteristics of the network structure.

Four cases, each unique in context, were selected based on a set of broad criteria

that were derived from the study's purpose and research questions. The criteria assured

that the individual cases would share important key characteristics and are "categorically

bound" (Stake 2006, p. 6) in their representation of the study's focus. The overall criteria

were as follows:

1. The cases should reflect programs with widely decentralized implementation networks.

2. The cases should be domestic (e.g., within the U.S. and its territories) and include programs that fund thirty grantees or more[21]. The large number of grantees funded by a single program poses unique challenges in identifying a common set of national indicators.

3. The cases should involve cooperative agreements[22] or grants as the policy tool.

---

[21] Many of CDC's national programs are funded across all 50 U.S. states. Tribal organizations and U.S. territories are sometimes funded as well.
[22] Cooperative agreements are typically used by CDC as the policy tool to provide funding to grantees. Cooperative agreements are a type of grant and require that grantees "cooperate" with CDC for various programmatic decision-making. The funding announcement for each program specifies the roles and expectations for both the grantee and CDC.

4. The cases should reflect programs with a program-specific, performance measurement system either in development or fully implemented. The performance measurement system does not necessarily need to include GPRA measures.

5. The cases should provide ample and rich opportunities for studying performance measurement systems in contexts of networked public management environments. This implies that the maturity of the performance measurement system should be considered.

The cases were stratified as follows: two cases represent programs with primarily vertical, integrated networks and two others represent programs with more extensive horizontal, network composition. Unfortunately, CDC does not maintain an inventory of its national programs with performance measurement systems. Therefore, the researcher sought advice from several CDC colleagues about case selection. All those who were consulted work at senior levels and are well versed with programs across the agency. Through this first set of discussions, a group of ten potential CDC programs were identified meeting the criteria above. Next, individual meetings were held with staff from most of those programs to further explore their program structure, performance measurement system, and interest in participation.

Based on the information collected, four cases were proposed for inclusion: (1) the National Diabetes Prevention and Control Program (NDPCP); (2) the Comprehensive STD Prevention Systems (CSPS); (3) the National Breast and Cervical Cancer Early Detection Program (NBCCEDP), and (4) the National Tobacco Control Program (NTCP). Upon initiating field work, the NDPCP was dropped and replaced with the Public Health and Emergency Preparedness (PHEP) program. This change was made based on advice from representatives in NDPCP. The evaluation team in the Division of Diabetes Translation was revisiting their performance measurement development process

109

and suggested that the researcher identify a more fully developed system. The replacement of cases is appropriate in qualitative research, especially when the change will enhance the utility of the study (Stake 1995).

In regard to stratification, the CSPS and the NBCCEDP represent vertically integrated networks while the NTCP and PHEP include vertical along with more extensive horizontal network dimensions. Each program included in the study is briefly introduced below.

### 3.4.1 Division of State and Local Readiness (DSLR), Public Health Emergency Preparedness (PHEP) Program

DSLR is located in CDC's Coordinating Office on Terrorism and Preparedness for Emergency Response (COTPER). Through DSLR, over $700 million is awarded to 62 state, local, and territorial health grantees in the Public Health Emergency Preparedness (PHEP) Program. The program aims to build public preparedness for a range of hazards including natural disasters as well as terrorist, chemical, biologic, radiologic, and nuclear emergencies. The PHEP program was initiated in 1999 and has changed substantially from year to year. The program is generally viewed as 'new' and the discipline of public health preparedness is still evolving. Working in a complex political environment, grantees collaborate with both vertical and horizontal partners. Six national-level performance measures are currently place and an effort is underway to further develop the indicator set.

**3.4.2 Division of Sexually Transmitted Diseases Prevention (DSTDP), Comprehensive STD Prevention Systems**

CDC's DSTDP is located in the National Center for HIV, STD, and TB Prevention (NCHSTP). The primary grantee program of the DSTDP is called the Comprehensive STD Prevention Systems (CSPS). The purpose of the program is to support STD programs in designing, implementing, and evaluating comprehensive STD prevention systems. Some grantees receive additional funding to address STD-related infertility, syphilis elimination, and specific surveillance activities. Just over $104 million dollars is distributed to all fifty U.S. states, the District of Columbia, six U.S. cities, and seven U.S. territories. Program activities vary by grantee, but include the implementation of community and individual behavior change programs, medical and laboratory services, partner notification and counseling services, surveillance activities, and planning to address STD outbreaks.

The program has a strong vertical dimension in its implementation structure with funding often moving from CDC to state health departments to county or other local health jurisdictions. However, horizontal features are also present as programs are encouraged to collaborate closely with other public health programs including those addressing HIV and hepatitis, the correction and detention systems, drug treatment facilities, family planning clinics, and private medical providers. CDC began requiring grantees to report on a set of twelve core indicators in 2004.

### 3.4.3 Division of Cancer Prevention and Control (DCPC), National Breast and Cervical Cancer Early Detection Program

CDC's DCPC is located in the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). DCPC funds the National Breast and Cervical Cancer Early Detection Program (NBCCECP) which was established by Congress in 1990 with the passage of the Breast and Cervical Cancer Early Prevention Act. The law requires that CDC provide funds to state, territorial, and tribal health agencies to carry out the program. The law specifies six activities for implementation: (1) breast and cervical cancer screening; (2) referrals for medical treatment and other support services; (3) public information and education; (4) professional education; (5) quality assurance; and (6) monitoring and evaluation, including surveillance. Although NBCCEDP funds can be used for some diagnostic procedures, funds may not be used for cancer treatment. The NBCCEDP serves low-income, uninsured, or under-insured women[23].

Currently, the program provides approximately $160 million dollars in funding to 68 state, tribal, and territorial grantees. Since the program's inception, over 3.2 million women have been screened and nearly 8 million screening tests provided. State, tribal, and territorial agencies manage the NBCCEDP through varied implementation structures, but these typically involve a fee-for-service model whereby providers of breast and cervical screening services are reimbursed by the state, region, or county for services rendered.

Although the service delivery structure is primarily vertically integrated, programs collaborate at more horizontal levels with organizations such as the American

---

[23] HHS, CDC, Request for Application DP07-703

Cancer Society (e.g., local chapter), the Komen Foundation, the state comprehensive cancer control programs, and local level community based organizations. In addition, given shared risk factors, NBCCEDP programs typically collaborate with other chronic disease programs such as those addressing diabetes, heart disease, physical activity, nutrition, and tobacco prevention. These collaborations often facilitate coordinated efforts for public and provider education and client outreach. Extensive patient-level screening and diagnostic data are collected and reported semi-annually to CDC. A performance measurement system has been in place since 2005 and is used by CDC for program monitoring and as part of a performance-based budgeting formula.

### 3.4.4 Office on Smoking and Health, National Tobacco Control Program

CDC's Office of Smoking and Health (OSH) supports the National Tobacco Control Program (NTCP) providing a total of $63 million in funding to all fifty U.S. States, the District of Columbia, and seven U.S. territories. Program goals include preventing the initiation of tobacco use among young people, eliminating exposure to second hand smoke, promoting tobacco cessation among adults, and addressing tobacco-related disparities among certain population groups. Funding supports implementation of community interventions and mobilization, counter-marketing efforts, policy development, and surveillance/evaluation[24].

The implementation structure for the NTCP is fairly networked at the state and local levels often involving extensive collaboration with partners representing a broad spectrum of sectors and levels of government (e.g., school system, non-profit

---

[24] http://www.cdc.gov/tobacco/tobacco_control_programs/stateandcommunity/index.htm#about. (accessed March 7, 2007). In addition, HHS, CDC, Program Announcement 03022.

organizations such as American Cancer Society and the American Heart Association, health maintenance organizations, business associations, and other chronic disease programs). In many states, the CDC awards are augmented by other funding sources (e.g., state legislature allocations) often derived from tobacco excise taxes and/or the 1998 Master Tobacco Settlement Agreement.

In 2005, OSH published *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs* (CDC 2005) which identifies120 evidence-based, outcome-level, performance measures. Given the variability in implementation activities among grantees, programs have flexibility in selecting measures for which to report to CDC. More recently, OSH identified a subset of twenty-five of these key measures (i.e., core indicators) for which all grantees will be required to track and report to CDC.

### 3.5 Sample Selection: The Individuals

Persons involved in developing, implementing, and or managing the performance measurement system for each of the four programs were recruited for study participation. These persons included CDC staff, contract employees, and external stakeholders, including directors for state-based programs. It is common for CDC to contract with agencies for a variety of evaluation and research activities, including developing performance measurement systems. Likewise, CDC often involves external stakeholders (e.g., representatives of state or national organizations) in an advisory capacity for such efforts as these. Miles and Huberman (1994) recognize the value of including those "peripheral" to the case as well as those closer to the phenomenon being studied, and suggest that those on the periphery of the case offer a unique perspective and contribute to a stronger contextual understanding.

Purposeful sampling was used to select interviewees for each case. Interviewees represented a group with experience in the development, implementation, or management of the performance measurement system. Again, these individuals represented persons who had the most extensive experience with the performance measurement system or who otherwise offered an important perspective. To increase the comparability of data, an attempt was made to interview persons in similar positions or roles across the four cases. These roles and the number of people interviewed for each[25] included:

1. The person(s) involved in the development and/or implementation of the performance measurement system (n=19)

2. The person(s) responsible for the CDC data management system used to manage performance data reported by grantees (n=7)

3. The Branch Chief or Section Chief for the program services branch (i.e., the branch that oversees the cooperative agreements or grants for the national program) (n=7)

4. Program consultants from the program services branch who work directly with funded grantees to oversee the cooperative agreements or grants (n=11)

5. Policy staff responsible for managing GPRA and PART reporting (n=5)

6. Program Directors who manage the state-based programs[26] or CDC field staff that work in the grantee program (n=7)

The researcher identified persons who directed the development of the performance measurement systems for all four programs. These people served as "gatekeepers" to the larger program staff and stakeholders. Once human subjects' approvals were obtained in the fall 2007, the researcher met individually with each gatekeeper and formally requested their participation in the study. The researcher

---

[25] Some participants fit two roles (e.g., a program consultant who was on a performance measurement workgroup to help develop the measures).
[26] The researcher was limited by the number of non-federal employees that could be included in the study given the absence of OMB Paper Reduction Act approvals.

provided the gatekeeper an invitation to participate by email with information about the study. The gatekeeper assisted the researcher in securing Division approval for participation and identifying others important to the development, implementation, and/or management process. Potential interviewees were contacted by email and invited to participate in the study. The email explicitly stated that the research was being conducted to fulfill dissertation requirements and that human subjects' review had been obtained from both CDC and Georgia State University (GSU). Individual emails to participants were modified slightly to personalize them as appropriate.

## 3.6 Data Collection Methods

The research protocol was approved by the institutional review boards for both CDC and Georgia State University (GSU). As noted in the section addressing the strengths of qualitative research, the approach allows for a variety of methods. Rossman and Rallis (2003) suggested that case studies in particular are "methodologically eclectic." In fact, case study *requires* that multiple methods or sources of evidence be used (Yin 2009). The most common data collection methods used in case study are interviews, document review, and observation, although quantitative methods are also often incorporated (Patton 2002; Stake 2006; Yin 2009). These three methodologies were adopted here as a means to facilitate a holistic understanding of each case while also allowing for methodological triangulation of data, which supports the reliability and validity of findings (Yin 2009). Most importantly, a systematic, objective approach to data collection was practiced so that generalizations could be made about the case(s) as patterns and themes were identified (Stake 1995).

116

### 3.6.1 Interviews

In case study, interviews are an extremely important data source (Yin 2009), in part because they can provide such rich data. Interviewing aims to "discover the informant's experience of a particular topic or situation" (Lofland and Lofland 1995, p. 18). While various approaches to interviewing are described (Patton 2002), a semi-structured interview guide approach was adopted here. This approach involved combining topic-initiating questions that were derived from the research questions with follow-up questions or probes aimed at gathering more detailed information (Rapley 2001). The interview guide is included as appendix C. The guide assured that relevant subject areas were addressed consistently across interviews, while allowing flexibility for the researcher to probe in order to further clarify particular topics (Miles and Huberman 1994; Patton 2002). For multiple cases, Miles and Huberman advocate for a structured data collection instrument to improve internal validity and assure that comparisons can be made across cases.

Interviews were conducted in person, in a private office, typically at CDC. Written, informed, voluntary consent was secured from all interviewees. The consent form is included as appendix D. For individuals physically located in another city; telephone interviews were performed. Incentives were not provided for participation. All interviews were audio recorded with the permission of participants. Email was used to facilitate interview scheduling and to remind participants of interview dates and times. In all 55 people were invited to participate and, of those, 52 (95%) agreed to participate. Of the three who did not participate, two never responded to repeated emails and one recommended another individual as more appropriate for inclusion.

In sum, a total of 50 interviews were conducted with 52 individuals (table 7). Two people were interviewed together in two different interviews (i.e., a total of four people interviewed); the participants explicitly requested that they be interviewed together and represented similar roles (i.e., two policy staff persons; two data management staff). Interviews were conducted from January 22 through June 4, 2008. For the most part, interviews were conducted by "case"; that is, all 13 interviews with participants representing the CSPS were conducted before moving to the second case (PHEP), and so forth. The order in which cases were addressed in the study was established based on the availability of staff. The average interview length across the four cases was 67 minutes (range: 33-133 minutes). All interviews were transcribed verbatim by a professional transcriptionist.

**Table 7. Interviews Conducted by Case**

| Case | Number of Participants | | Total Interviewees | Average Interview Length (minutes) | Dates of interview data collection |
|---|---|---|---|---|---|
| | Male | Female | | | |
| Public Health Emergency Preparedness Program | 2 | 11 | 13 | 67 minutes | March 13, 2008 – April 4, 2008 |
| Comprehensive STD Prevention System | 8 | 5 | 13 | 69 minutes | January 22, 2008 – March 24, 2008 |
| National Breast and Cervical Cancer Early Detection Program | 5 | 8 | 13 (12 interviews) | 70 minutes | March 31, 2008 – May 14, 2008 |
| National Tobacco Control Program | 5 | 8 | 13 (12 interviews) | 57 minutes | April 23, 2008 – June 4, 2008 |
| Totals | 20 | 32 | 52 (50 interviews) | 67 minutes | January 22, 2008 – June 4, 2008 |

### 3.6.2 Document Review

In case study, documents provide valuable information that can corroborate data collected from other sources, provide additional evidence, and spawn new paths of inquiry (Patton 2002; Yin 2009). Documents may also, however, contradict data gathered through interviews and observation, encouraging the researcher to explore these

discrepancies in greater detail (Rossman and Rallis 2003). For this study, relevant

documents were systematically collected and reviewed. Program-specific web sites were

reviewed to identify potential documents for review, and participants in individual

interviews were queried about relevant documents.[27] Government websites were also

accessed and reviewed (e.g., Office of Management and Budget's ExpectMore.gov). A

document review or summary form (appendix E) was completed for each document

selected for study inclusion. Documents included the funding announcement for the

specific program, reports on the development of the performance measurement system,

summaries of stakeholder input on the performance measures, minutes from relevant

meetings, presentations about the performance measures, relevant policy documents,

guidance documents summarizing the selected performance measures, and monitoring

reports used to provide feedback to grantees among others. Appendix F lists the 57

documents and websites collected and reviewed for each case. Some cases provided a

greater number of documents than others. Given the researcher's role as a CDC-

employee, participants were generous in providing varied documents, including drafts of

documents and internal reports for which an external researcher may not have had access.

### 3.6.3 Direct Observation

Observation allows the investigator to enter the research context or setting and

explore its complexities (Rossman and Rallis 2003). For this study, data collected

through observation provided additional information about the performance measurement

systems, the stakeholders involved in their development, and the context in which they

---

[27] As a CDC employee, the researcher has access to both internal (CDC Intranet) and external (CDC Internet) web sites for CDC programs.

have been developed. Key stakeholders for each program were queried about

opportunities for observation during the data collection phase of the study. Observational

opportunities included internal CDC meetings of performance measurement workgroups,

external stakeholder meetings to assist in developing performance measures, conference

presentations, other presentations, and a Congressional hearing. A total of 12 formal

observations were conducted across the four cases with at least two conducted for each

case. Table 8 summarizes the observations conducted for each case.

**Table 8. Formal Observations Conducted by Case**

| Case | Observation | Date of Observation | Length of Observation |
|---|---|---|---|
| **Public Health Emergency Preparedness** | American Evaluation Association Conference presentation, "Getting from War Stories to Science: Developing Performance Measures in Public Health Emergency Preparedness" | November 2007 | 1.5 hour |
| | CDC Evaluation Forum Presentation, "Strategies for Ensuring Data Integrity in Performance Measurement: Lessons Learned from the Public Health Emergency Preparedness Cooperative Agreement" | December 2008 | 1.5 hours |
| | National Association of County and City Health Officials (NACCHO) Public Health Preparedness Summit presentation, "Are We Prepared?" | February 2008 | 1.5 hours |
| | National Association of County and City Health Officials (NACCHO) Public Health Preparedness Summit presentation, "Developing and Implementing a National PHEP Measurement System to Support Accountability and Program Improvement" | February 2008 | 1.5 hours |
| | Meeting of Evaluation Workgroup Sub-committee– Incident Management workgroup | April 2008 | Two days |
| | | | |
| **Comprehensive STD Prevention Program** | American Evaluation Association Conference presentation, "How Do You Keep It Going: Steps that One CDC Program Takes to Keep Performance Measures Relevant" | November 2007 | 1.5 hour |

| Case | Observation | Date of Observation | Length of Observation |
|---|---|---|---|
| | STD Performance Measurement Workgroup meeting | March 2008 | 2 hours |
| | | | |
| **National Breast and Cervical Cancer Early Detection Program** | U.S. Congressional Hearing – Committee on Oversight and Government Reform, U.S. House of Representatives: The National Breast and Cervical Cancer Early Detection Program | January 29, 2008 | 3 hours |
| | Meeting of CDC Minimum Data Element (MDE) Committee | April 2008 | 1.5 hours |
| | Montana Statewide Comprehensive Cancer Meeting – Presentation "The Montana Breast and Cervical Health Program) | May 2008 | 45 minutes |
| | | | |
| **National Tobacco Control Program** | Meeting of the Core Indicator Workgroup | May 16, 2008 | 1 hour |
| | Meeting of the Core Indicator Workgroup | May 28, 2008 | 1 hour |

Observation length varied depending on the context. A meeting of the STD performance measurement workgroup lasted approximately two hours, while an external stakeholder meeting convened by COTPER lasted two full days. The researcher took an unobtrusive approach to conducting observations rather than a participatory one (Patton 2002). An observation guide (appendix G) was used to collect field notes, including descriptive data about the physical environment, interactions among those observed, and

investigator's impressions and analytic insights (Rossman and Rallis 2003). The researcher took extensive field notes for each observation. These notes represent the observation as data or evidence, comparable to an interview transcript (Schwandt 2001). Raw notes compiled during the observation were further refined into more detailed summaries recorded on the observation guide following each observation.

### 3.6.4 Summary of Data Collection Methods

The majority of data collection was conducted between November 2007 and June 2008, although some document and website review continued through December 2008. Table 9 summarizes data collection for the entire study, which included a total of 50 interviews with 52 individuals, 57 document or web-site reviews, and 12 formal observations.

**Table 9. Summary of Data Collection**

| | Public Health Emergency Preparedness | Comprehensive STD Prevention Program | National Breast and Cervical Cancer Early Detection Program | National Tobacco Control Program | Total |
|---|---|---|---|---|---|
| Number of Interview Participants | 13 | 13 | 13 | 13 | 52 |
| Number of Interviews Conducted | 13 | 13 | 12 | 12 | 50 |
| Number of Documents Review | 15 | 15 | 14 | 13 | 57 |
| Number of Observations with Field Notes | 5 | 2 | 3 | 2 | 11 |

### 3.7 Data Management

Given the extensive data that were collected through the methods detailed above,

attention to effective data management was essential. A well-organized database supports

the study's reliability by assuring that other investigators can directly review the case

study evidence (Yin 2009). In addition, effective data management is critical to

supporting analysis. A number of strategies were used to ensure an efficient approach to

data management.

1.  Microsoft Excel – Detailed Excel spreadsheets were maintained to track all data
    collection and analysis efforts according to the individual case. The spreadsheets

included details about the data collection method (e.g., observation, interview, document review); relevant dates and times (e.g., dates of correspondents with participants, interview date/time/place, observation date); participant names, pseudonyms, and contact information; data file names for audio recordings and transcriptions; document names with corresponding electronic document file names; etc. Excel was also used to develop the analytic codebook and facilitate case and cross-case analysis.

2. Atlas.ti Scientific Software Atlas.ti was used to facilitate data analysis. Atlas.ti allows for the efficient coding and retrieval of data, along with other analytic functions (e.g., content analysis, memoing, mapping relationships). All text-based data was stored in Atlas.ti, including interview transcripts, document review summaries, and observation summaries.

3. Microsoft Word Microsoft Word was used to maintain all electronic documents including interview transcripts, document review summaries, observation summaries, a researcher's journal, the analytic codebook, and the written chapters developed for inclusion in the final dissertation.

4. Digital audio recordings – Sony digital audio recorders were used to record all interviews. The digital audio files were maintained on a personal computer and deleted once a written transcript was completed.

5. Electronic data – All electronic data were maintained on a personal computer, password protected, and backed up daily to a FireLite Smartdisc.

6. Hard copies Hard copies of some documents were maintained separately (e.g., reports, power point presentations received at observations events, signed informed consent forms) in a locked storage cabinet.

### 3.8 Data Analysis Procedures

Although analysis is, to some extent, intuitive and an "art," there are systematic, analytic methods that can be applied to all evidence gathered that assure a "scientific" approach (Creswell 2007; Merriam 2009; Miles and Huberman 1994; Schwandt 2001; Stake 2006; Yin 2009). As noted earlier, in qualitative research, analysis is conducted simultaneously with the data collection effort. Emerging themes generated from one interview are explored in future ones; something observed leads to reviewing another document; analytic insights are recorded; and potential themes are explored with other

participants. This is possible given both the researcher's role as the primary instrument for data collection and the flexible nature of qualitative research. The researcher terminated data collection based on practical considerations related to resources (i.e., time) and also influenced by the number of available sources, research limitations (i.e., inability to interview more than nine non-federal stakeholders), and data saturation.

Rossman and Rallis (2003) identified seven analytic procedures to guide analysis that were followed for the study. These include: (1) organizing the data; (2) becoming familiar with the data; (3) generating categories and themes; (4) coding the data; (5) interpreting the data; (6) searching for alternative explanations; and (7) writing the dissertation. As described below, these procedures are not necessarily conducted in a linear manner, but rather, iteratively.

### 3.8.1 Organizing the Data

A central challenge in qualitative research is making sense of such vast amounts of data (Miles and Huberman 1994; Patton 2002). This challenge was relevant for the current study given the vast amount of data collected and the involvement of a single researcher. Several approaches, described above in section 3.7, were used to organize the data, making it more manageable. In particular, the data were organized by each case and Atlas.ti allowed the researcher to filter data in various ways (e.g., by creating "families" of data by case). Therefore, data were easily grouped and sifted to facilitate analysis (e.g., by case, participant role, related codes).

### 3.8.2 Familiarizing Yourself with the Data

Knowing the data intimately requires immersion in the data itself. For this study, the investigator conducted all research activities including data collection and the development of analytic memos maintained in a journal. In addition, once the written interview transcript was received back from the transcriptionist (typically within 2-3 days), the researcher vetted it against the original audio recording, allowing a second "listen" of the interview and assuring accuracy of the transcription..

Further immersion occurred as part of codebook development and data coding. For instance, to construct the codebook (see 3.8.3), the researcher "open coded" over half of the interview transcripts (n=29), again offering an opportunity to closely review the data. Next, the researcher coded all interview transcripts, document review forms, and observation field notes – this work continued to facilitate a deeper understanding of the data. Finally, one other strategy supported the researcher's constant reflection on the data. The investigator maintained an analytic journal throughout the study where ideas were recorded about emerging insights, potential themes, and methodological issues (Glesne and Peshkin 1992; Stake 2006). An example of an entry in the journal is noted below:

> *March 29, 2008: Lack of direct line control over local level implementers (and network partners) diminishes perceived influence on the part of CDC on grantees' performance, data collection and reporting for the performance measures, and resulting performance. CDC managers and grantees seem very cognizant of what is and isn't within the grantees' control -- issues of "fairness" are frequently cited by grantees.*

As part of the journal, the researcher tracked key decision points made by the researcher and important reflections on the researcher's role in data collection and analysis (Janesick 2000).

*April 20, 2008: I initiated NBCCEDP interviews in mid March – Feona invited me to present on the study at one of their all-staff branch meetings. Everyone has been very supportive of the study. As with the other cases, I interviewed two program directors (PDs) who were well versed in the topic; they were two long-serving PDs, one of which is the current NBCCEDP Program Director Council chairperson.*

All of these efforts – data collection, vetting the transcripts, open coding to develop the codebook, final coding of all data, and maintaining an analytic journal – contributed to the researcher's effort to maintain an intimate understanding of the data.

### 3.8.3 Generating Categories and Themes

As data were collected, potential categories and themes were identified and noted in the researcher's analytic journal. For instance, while interviewing participants in the first case, a potential theme emerged around the notion that, given the network context, performance measurement systems evolve over time – that is, the systems may develop in an incremental fashion, gaining in complexity and sophistication. As themes like this were identified, the researcher documented them in the analytic journal and began to explore them further in other interviews, document review, and observations. Similarly, categories emerged during data collection. For instance, the researcher observed varied descriptions of the *use* of performance measurement data in networked contexts as well as different aspects of the *process* of developing performance measures in networks.

The process used by the researcher to develop the codebook also facilitated the identification of categories and themes. Although there are different approaches to building codes, two methods were used here. First, a priori codes were identified deductively based on the theoretical framework, research questions, and interview questions (Miles and Huberman 1994). For example, the descriptive codes "horizontal

networks" and "vertical networks" were drafted given their significance to the research questions themselves.

Next, codes were developed inductively from the text using open coding, a technique first described by Strauss and Corbin (1990), but also by Charmaz (2006). Open coding involves creating codes tied to the data itself by reading the text and identifying codes directly related to that text, thus preserving elements of the study context (Miles and Huberman 1994). The approach helps ensure that the analysis is "grounded" in the data. This two-level strategy to coding, a priori coding and open coding, results in both "etic" codes that are more conceptual and broad in nature, as well as "emic" codes, those closer to the data and participants' perspectives (Miles and Huberman 1994). The open coding technique was especially valuable in deriving categories and themes from the data itself.

While some (Charmaz 2006) advocate the use of gerunds and a rigorous line-by-line coding for the first step of open coding, a less painstaking approach of deriving codes or categories was used here (Hsieh and Shannon 2005; Patton 2002). More specifically, an initial round of open coding was conducted on a subset of just over half the transcripts for each case using Atlas.ti (n=29 total transcripts). In this exercise, initial codes were developed based on the actual text although every line of text was not coded. Based on the open coding of 29 interview transcripts, over 1400 initial codes were developed and were entered into an Excel spreadsheet. At this first stage, the "codes" were fairly raw in form including some that were in-vivo text (e.g. "at CDC this program is so decentralized," "networks challenge ability to see who has responsibility for performance") and others that were more descriptive (e.g. challenge in measuring a less

direct service, no consensus on what to do). This inductive approach to coding resulted in codes that were emergent and, as noted earlier, directly tied and "grounded" to the data.

Using the constant comparative method, originally described by Glaser and Strauss (1967) as part of their grounded theory approach, the researcher reviewed and compared the initial 1400 open codes. Atlas.ti allowed the researcher to move easily from a code back to the original text in order to maintain a contextual perspective. By comparing codes with codes, data with data, and codes with data, the researcher was able to begin sorting and grouping codes into broader themes and categories. This step allowed the researcher to move toward developing codes that were more conceptual in nature. Charmaz (2006) describes this phase as "focused coding"; Miles and Huberman (1994) describe it as "pattern coding" and suggest it is especially valuable for multiple case studies as it helps begin to identify common themes.

This second phase of reviewing, sorting, and organizing reduced the 1400 initial codes to 147 codes that were grouped around topics or categories. For instance, under the broad category of "networks," the researcher identified the following codes: building networks; conflicts in networks; policy tools to support networks; managing networks; value of networks; challenges of networks; unintended consequences of networks; compromise in networks; and competition in networks.

Microsoft Excel® was used to help sort and organize the codes. The 1400 original open codes were maintained in the Excel spreadsheet as data were grouped allowing for detailed examples to be maintained. For instance, as illustrated in table 10, the researcher identified an emerging theme during this phase of analysis and codebook which she titled

"conflicts in networks." The individual cells in the table are filled with some of the

original open codes that the researcher grouped under this theme.

Table 10. Codebook Development: Conflicts in Networks

| Can be competing priorities to PM at grantee level* | Can be different agendas between states and locals | Local -federal tensions | Differences in priorities for feds vs. states | Different missions for jails | Do we have any agreement on what we're trying to do? |
|---|---|---|---|---|---|
| I don't even know if CDC is on the same page | Network partners can impede public health performance FBI wouldn't let CDC in during Katrina | Public health's voice can be lost in preparedness network/arena | Philosophical battle between control and influence | Political demands to bend to grantees' demands too often | Politically it would be hard to raise the targets |
| Problems of turf | Some doctors view toward the health department is I don't have to talk to them | Strongly influenced by Dept of Homeland Security and Dept of Defense | Varied disciplines in preparedness result in different interpretations, reflect different cultures that effect PM | We disrupt the routine of jails when we come in | How does public health contribute in the larger network response? |
| Who pays for testing in the jails? | Whose performance are you measuring? | | | | |

*Each cell is filled with an "open code" which has been organized under the broader code "Conflicts in Networks"

Next, this smaller set of 147 codes was again closely reviewed by the researcher and new categories or themes were identified by collapsing, condensing, or expanding certain codes (Creswell 2007). During this third wave of comparisons, the 147 codes were reduced to a group of 60 codes. Again, the original open codes were maintained in the Excel spreadsheet, providing detailed examples for each of these 60 codes. During this phase, conceptual categories emerged for the codebook; Miles and Huberman (1994) suggest such a structure is essential. Groups of codes began to naturally fit together under these larger categories. For instance, there were a number of codes which reflected the *process* of developing performance measures; other codes congealed around the issue of *measurement* while still others reflected notions of *cultural shifts* and *network characteristics*. Particular attention was given to developing codes not only at the descriptive level, but also at deeper, more analytic levels reflected in the categories and themes (Charmaz 2006; Schwandt 2001).

After this round of sorting and grouping, the code "conflicts in networks" that was developed in the second stage (see table 10) was abandoned. Instead, a new code of "value and goal conflicts" was created under a broader category of "network characteristics." Several dimensions of networks were included under this category, including the a priori descriptive codes such as "network characteristics: vertical relationships" and "network characteristics: horizontal relationships," but also inductive codes related to process such as "network characteristics: bargaining, consensus building, and relationship building." Table 11 below reflects the original open codes maintained for the new code, "Network Characteristics: Value and Goal Conflicts."

**Table 11. Codebook Development: "Network Characteristics: Value and Goal Conflicts"**

| | | FEMA region vs. public health regions are different and have different focus | Program crosses so many disciplines big challenge | Their mission is public safety, ours is public health |
|---|---|---|---|---|
| There are so many fingers in the pot | Jail partners need to see a benefit | | | |
| Trying to understand the critical linkages between public health and hospitals | Networks can be complicated by things like mixed areas of responsibility | Can be competing priorities to performance measurement at grantee level | Can be different agendas between states and locals | COTPER influenced by homeland security value system vs. public health has a different value system |
| Differences in priorities for feds vs. states | Different missions for jails | I don't even know if CDC is on the same page | Making arguments with jails that public health is important | Syphilis elimination is HIV prevention |

The final phase in this step of analysis to form the codebook involved the development of operational definitions for all 60 codes. These definitions are essential to ensure that the researcher applies the codes consistently over time (Miles and Huberman 1994). The researcher adopted an approach to codebook development described by McQueen and colleagues (2008), public health researchers. Definitions included seven components – a code name, brief description, long description, when to use, when not to use, coding rules, and examples. The Excel spreadsheet, with its elaborate detail of sorted open codes, proved invaluable to the development of the operational definitions. Table 12 illustrates the codebook definition for "network characteristics: value and goal conflicts."

**Table 12. Example of Codebook Definition**

| | |
|---|---|
| **Code Name**: | Network Characteristics: Value and Goal Conflicts |
| **Brief Description**: | Conflicts/issues between network partners around values, goals, mission, priorities, etc. |
| **Long Description**: | Given the network structure, conflicts or issues may arise in developing performance measurement related to differing agency values, goals, mission, priorities, areas of responsibility / turf, culture, etc. These conflicts may impede the development of a common set of measures or the collaboration needed to collect/report performance related data. |
| **When to Use**: | Apply this code when text refers to or describes tension between network partners that is related to fundamental differences in values, goals, etc. |
| **When NOT to Use**: | N/A |
| **Coding Rules**: | N/A |
| **Example**: | "Human nature, again, I could be part of this myself if I was out there, I have been out there in the past, but you've got your own little kingdoms and queendoms and fiefdoms and all that and the HIV STD programs are a great example of that. They're still not really working together out there and it's now 2008."<br><br>"You know, I think each State determines who their screening population is going to be, and in [state name] it's not just the Department of Health at the site, we have stakeholders, and we have a really strong Komen presence in our State" [referring to partners who want women aged 40-50 screened vs. CDC policy of screening women 50-64]. |

Once the draft codebook was completed, the researcher coded a subset of eight interviews (i.e., two from each case), in order to further refine the codebook. During this time, definitions were revised and four new codes were added. The final codebook, then, included 64 codes (see appendix H).

135

### 3.8.4 Coding the Data

The final codebook was entered into Atlas.ti so that the full definition (i.e., all seven components) was easily visible as the researcher coded the textual data. Coding data involves chunking the text into more manageable segments and attaching a code to it (Bogdan and Biklen 2007; Miles and Huberman 1994; Patton 2002). Charmaz (2006) suggested that coding represents the "analytic frame from which you build the analysis" (p. 45). The researcher coded all data for each case in sequence; that is, all interviews, document review forms, and observation field notes were coded for one case before moving on to the next. This approach supported the tenet of immersion in qualitative data the researcher was able to focus entirely on the data collected for each case. Throughout the coding process, the researcher developed "comments," a feature supported by Atlas.ti. The software allowed the researcher to attach comments to particular segments of text, offering the opportunity to record insights during the coding process. Those comments were saved in the database and could be easily retrieved. Data coding proved a critical analytic process allowing the researcher to both examine the whole as she reviewed each transcript, document review form, and observation field notes, but also to extrapolate data and attach codes to build empirical evidence for categories and themes observed in the data.

### 3.8.5 Interpreting the Data

Rossman and Rallis (2003) differentiate analysis from interpretation. They suggest that while analysis involves coding to organize the data, interpretation involves the more complicated process of making meaning from the data (Rossman and Rallis 2003). As a process of meaning-making, interpretation is shaped by the research

questions, conceptual framework, and ideas that were the starting point of the inquiry (Peskin 2000). The efforts to develop comments during the coding process (described above) and maintain an analytic journal throughout the study represent strategies that support interpretation.

An important analytic strategy involved "memoing" throughout the coding process. Atlas.ti includes a memo feature that allowed the researcher to develop and maintain numerous memos. A method adopted from grounded theory, the writing of memos promotes continual reflection on the data. For this study, memo writing helped maintain a focus on the data, capture analytic insights, note relationships between and among codes, make personal reflections on the researcher's analytic process, and continually document new ideas and impressions. Given that the research was conducted by a single investigator, memoing proved invaluable in increasing the level of abstraction in analysis. Although memos are maintained as part of the Atlas.ti database, memos were viewed as an extension of the analytic journal maintained by the researcher.

Specific "memos" were developed for each case (e.g., NBCCEDP, PHEP) as well as for some key topics (e.g., "control within networks," "accountability"). The writing process itself, including efforts to memo, was fundamental in elucidating a more nuanced analysis. The researcher used the memo function of Atlas.ti to record free flowing thoughts as the coding proceeded rather than fully developed analytic deductions. An example of a memo is provided below:

> Coding Memo: June 2008 "The data here seems to suggest that with "wicked problems" and in a networked context, it takes **time** to develop more sophisticated performance measures -- the measures evolve over time from more simple, "low hanging fruit" to more complex measures that stakeholders at different levels can accept and deal with. Over time (i.e., it's a process!), data collection and management systems are

developed, the field integrates the practice of data collection and reporting at various levels (with PHEP they're starting just at the state level, but want to move to measures at the local level), developmental measures are piloted and tested, analytic methods develop that improve measures, the science base evolves, capacities develop in the field to collect and report data in ways that support data validity, and stakeholders build an appreciation for the utility of the data."

Again, the constant comparative method was used extensively to aid interpretation. Comparisons of data were made at various levels. At first, data were compared with data within an individual interview to identify similarities and differences and to assess its relevance (Charmaz 2006). Using Atlas.ti, coded data were easily sorted in multiple ways to help make comparisons across interviews, cases, and data collection methods. Atlas.ti also provides a function to calculate code frequencies which helped assess the potential strength or importance of specific categories and themes. Similarly, co-occurrences of specific codes could be assessed; for instance, the researcher could examine all text that was coded as both "network characteristics: horizontal dimension" and "design: control over performance on measures."

In multiple case study, data analysis must attend to both the individual case and the aggregate of cases. Stake (2006) addresses the tension that exists in balancing these two imperatives. Given that performance measurement will be better understood through its examination in unique contexts, analysis of each individual case is essential (Stake 2006). Yin (2009), Merriam (2009), and Stake (2006) all emphasize the importance of first conducting "*with-in* case analysis" that includes individual case descriptions that convey a holistic understanding. Each case provides an opportunity to study how the phenomenon operates under specific conditions (Stake 2006).

Consequently, attention to within case analysis preceded cross-case analysis. Once coding was complete for a particular case, the researcher developed an individual case description and summary before moving forward to code the next case. This approach allowed the researcher to maintain a concentrated focus on each case. The case description and summary were organized in a report format so that it could be returned to representatives from each program for review. With that audience in mind, the format used included a brief introduction to the study and its purpose; a review of the methods and data collection; three main sections organized around 1) the program, 2) the networks, and 3) the performance measurement system; and an overall summary of the case. In writing the case reports, the researcher reviewed data associated with specific codes to develop the individual report sections (e.g., the networks); that is, data for the same set of specific codes associated with a report section were reviewed and analyzed for each case. This approach was meant to ensure a consistent approach to analysis across the four cases.

As the case description reports were completed, they were sent electronically to the person who had helped coordinate the case study for his or her review and member checking. The investigator encouraged that individual to share the report with others for review and comment. Member checking involves engaging the research participants in a review of tentative findings or interpretations in order to assess their plausibility (Merriam 2009; Creswell and Miller 2000). Lincoln and Guba (1985) suggest that member checking is the most important technique to establish credibility and support internal validity. Following the review of the report, the researcher met individually with the key program or evaluation staff to discuss the accuracy of the report. A minimum of

two people from each case reviewed their report and provided comment. Appropriate revisions were made based on the meeting and the final report was provided to the individual program.

Once the individual reports were verified through the member-checking process, the researcher developed more in-depth findings for each case based on the research questions. Using Excel spreadsheets, the researcher first developed extensive and detailed matrices for each of the three major areas used to describe the case – the program, the networks, and the performance measurement system. The matrices summarize descriptive data for each of these three areas and are included as appendices (program characteristics are summarized in appendix I; network characteristics in appendix J; and the performance measurement systems in appendix K).

Next, the researcher developed a list of potential findings for each case according to each of the three research questions – this process was aided by the matrices described above. The potential findings for each case were organized in an Excel spreadsheet (appendix L) and then sorted and combined into more formal statements that comprise the findings for each case. "Cross-case analysis" was principally guided by the research purpose, theoretical framework, research questions, and findings from the individual cases (Stake 2006). More specifically, the cross-case analysis considered the findings across the four cases in order to make assertions that fit across the four cases while also preserving the "situationality" of the individual case findings (Merriam 2009; Stake 2006; Yin 2009). A process described by Stake (2006) to facilitate the cross-case analysis was used.

In the cross-case analysis, the researcher applied findings from the individual cases to the study's topic of interest – that of performance measurement applied in networked contexts. Working from the individual case findings, the researcher assessed the potential prominence of one case or another in addressing the research questions. Given the unique aspects of each case (e.g., the extent of its network, the sophistication of its performance measurement system) some cases were more relevant to a specific research question than another. As before, Excel spreadsheets helped to organize this process – individual sheets were used to develop a matrix for each hypothesis (Jennings and Haist 2006) and related research question. Matrix columns included those for case findings (tagged by case name), the effect of the specific finding on the performance measurement system, potential broader implication of the finding for performance measurement, specific evidence for the finding, relevant information about the case context (i.e., case situationality), other cases that support the finding, and other cases that counter the finding (see appendix M).

By using this systematic approach, the researcher was able to move from the individual case reports to develop the cross-case findings and related assertions addressing the study's research questions. Also, by tagging the individual findings by the case from which they were derived, the researcher could take into consideration the importance of particular cases in addressing specific research questions and how typical or atypical that particular case was to the study. Developing the assertions related to the research questions was not only a matter of assessing what was common across the four cases, but also considering the unique findings of each case (Stake 2006) – both advanced understanding. The intention was to develop assertions that had a single focus, provided

an orientation for understanding the research topic, and were supported by evidence (Stake 2006). In summary, the multiple sources of data represented in the four cases were used to help clarify meaning and build more complex explanations reflected in the key assertions. The data displays (i.e., matrices developed as excel spreadsheets) described above helped to draw conclusions at this stage.

### 3.8.6 Searching for Alternative Explanations

During analysis, rival explanations were identified and considered (Yin 2009). For example, as data were collected for the NBCCEDP, the researcher formed an explanation about the use of the performance measures for budgeting purposes. In subsequent interviews, rival explanations were explored to further understanding of the issue. In addition, negative instances those situations that are not consistent with the emerging explanation were explored and used to refine understanding (Merriam 2009). For instance, the researcher explored why one participant had a different perception about potential gaming of performance measures than the others interviewed. The constant comparative method described above helped to identify negative or discrepant cases. By consciously exploring competing explanations and negative cases, a stronger, more logical explanation that is well grounded in the data was developed (Rossman and Rallis 2003).

### 3.8.7 Writing the Dissertation, Report, or Manuscript

The analytic process continues through writing and reflects the interpretive act of bringing meaning to the data through narrative (Rossman and Rallis 2003). Good writing is critical in qualitative research as understanding, explanation, and findings are

interpreted through text (Richardson 2000). In particular, the researcher must represent

the findings in rich detail in order to communicate meaning to the reader (Brower,

Abolafia, and Carr 2000). Qualitative writing should reflect three qualities according to

Golden-Biddle and Locke (1993) – authenticity, plausibility, and criticality. First, the

account must be authentic, providing a rich description to assure the reader that the

investigator has done sound research. Second, the account should provide a plausible

explanation to the reader, that is, have face validity. And third, the writing should

encourage and challenge the reader to think critically and deeply. For this study, the

researcher attempted to attend to these qualities as guiding tenets for the writing.

As noted above, individual case reports were developed for representatives of the

participating cases. The researcher, in consultation with her dissertation chairman, chose

to present findings of the four cases in individual chapters (chapters 4-7). This approach

seemed most appropriate given the unique characteristics of each case, their importance

to the study, and the length of the initial reports. Others (Yin 2009; Stake 2006) have

supported this approach. Like the initial reports developed for each site, a common

structure was adopted for these four chapters to facilitate both readability and cross-case

analysis – such an approach is consistent with a linear-analytic structure described by Yin

(2009).

The findings for each of the four programs are presented in chapters 4 through 7.

A summary of the findings for all four cases is provided in appendix N. Each chapter

begins with a case description followed by a detailed presentation of the findings and

then a chapter summary. The case description includes a synopsis of the overall program,

its implementation network, and its performance measurement system. The typology used

is organized based on the literature review as well as inductive analysis of the data

collected (table 13).

**Table 13. Typology Used for Case Description**

| The Program | CDC organizational context |
| --- | --- |
| | Program goals |
| | Stage of program development |
| | Budget stability |
| | Stakeholders |
| | Political context |
| The Implementation Network | Network structure: vertical relationships |
| | Network structure: horizontal relationships |
| | Network function: authority and control within the network |
| | Network function: shared organizational goals and priorities within the network |
| | Network function: variability in context, resources, capacity |
| Performance Measurement System | Process to develop the performance measurement system |
| | Performance measurement system design including:<br>• Purpose<br>• Level of measurement<br>• Types of measures<br>• Use of targets or standards<br>• Quality assurance efforts |
| | Use of performance measurement system and data |

The literature review suggests that network public management has emerged, in

part, based on the complexity of the social and health problems faced today.

Consequently, the description of each case begins with a narrative about the program. Inductive analysis identified six common characteristics of each program: CDC organizational context, program goals, stage of program development, budget stability, stakeholders, and political context.

Next, the implementation network is described based on five characteristics, two related to the structure of the network and three related to the network's function. These characteristics were identified both through the literature review and inductive data analysis. For network structure, both the vertical and horizontal dimensions of the network are addressed. For network function, three aspects are considered – issues of authority and control within the network; shared goals and priorities between network members; and topics of context, resources, and capacity.

Finally, the performance measurement system is described including the process for developing the system used by each program, its overall design, and use of the performance measurement system and data. Five aspects of the performance measurement design are described including its purpose, level of measurement, types of measures, use of targets or standards, and quality assurance efforts.

Following the case description, each chapter concludes with a summary of case findings, including evidence for each. As noted above, development of individual case findings were directed by the study's research questions. Chapter 8 includes a descriptive summary of the four cases and presents findings from the cross case analysis organized by the three research questions.

### 3.9 Data Validity and Reliability

Trustworthiness is proposed by Lincoln and Guba (1985) as an appropriate criterion to assess qualitative inquiry that differs from traditional positivist criteria for assessing research quality. The trustworthiness construct replaces traditional views of validity and reliability with credibility, transferability, dependability, and confirmability (Lincoln and Guba 1985). Others offer varied typologies of validity (Maxwell 1992; Schwandt 2001). Merriam (1995) states, "Unlike experimental designs in which validity and reliability are accounted for before the investigation, rigor in qualitative research derives from the researcher's presence, the nature of the interaction between researcher and participants, the triangulation of data, the interpretation of perceptions, and rich, thick description" (p.151).

### 3.9.1 Internal Validity

Internal validity is conventionally defined as the congruence between one's findings and the reality of what is studied (Merriam 2009). Validity refers to the interpretations drawn from the data, not the data themselves (Creswell and Miller 2000). For this study, internal validity is based on how well the investigator represents the various perceptions and interpretations of those included in the study (Lincoln and Guba 1985; Merriam 2009; Creswell and Miller 2000). Several strategies were used to improve the internal validity of the proposed study, including emersion in the field, triangulation, member checking, searching for disconfirming evidence, and colleague examination.

As discussed earlier, emersion in the field refers to the researcher's engagement in the research context in order to understand the phenomenon of interest holistically and in

146

all its complexity. For this study, the investigator gathered primary data over a seven-month period and used varied data collection methods to facilitate a deep understanding of the research topic. In addition, the researcher conducted all aspects of analysis, many of which supported a continued "closeness" to the data that benefits internal validity.

Triangulation verifies the repeatability of an interpretation or observation to confirm findings (Merriam 2009; Stake 2006). Different types of triangulation have been described in the literature (Mathison 1998; Patton 1999), and two approaches were applied here. First, methodological triangulation, considered the strongest form of triangulation, was applied. Methodological triangulation involves assessing data from the multiple data collection methods (e.g., interviews, document review, and observation) (Lincoln and Guba 1985; Patton 2002). Another form of triangulation, data source triangulation, was also used; this form uses several data sources (e.g., data from more than one person or data collected from more than one point in time) to strengthen findings (Lincoln and Guba 1985; Patton 2002). Both approaches increase validity by ensuring a systematic process of searching through the data to identify common themes and by relying on more than a single source of evidence (Creswell and Miller 2000). Atlas.ti was helpful in assessing triangulation given its capability to sort text in varied ways and produce reports that identify the unique data sources.

Aside from concerns for validity, triangulation is also useful during data analysis to clarify meaning by considering multiple perspectives (Stake 2006). Mathison (1988) suggested that triangulation can result in three possibilities including convergence, inconsistency, or contradiction. As triangulation identified inconsistencies or

contradictions in results, the investigator considered plausible explanations to account for them. This approach actually helped to extend and enrich meaning and understanding.

Member checking, discussed earlier, was also used to strengthen internal validity (Merriam 2009; Creswell and Miller 2000). Lincoln and Guba (1985) suggest that member checking is the most important technique to establish credibility. As noted above, individual case reports were developed and disseminated to each program. The researcher then met individually with key representatives to discuss the accuracy of descriptions, the plausibility of case summaries, and whether adequate evidence was provided in support of the findings. As noted, a minimum of two participants from each case reviewed the reports. Participants provided detailed feedback in the face-to-face meetings, typically providing written edits on the report or an electronic copy of the report with tracked changes.

Identifying negative or disconfirming evidence involves searching through the data to either confirm or disconfirm preliminary themes or categories (Miles and Huberman 1994). Patton (1999) describes both inductive and logical approaches to searching for rival explanations and negative cases. This approach strengthens validity by further supporting the credibility of findings (Creswell and Miller 2000). For this study, the identification of negative evidence served to facilitate deeper exploration of the data and generate more insightful and nuanced understanding.

Finally, the researcher enlisted a colleague at CDC to examine findings and comment on their perceived credibility (Merriam 2009; Patton 2002). The colleague reviewed all case-specific chapters and the cross-case analysis chapter, offering suggestions and identifying areas of clarification. In addition, presentations of

preliminary results were made at the American Evaluation Association annual conference (November 2008) and at an evaluation forum at CDC (December 2008). Colleagues were especially helpful by raising important issues that encouraged thoughtful consideration about the analysis, the researcher's role in the study, and about how findings are conveyed in the writing.

### 3.9.2 External Validity

In research involving quantitative methods, external validity refers to the generalizability of the findings. As discussed earlier, qualitative findings are not intended to be generalized but to expose a deeper understanding of the phenomenon of interest at a particular time and in its unique context. Although efforts have been described to strengthen generalizability in qualitative research (Miles and Huberman 1994; Yin 2009), an alternative conceptualization was applied for this study. "User or reader generalizability" is a way to conceptualize external validity. Lincoln and Guba referred to this as "transferability" (Lincoln and Guba 1985). This concept is based on the idea that the reader or user of the findings has responsibility for determining how well the results transfer to his or her own situation (Merriam 2009). Given this view, two techniques were applied to strengthen the external validity of this study. First, the dissertation findings reflect rich, thick description that enable readers to determine how well the findings might fit their own situations (Lincoln and Guba 1985; Merriam 2009). Second, the multiple case study design contributes to extending variation in the study of performance measurement and allows readers to assess a greater range of experiences and contexts (Merriam 2009).

### 3.9.3 Reliability

In quantitative research, reliability refers to whether results would be the same if the study were replicated. Given the nature of qualitative research and the fact that human behavior is dynamic and fluid, this view of reliability is inconsistent with the approach (Merriam 2009). The idea of reliability has been recast by some as "dependability" and "consistency" (Lincoln and Guba 1985). That is, reliability involves assessing whether the findings are consistent with the data collected (Merriam 2009). To strengthen the reliability in this study, the researcher: (1) maintained a detailed audit trail as part of the researcher's analytic journal to assure transparency in research methods; (2) used data source triangulation; (3) used methodological triangulation; and (4) involved peer review (Merriam 2009). All of these have been described previously except the use of an audit trail. An audit trail was maintained as part of the researcher's analytic journal to document the research and included the following: (1) listing of all data collection efforts; (2) summary of data reduction and analysis products (e.g., codebook, field notes); and (3) process notes including key decision points made throughout the research process. Other written sources supporting reliability include the study proposal, data collection instruments, and final reports and written analysis (Lincoln and Guba 1985).

### 3.10 Ethics

As noted earlier, the study protocol was approved by institutional review boards at both GSU and CDC. High ethical standards were maintained throughout the research process. The written, verbatim transcripts were returned to the interview participants whom were offered the opportunity to review the transcript and remove any statement(s)

he or she was not comfortable including. Pseudonyms were used to help protect

confidentiality – all names were replaced with pseudonyms in transcripts, document

review forms, and observation field notes.

The thick description inherent in qualitative research presents an important

challenge to "inferred" identification (Anastas 2004). Even without identifying

information, there is the possibility that certain readers may infer a specific participant.

Participants were made aware, in the informed consent form and verbally at the time of

the interview, that their name would not be used in any written reports, but that only the

overall program (e.g., NBCCEDP) would be associated with specific statements. And, as

described earlier, several steps were taken to assure proper data storage (e.g., password

protected laptop for electronic data, locked file cabinet for hard copies of data).

## 3.11 Researcher Assumptions and Biases

While limitations refer to issues related to methodology, qualitative research must

also address the researcher's personal assumptions and biases that he or she brings to the

research process (Peshkin 1998). Addressing researcher reflexivity is another strategy

viewed as strengthening the credibility and validity of the research (Creswell 2000;

Mauthner and Doucet 2003). Mauthner and Doucet (2003) advocate for researchers to

practice greater reflection and accountability in their work, suggesting that a higher level

of self-consciousness will lead to improved confidence in the findings. The need for self

reflection is essential given that the researcher represents the primary instrument of data

collection in qualitative research. For this section, then, first person is used to describe

my own subjectivities that may potentially have influenced the conduct of this research

and my interpretations of findings.

Although I am a doctoral student at GSU and GT, I am also an employee at CDC. Therefore, while I conducted this research as an affiliate and representative of GSU, I also was conducting it as a CDC employee, which raised a potential conflict. One result of this dilemma was to disclose these relationships to participants which I did in the introductory e-mail to all participants and within the informed consent form. Given my working relationship at CDC, there were also concerns about the willingness of participants to be open about their experiences. However, I found participants to be incredibly transparent, openly sharing their perspectives and providing me internal documents and materials unlikely to have been made available to an "outsider."

Another issue related to my fifteen years of work experience at CDC. This experience contributes to underlying generalizations, assumptions, and perspectives I hold about CDC's national public health programs. For instance, I served as a program consultant for 7 years, working closely with state and city health department officials. While this experience provides me a deeper understanding of the program context, I also had to recognize my own preconceived judgments about those programs and remain vigilant in my awareness of when these assumptions affected my perceptions. For instance, I have developed strong views about the importance of providing state and local partners scientific guidance while also allowing for local tailoring of community-based public health programs in order to fit their unique context. While my experiences at CDC certainly enhanced my ability to relate to participants and more quickly assess and understand certain situations or dynamics, I tried to stay aware and avoid making inappropriate conclusions based on preconceived notions.

In my proposal, I wrote that I had concerns about the current emphasis on indicator development at CDC, which I often saw at the expense of other evaluation approaches. While I stayed attuned to this perception, my conversations with participants led me to overcome this particular bias and recognize the importance this strategy offers public health programs. Although I believe that there are unique challenges to implementing performance measurement systems at the federal level and am concerned that advocates lack an understanding of the complexity of its application in these setting, I realized a new appreciation for performance measurement for program management and evaluation.

Finally, my theoretical framework introduced a necessary bias that framed both what I explored and how I interpreted the data collected. While this is appropriate and necessary, it is important to remember that I explored this issue from a particular and specific lens.

### 3.12 Study Limitations

Several limitations must be acknowledged. First, the results of this study are applicable only to the four cases studied. The small sample of cases that were included in the study, as well as the limited number of participants, precludes broader generalization in the statistical sense. As addressed above, generalizability in case study research is better conceptualized as "user generalizability." Second, there are limits given that one investigator conducted all aspects of the study. My experience on team-based qualitative research efforts has demonstrated the value additional researchers bring to interpretation. The team approach allows for critical thinking and dialogue across team members that is largely absent here.

Third, given that the researcher did not pursue OMB approvals, she could only include up to eight non-federal participants in the study. OMB approval was not pursued given the lengthy review process involved (i.e., twelve to eighteen months or more following IRB approvals). The inability to include more representation from the grantee programs, in particular, is an important limitation to the study. The researcher attempted to address this limitation by including CDC program consultants who work closely with the grantees (two from each case), CDC field staff assigned to specific health departments (this was limited to the CSPS), and a small number of program directors working for state health departments. In addition, several program consultants who were interviewed for the study had, in the past, worked for state programs.

Next, the investigator conducted the research in her workplace. Although the researcher's role at CDC was perceived to facilitate access, her role as a colleague may have inhibited some respondents. However, as noted above, the researcher found her position afforded greater access to people, documents, and opportunities for observation, and also facilitated more open and frank dialogue from participants. In the end, the researcher's "insider" position was perceived as a valuable asset in conducting the research.

Fifth, a majority of those interviewed were involved in the development and implementation of the performance measurement system and may have conveyed a more positive impression than others. However, the study included participants (in each case) that were not directly involved and invested in the performance measurement effort (e.g., program consultants, grantee representatives).

Sixth, federal performance measurement is, to some extent, politically motivated and politically charged given policies such as GPRA and PART and the use of performance measurement data for decision making related to individual grantee budgets. The nature of the topic may have inhibited some respondents from speaking freely about their perceptions about it. In particular, there was political sensitivity in writing about some of the programs and their political contexts. After working at CDC for fifteen years, I am acutely aware of these sensitivities and had an inherent conflict of not wanting to "bite the hand that feeds me." However, I have attempted to honestly reflect the political context for these programs while also remaining sensitive to the political issues inherent to performance measurement. I found the member checking phase an important opportunity to assess these issues with the program staff and garner feedback and advice about their representation. Finally, limited resources in terms of time limited the data collection and analysis process.

# CHAPTER 4

# PUBLIC HEALTH EMERGENCY PREPAREDNESS (PHEP) PROGRAM

## 4.1 PHEP Case Description

As outlined in Chapter 3, Section 3.8.7, the case description is presented for this case and all others using a standard typology. The description includes a summary of the program, the implementation network, and the performance measurement system.

## 4.1.1 The Program

### 4.1.1.1 CDC Organizational Context

In 1999, CDC began funding state, local, and territorial health agencies to address terrorism-related public health emergencies. In 2002, following the September 11th and the 2001 anthrax attacks, Congress authorized funding for the PHEP cooperative agreement to support public health preparedness activities nationally, and funding for the program skyrocketed from about $45 million in FY 2001 to nearly $1 billion in FY 2002. During same time, the overall preparedness budget at CDC exploded from roughly $180 million to over $3.2 billion.

CDC's Division of State and Local Readiness (DSLR) in the Coordinating Office for Terrorism Preparedness and Emergency Response (COTPER) administers the PHEP cooperative agreement, providing consultation and technical assistance to grantees. The current cooperative agreement (PA #AA154) has been in place for eight years. In fiscal year 2008, CDC awarded over $700 million in funding through the PHEP program to 62

state, local, and territorial health agencies[28]. Awards are based, in large part, on population size and ranged from $330,743 to $50,161,370 with an average award of $11,368,829 (median award $8,897,688).

PHEP's organizational context can be distinguished, in part, by the fact that the PHEP program is situated in a "coordinating office" (i.e., COTPER) rather than a CDC center, reflecting a horizontal dimension of the program within CDC itself. As a coordinating office, COTPER provides funding to programs across the CDC centers, and DSLR staff draw on expertise (i.e., subject matter experts) from diverse areas such as infectious disease, influenza, environmental health, quarantine, and laboratories to support the PHEP program. For instance, scientists from other parts of CDC who specialize in infectious pathogens such as E. coli assist in the development of policies, provide technical assistance to grantees, and advise on the development of performance measures.

The organizational context for the PHEP program can also be characterized by constant change, high political visibility, and significant demands from federal government levels above DSLR. In particular, the political requirements imposed on the program from Congress, other federal agencies [e.g., Department of Homeland Security (DHS), Department of Health and Human Services (HHS)] as well as CDC's Office of the Director, are constant. PHEP staff are inundated by seemingly endless Congressional inquiries, new policy directives, and changing priorities (see political context, section 4.1.1.6 below).

---

28 The 62 grantees represent public health agencies for all 50 U.S. states, Chicago, New York City, Los Angeles County, Washington D.C., and eight U.S. territories.

Several of those interviewed used the term "wild, wild west" to describe the program culture of PHEP, reflecting an environment where unexpected demands from "above" were common and staff were challenged to feel a sense of control. Some expressed frustration over a situation in which efforts to implement a thoughtful, science-based approach were often derailed by demands from levels above them.

4.1.1.2 Program Goals

Since its inception, the PHEP program has morphed and expanded in its scope, often in response to critical events (e.g., 9/11, anthrax, Hurricane Katrina) or emerging threats (e.g., West Nile, pandemic flu). Today the program is tasked to address "all-hazards" which includes natural disasters as well as terrorist, chemical, biologic, radiologic, and nuclear emergencies. Addressing such a diverse range of hazards only begins to hint at the complexity of public health preparedness.

And unlike many other programs at CDC, the PHEP program's focus has been significantly shaped by a number of federal-level policy initiatives such as the DHS' National Response Plan, National Preparedness Guidelines, and Target Capabilities List; the Federal Emergency Management Agency's (FEMA) Homeland Security Exercise and Evaluation Program (HSEEP); the 2007 Homeland Security Presidential Directive (HSPD) 21; and the Pandemic and All-Hazards Preparedness Act of 2006 (PAHPA). In addition, staff are sensitive to the need to define the program in alignment with nine strategic preparedness goals that have been established by CDC.

One of the most significant challenges for DSLR staff is in defining public health preparedness, which can be described as sitting at the intersection of several disciplines including emergency management, defense, and medical services. Given that an "all-

hazards" program encompasses such diverse events, the dimensions of public health preparedness are expansive – surveillance, epidemiology, laboratory operations, response and recovery, and program implementation. At the same time, preparedness reflects many standard public health functions; consequently, differentiating it from CDC's typical public health work is also a challenge.

Acknowledging that the field is relatively new, DSLR staff have struggled to build an operational definition of what it means to be "prepared" by a limited science base and a lack of federal standards. Consensus on key aspects of what people viewed as a somewhat nebulous concept had not been achieved, although a concerted effort to do so is being led by DSLR's Outcome Monitoring and Evaluation Branch (OMEB) and a framework is emerging (see performance measurement system, section 4.1.3 below). Unlike most other public health programs, preparedness cannot be easily described using incidence and prevalence data or morbidity and mortality rates. Preparedness more accurately represents a *process* to build infrastructure, plans, and partnerships rather than an "end state" that could be eventually achieved and easily measured.

4.1.1.3 Stage of Program Development

As noted above, what is now called the PHEP program was initiated in 1999 and has changed substantially from year to year. The program is generally viewed as "new" and the discipline of public health preparedness is still evolving. In part, the program has been challenged to define itself given the lack of a solid science base and the inherent complexities reflected in the issue itself. The characterization of the program as the "wild, wild west" reflects participants' feelings that they were often "blazing new paths" as they helped to define the field of public health preparedness.

Overall, PHEP is a program still in development with more learning needed by CDC in collaboration with its grantees to better define the program's parameters. Accordingly, CDC expectations of grantees consider the program's developmental stage, recognizing the need for flexibility in regard to expectations. At the same time, however, PHEP faces significant accountability demands and policy requirements from HHS, many of which are viewed as unrealistic and insensitive to program realities (see political context, section 4.1.1.6 below).

4.1.1.4 Budget Stability

As noted earlier, the PHEP program experienced a phenomenal increase in funding in FY 2002, reaching $999 million. Effectively managing the influx of funding was challenging at all program levels, including CDC. At the state level, programs receive funds from multiple sources to support public health and medical preparedness (e.g., CDC, DHS, Assistant Secretary for Preparedness and Response) and are stretched to manage the reporting requirements for multiple federal agencies while also distributing funds to local levels needed to implement the program. Since its peak in 2002, annual funding allocations for the PHEP program have been consistently reduced, and staff recognize the importance of demonstrating accountability in order to defend current funding levels.

PAHPA, which was passed as law in 2006, requires that performance-based budgeting be instituted for the PHEP program in 2009 and mandates that funds be withheld for programs who "fail" to meet established targets or submit a pandemic flu plan meeting required criteria. Recently, HHS requested that CDC provide draft

160

performance measures that will be used to meet PAHPA requirements so that the measures could be published in the Federal Register for public comment.

DSLR staff expressed concerns about a shift to performance-based budgeting, especially given the stage of program development, lack of a strong programmatic science base, and on-going data quality concerns with the program-level performance measures currently in place. Many are afraid that the performance measures required by PAHPA will be used as "sticks" to reduce funds for "failing" programs. Philosophically, participants have concerns that reducing funds from a public health program may only worsen a grantee's capacity to protect the public. Others fear that PAHPA's approach to performance-based budgeting will damage relationships with grantees, lead to a "high stakes" testing environment where programs will simply "teach to the test," or result in gaming of the measures.

4.1.1.5 Stakeholders

A significant group of stakeholders at both the federal and state level exert influence on the PHEP program through their relationships with CDC. Within CDC, members of the Office of the Director represent a major stakeholder internal to the organization. PHEP is the single largest cooperative agreement administered by the agency, and, subsequently, the CDC Director works closely with the Director of COTPER.

Executive agencies at the federal level, including DHS, HHS, ASPR, and FEMA all represent important stakeholder groups for the program. In addition, Congress has a stake in the program given all 50 states are funded and that U.S. constituents across the country and U.S. territories are affected. Data suggest a top-down, hierarchical approach

to the PHEP program from those in Washington DC, with more extensive stakeholder involvement than typically experienced by CDC programs. For instance, program funding announcements for PHEP are closely reviewed by HHS. Some participants described a "cultural clash" between the federal agencies involved in preparedness and those in public health; in particular, they felt the preparedness field, in general, was more closely aligned with the military and its characteristic command and control approach. In contrast, public health typically assumes a collaborative stance with its partners and stakeholders to accomplish its goals.

The PHEP grantees represent another group of key stakeholders for the program. And national groups representing the grantees such as the National Association of City and County Health Officials (NACCHO), the Association of State and Territorial Health Officials (ASTHO), the Council of State and Territorial Epidemiologists (CSTE), and the Association of Public Health Laboratories (APHL) are all important groups with a stake in the PHEP program. ASTHO has an active subgroup – the Directors of Public Health Preparedness – comprised of state and territorial public health preparedness directors that advocates on behalf of its representatives. Both the individual grantee programs and the national organizations attempt to exert political influence on the PHEP program through regular consultation and interaction.

4.1.1.6 Political Context

The political context represents a dominant theme for the PHEP program. In fact, analytic codes related to the topic were the most frequently applied of any in the dataset. The sections above already hint at the political environment for the program. In part, the political context for PHEP reflects the public visibility and perceived importance of

preparedness in today's culture. But the program's political context may be impacted to an even greater extent by the substantial and complex intergovernmental arrangements at federal, state, and local levels that comprise the homeland security system in general. That system requires the integration of efforts from a myriad of federal agencies and offices including DHS, Department of Defense (DoD), FEMA, HHS, ASPR, CDC, Health Resources and Services Administration (HRSA), and others. Congress and the former President (George W. Bush) also instituted policy initiatives, legislation, and executive measures that directed or influenced preparedness efforts.

DSLR faces a difficult political climate that seems to regularly shift the program from one focus to another. As noted, political actions from those "above" CDC in the institutional hierarchy are frequent and limit CDC's control over the program's implementation. At the time of data collection for this study, DSLR staff were dealing with two initiatives that had direct effects on the program – PAHPA and HSPD-21. Under PAHPA (PL 109-417), the Assistant Secretary for Preparedness and Response (ASPR) was established within the Department of Health and Human Services (HHS). PAHPA directs CDC to require grantees to contribute non-federal matching funds beginning in FY 2008 and, as noted earlier, to institute a performance-based budgeting formula for the PHEP program beginning in FY 2009. PAHPA also requires that PHEP grantees develop pandemic flu plans that meet criteria established by HHS and requires "evidence-based benchmarks and objective standards" to measure levels of preparedness (to include outcome goals). The law dictates standardized funding cuts (i.e., 5% cut for first year of failed performance, 10% for second consecutive year of failed performance) for grantees that fail to meet established targets.

HSPD-21, drafted by the Homeland Security Council and issued by the White House in October 2007, establishes the National Strategy for Public Health and Medical Preparedness. The national strategy is built on key principles set forth in the *Biodefense for the 21st Century* (April 2004) and its intent is to articulate an approach to protect the health of all Americans against all disasters. Among other requirements, the presidential directive commands that the PHEP program assure state and local capacity to distribute countermeasures (e.g., drug treatment for anthrax exposure, prophylactic treatment for pandemic flu) within 48 hours after a decision to do so. Furthermore, HSPD-21 requires that standards and performance measures be developed within 270 days of the date of the directive "for state and local government countermeasure distribution systems, including demonstration of specific capabilities in tactical exercises in accordance with the National Exercise Program, and establish a process to gather performance data from state and local participants on a regular basis to assess readiness." The requirement related to the National Exercise Program reflects a DHS initiative called the Homeland Security Exercise and Evaluation Program (HSEEP) that sets out guidelines for planning and conducting security exercises at all levels of government. Finally, as part of HSPD-21, CDC is directed to begin collecting performance data within 180 days after the development of the measures and to use the data for determining future grant funding.

In summary, PHEP is situated in a complex political context with a strong top-down, compliance orientation. Some of the political directives have been difficult to translate into the public health realm given their orientation in emergency management or defense. And finally, COTPER faces frequent inquiries from Congress, the Inspector General, the Government Accountability Office (GAO), as well as on-going reporting

requirements for the Government Performance and Results Act (GPRA) and OMB's

Performance Assessment and Rating Tool (PART). In the end, the constant political

demands have been wearing for both DSLR and for PHEP program grantees. More

importantly, those demands have compromised a planned, science-based approach to

public health preparedness and a collaborative implementation process with state, local,

and territorial partners.

**4.1.2 Implementation Network**

4.1.2.1 Network Structure: Vertical Relationships

The vertical relationships that comprise the PHEP program begin at the federal

level with Congress, the President, and executive agencies including HHS, DHS, and

others. From a hierarchical perspective, three primary levels cascade below represented

by 1) CDC and DSLR, 2) the 62 state, city, and territorial health agencies, and 3)

thousands of local level jurisdictions across the U.S. These vertical relationships are

primarily intergovernmental and formalized by legal and fiscal arrangements: Congress

authorized funding for PHEP in 2002; CDC, in turn, receives an annual budget allocation

for the PHEP program from HHS; the 62 grantees, in turn, receive annual cooperative

agreement funding awards from CDC; and those grantees use contracts or other

mechanisms to distribute funds to regional or local levels. Over 50% of CDC funds are

distributed by grantees to more local levels.

Many participants used the catchphrase, "all preparedness is local," reflecting the

idea that public health responses will typically be initiated at the local level. A more

accurate description, however, is that "initial response is local" and that all levels are

needed in response to large scale events. For instance, an individual at the local level will

be the first to encounter a person with anthrax infection, but network members at other levels (state, federal) will be needed to implement activities (distribution of prophylaxis treatment) to reduce morbidity and mortality.

4.1.2.2 Network Structure: Horizontal Relationships

While vertical relationships ensure financial and other technical support from the federal government down to the local level, horizontal relationships are equally as important at all levels of government in preparing for and responding to public health emergencies. In nearly any scenario imaginable, multiple disciplines and sectors will be required to work collaboratively in order to effectively respond to an emergency for which public health may be only one relevant component. Using the anthrax example from above, participation of partners in the education, transportation, senior services, postal services, police and security, health care, and business may all be needed to manage the distribution of prophylactic treatment on a large scale. Consequently, PHEP has a significant horizontal dimension at every level in the vertical implementation chain.

The importance of collaboration with horizontal partners is evident in federal policies that affect many parts of the emergency management framework, including public health. HSPD-21 cites key principles derived from other federal strategy documents (e.g., *Biodefense for the 21st Century*, April 2004, *National Strategy to Combat Weapons of Mass Destruction*) which include "vertical and horizontal coordination across levels of government, jurisdictions, and disciplines" as well as "engagement of the private sector, academia, and other nongovernmental entities in

preparedness and response effort." In support of collaboration at the federal level, HSPD-21 required the establishment of the Public Health and Medical Preparedness Task Force to include broad representation from relevant agencies. The Task Force is chaired by the Secretary of HHS and include in its membership the Secretaries of State, Defense, Agriculture, Commerce, Labor, Transportation, Veteran's Association, and Homeland Security as well as the Director of OMB, Director of National Intelligence, and the U.S. Attorney General.

At CDC, COPTER works horizontally across the broader agency to coordinate efforts in support of preparedness activities. As noted earlier, PHEP draws on subject matter experts from many parts of CDC to make technical and scientific contributions. At the state-level, coordination is needed across government departments and with other sectors to carry out activities including planning and preparedness exercises. Such coordination may involve working with departments of education, transportation, emergency management, and environment, but also with hospital systems, commerce, and others. And at the local level, public health must collaborate with many of these same partners as well as the first responder agencies (e.g., fire, police).

### 4.1.2.3 Network Function: Authority and Control Within the Network

Control within the vertical network is exerted, in part, based on the institutional relationships – for instance, as a federal agency, CDC is directed by Congress, HHS, and others in the Executive Branch. And, as noted earlier, DHS, ASPR, and others also have influence over CDC's PHEP, primarily through the imposition of policy initiatives that dictate programmatic requirements. Cooperative agreements are used by CDC to fund the 62 PHEP grantees. This particular funding mechanism defines explicit responsibilities for

both the grantee and the CDC, allowing for substantial federal involvement in program implementation, including on-going programmatic monitoring and the provision of technical assistance. States have discretion to use the policy tools of their choosing in awarding CDC funds to local jurisdictions.

A primary challenge of the vertical structure is CDC's diminishing level of influence on actual program implementation with each step down the intergovernmental chain. While CDC has some formal authority over the grantee by virtue of the cooperative agreement, CDC's control over local level partners in the PHEP network is weak to nonexistent. CDC and grantees have even less control over the horizontal network dimension, but are nonetheless reliant on their participation to provide a seamless emergency response. Most of these horizontal relationships tend to be informal and unfunded. Developing effective relationships with these horizontal partners, therefore, becomes a critical part of public health managers' jobs and involves working across disciplinary cultures.

4.1.2.4 Network Function: Shared Organizational Goals and Priorities Within the Network

As noted above, preparedness demands the involvement of multiple disciplines and sectors working collaboratively across both vertical and horizontal networks. But differences or conflicts between agencies about priorities and goals can make collaboration difficult. For PHEP, there was more evidence of goal conflicts between vertical partners than across horizontal ones – probably because participants spoke more directly about those vertical relationships and few state-level representatives were included in the study. Given the top-down nature of the program itself, goals and

priorities of those upstream (e.g., HHS, ASPR, DHS) were privileged, but were often

viewed as "out of touch" with programmatic realities and lacking sensitivity to variability

in grantee context.

In addition, as in any vertically decentralized structure, implementation "drift"

occurs – that is, at each level (e.g., state, regional, local) the priorities and goals of a

particular grantee influence how activities are, in fact, carried out. In terms of the

horizontal dimension, participants spoke more about the challenges of earning a "place at

the table" and engaging partners than of specific conflicts across agencies in defining

common goals and objectives. Public health is a relative newcomer to the emergency

management arena and has struggled to define its role and earn recognition from the

others who have a longer history in the preparedness or emergency response arena. The

lack of available time on the part of horizontal partners to engage in planning efforts or

preparedness exercises led by public health is another barrier to their engagement. Even

at CDC, content experts in other parts of the agency have other priorities in terms of their

work responsibilities that conflict with their participation in the PHEP program efforts.

Similar challenges were faced at the state and local levels.

4.1.2.5 Network Function: Context, Resources, and Capacity

The 62 PHEP grantees vary considerably in terms of context, agency capacity,

and resources, all of which influence network functioning within each grantee's

jurisdiction. In terms of context, grantees differ in both geography and in their risk for

specific "hazards" (e.g., bioterrorist attack, natural disaster, radiation disaster) – just

consider the gulf of differences between New York State and South Dakota.

Consequently, the relevance of specific hazards influences the agencies that participate in

the network – a state prone to natural disasters may have a network comprised of uniquely different agency participants than a state seen at higher risk for a bioterrorist attack. In overseeing the PHEP program, CDC must remain sensitive to these contextual distinctions that require unique program models and program activities.

Resources and capacity also vary across grantees and, to some degree, influence grantees' ability to effectively work within the network structure. Resource levels are, in part, a consequence of the CDC funding award size, but also influenced by whether the individual grantee receives support from its own state or territory. Participants suggested that the current funding cuts in PHEP were negatively affecting grantees' capacity to carry out the requirements of the cooperative agreement, including collaborating with network partners.

**4.1.3 Performance Measurement System**

4.1.3.1 Process to Develop the Performance Measurement System

The PHEP program has had some form of a performance measurement system in place for several years. The measures have evolved extensively over that time; currently, six performance measures are in place for the national program and a dedicated effort is underway to refine and expand them. The current measures are summarized in table 14; the measures relate to three of CDC's nine Preparedness Goals that are also detailed in the same table.

**Table 14. PHEP Performance Measures (2008)**

| Performance Measure | Public Health Capability |
|---|---|
| **CDC Preparedness Goal 2: DETECTION AND REPORTING**<br>**Decrease the time needed to classify health events as terrorism or naturally occurring in partnership with other agencies.** | |
| **2A:** Percentage of Pulsed Field Gel Electrophoresis (PFGE) subtyping data results submitted to the PulseNet national database within 4 working days of receiving isolate at the PFGE laboratory.<br>Target: 4 days or less | Laboratory |
| **CDC Preparedness Goal 6: CONTROL**<br>**Decrease the time needed to provide countermeasures and health guidance to those affected by threats to the public's health.** | |
| **6A:** Percentage of key response partners that the public health agency successfully contacts without using electric grid power and primary land-line telephone service.<br>Target: Not established | Communication |
| **6B:** Time to notify all primary staff (secondary or tertiary staff as needed) with public health agency Incident Command System functional responsibilities that the public health agency's Emergency Operations Center is being activated.<br>Target: 60 minutes or less | Communication |
| **6C:** Time for primary staff (secondary or tertiary staff as needed) with public health agency Incident Command System functional responsibilities to report for duty at the public health agency's Emergency Operations Center.<br>Target**:** 2 ½ hours or less | Response |
| **CDC Preparedness Goal 9: IMPROVE**<br>**Decrease the time needed to implement recommendations from after-action reports following threats to the public's health.** | |
| **9A:** Time to complete a draft of an After-Action Report/Improvement Plan.<br>Target: 60 days or less | Program Implementation |
| **9B:** Time to re-evaluate response following approval and completion of corrective action(s) identified in an After-Action Report/Improvement Plan.<br>Target: Not established | Program Implementation |

Staff in DSLR's Outcome Monitoring and Evaluation Branch (OMEB) oversee development of the performance measurement system. OMEB is led by a PhD-level evaluator and comprised of staff with expertise in evaluation, epidemiology, data management, systems development, and public health. Most staff in the branch are relatively new to the Division (although not to CDC) having worked in OMEB for two years or less. The effort to develop performance measures is a high priority and a critical Division activity, and the branch receives strong support from management in both DSLR and COTPER's Office of the Director.

The history of the development of performance measures for PHEP reflects an evolution of measures over time, shaped largely by the expanding scope of the program and its political context. Table 15 below summarizes the various iterations of performance measures for the program by fiscal year. In general, previous performance measure sets were developed under difficult timelines and in response to political pressure or policy requirements.

**Table 15. PHEP Performance Measure System Development, 2003-2008**

| Fiscal Year | Number of Performance Measures |
|-------------|-------------------------------|
| 2003 | 120+ measures |
| 2004 | 47 measures |
| 2005 | 35 measures |
| 2006 | 23 measures |
| 2006 – 2008 | 6 measures |

The most recent development process was initiated in 2007 and is called the

PHEP Measurement Project. Led by staff in OMEB and supported by two government

contractors, the purpose of the project is to "develop, test, and implement measures of

public health capabilities for program accountability and improvement." A primary task

of the Measurement Project is to develop a conceptual framework for PHEP based on

priority program areas from which to identify performance measures and establish

program standards. OMEB estimates the development of revised and new measures will

be at least a two year process, although they recognize that performance measurement is

a dynamic, "self-correcting," and on-going process. The project is operating parallel to

other related, but separate, measurement efforts – the requirement to annually report

GPRA/PART measures, the new PAHPA requirement to institute performance-based

budgeting in 2009, and another OMEB initiative to develop a larger monitoring system

comprised of measures of capacity.

The involvement of stakeholders is viewed as critical to the success of the new

development process. In December 2007, OMEB invited representatives from ASTHO's

Directors of Public Health Preparedness to Atlanta to discuss the project, and in January

2008, OMEB assembled and convened a PHEP Evaluation Workgroup to provide

technical support and guidance to the measurement development process. The workgroup

includes 24 external stakeholders representing state and local health departments, tribal

communities, and national (e.g., APHL, ASTHO, CSTE, Centers for Public Health

Preparedness, NACCHO) and federal (e.g., DHS, ASPR, HHS' Assistant Secretary for

Planning and Evaluation) partners. During the January 2008 meeting, the Evaluation

Workgroup prioritized five PHEP capability areas for measurement development and

offered advice on a measurement approach. The capability areas were derived from DHS'

Target Capability List, and these five capabilities represent the foundation of the

conceptual framework that is intended to better define the PHEP program: (1) incident

management; (2) risk communications; (3) biosurveillance; (4) countermeasure delivery;

and (5) isolation and quarantine/community containment.

Once the five areas were defined, a nomination process was instituted to select

scientific and program experts to serve on topic-specific (capability areas) subgroups.

Subgroup members were not limited to the Evaluation Workgroup members. At the time

of data collection for this study, two of the five subgroups were formed and had convened

for two-day meetings to conduct process mapping exercises and begin revising existing

performance measures or developing new ones. Given the difficulty in defining

"preparedness" described earlier, the expertise of these subgroups was seen as critical to

the development of a conceptual framework and new or revised measures.

OMEB has applied a measurement framework focused on capacities, capabilities,

and performance to design their measurement system. PHEP defined these three

measurement categories as follows:

- Capacity Measures The acquisition, development, and maintenance of public

  health infrastructure and assets to support emergency preparedness and response.

  Measures of capacity include equipment, medications, supplies, staff, training,

  and preparedness plans.

- Capability Measures The ability to demonstrate public health emergency

  preparedness operations or actions. Measures of capability include documentation

of discussion-based and operational-based exercises and/or responses to real incidents. Capability = Capacity + Practice/Expertise

- Performance Measures Characteristics of public health emergency preparedness capability associated with the quality of operations or actions. Provides quantitative information on how well a process or outcome is performed. Measures of performance are often time-based but can also include measures of quality, completeness, and accuracy.

The intention of this latest measurement effort is to develop new performance measures by April 2010 along with related data collection guidance, training, and technical assistance materials. PHEP estimates that 6-8 new capability-based measures will be piloted and implemented annually. While the conceptual framework is perceived as a fundamental step in the development process, OMEB must also ensure that any new measures and measurement system align with various federal goals (e.g., CDC's nine preparedness goals, DHS target capabilities list) and are in compliance with federal reporting requirements (e.g., OMB, HHS, DHS, ASPR). In addition, OMEB must wrestle with other top-down political pressures, including the imposition of specific measures required by policies such as HSPD-21, which are often viewed as both nonscientific and unrealistic.

PHEP staff referred to the development of performance measures as a "scientific, social, and political process" in recognition of the complex array of factors affecting the effort. Collaboration with state and local partners, representatives of key national organizations (e.g., ASTHO, CSTE), and content experts was viewed as central to the performance measurement development process. But collaboration in the networked

environment of preparedness involves a multitude of stakeholders, including those

working along both vertical and horizontal dimensions and that represent funded and

voluntary partners. All these groups wield some level of political power and influence on

the development process. There have been previous missteps in working with grantees,

including poor communications, which have damaged trust and resulted in grantee

frustration. At the same time, everyone seems acutely aware of the need to demonstrate

accountability in a climate of declining funds, and data suggested a sincere commitment

from both CDC and grantees to work together in order to "get this right." CDC also

recognizes the experience and expertise grantees bring to the development process that

will help ensure that measures are feasible, valid, and relevant.

4.1.3.2 Performance Measurement System Design

In this section, the design of the performance measurement system is described

highlighting the following: purpose of system; level of measurement; types of measures;

use of targets or standards; and quality assurance efforts. Table 16 summarizes these

design features.

**Table 16. Design Features of PHEP Performance Measurement System**

| Design Feature | PHEP Performance Measurement System |
|---|---|
| Purpose of system | Accountability |
| Level of measurement | Grantee level |
| Type(s) of performance measures | Process measures that reflect program capabilities |
| Use of targets or standards | Yes – time-based targets for some measures |
| Quality assurance efforts | Yes |

4.1.3.2.1 Purpose

The primary purpose for the six existing performance measures is accountability. The PHEP program faces political pressure to demonstrate accountability "up" (e.g., to Congress, HHS, OMB) in order to justify and defend the resources allocated for the program. DSLR staff and grantees felt strongly that they *should* be held accountable, describing accountability in both fiscal and program performance terms. From a fiscal perspective, CDC's responsibility is as a steward of federal funds, assuring appropriate use of funds and budgetary practices. Programmatically, PHEP faces the difficult challenge of demonstrating accountability for "preparing" the nation, that is, to provide evidence that PHEP has improved our country's preparation for the varied disasters and attacks. Given the challenge in defining preparedness described earlier, this is no easy task. For now, the existing six performance measures are viewed as a means for CDC to satisfy demands for accountability and provide a "snapshot" of program performance to vertical network partners "above." These six measures are not thought to necessarily represent the most important priorities for the PHEP program nor are the measures perceived as particularly meaningful by grantees.

In contrast to the demands for "accountability up," CDC faces pressures from the grantees "below" for performance measures that reflect program priorities and that will support program improvement. There was a tension between the dual purposes of accountability and program improvement, with disparate views held about the ability of one set of measures to meet both purposes. PHEP staff view the two purposes as requiring very different approaches to performance measurement. From staff's perspective, satisfying accountability necessitates a small set of measures that can be

used to broadly reflect performance to those above CDC. Consequently, the six existing performance measures along with three PHEP-related GPRA measures are seen to address accountability.

For purposes of program monitoring and improvement, PHEP staff suggested a more comprehensive set of measures would be needed. While measures for program improvement would be correlated to the smaller set of accountability measures, the larger set of monitoring data would reflect a much deeper and broader view of program implementation that provides CDC and grantees more in-depth information to support system improvement. Consequently, staff believe that moving to a performance measurement system aimed at program improvement will require implementation of a more comprehensive *performance management* system inclusive of a larger set of monitoring data and the development of a *culture* around data use at all levels (e.g., CDC, grantee, local), neither of which currently exists.

Although there is an interest to have performance measures that also support program improvement, most recognize that the developmental stage of the program, including the absence of a strong program framework, means that such a system will not be in place in the short term. A more fully developed program model is needed to identify measures supporting program improvement, and time and resources are needed to build a multi-level performance management system and related staff capacity to support effective data use. In the meantime, PHEP is committed to the six existing measures that will continue to be used through 2008 for program accountability purposes, complimented by the three GPRA measures.

4.1.3.2.2 Level of Measurement

As it stands, the current set of measures reflect grantee-level performance rather than local level performance. The measures represent a set of programmatic capabilities viewed as important in public health preparedness and perceived as adequately responsive to calls for accountability. While PHEP recognizes the need to identify measures at the local level, the lack of a conceptual framework and concerns about inadequate capacity at the local level to collect and report data precludes the inclusion of this level of measurement.

4.2.3.2.3 Types of Measures

Because outcomes for PHEP had simply not yet been identified, the six performance measures are focused on process monitoring. Since multiple factors contribute to distal outcomes (mortality) and make it difficult to assess attribution, longer term outcomes will not be included in a revised set of measures. Instead, the program hopes to design future measures around proximal outcomes that more closely reflect an individual grantee's performance.

Given the rare occurrence of events within the preparedness arena, PHEP primarily relies on measures tied to preparedness "exercises" rather than real events. Five of the six measures are based on drills or exercises — CDC requires that exercises be conducted at least twice annually. Only measure 2A relates to laboratory testing in response to the receipt of actual isolates (e.g., E. coli, listeria).

4.1.3.2.4 Use of Targets or Standards

Measurement challenges led to the use of "time" as a proxy for quality in measures. For instance, measure 9A requires that an after-action report is completed within 60 days of an exercise or real event. Many of the time-based targets are developmental given the lack of science-based evidence supporting a standard and little trend data collected from programs. And although there was pressure from above to use time-based measures, many participants spoke to the limits of time as a proxy for quality. For instance, measure 6B relates to the time taken to notify key staff that the Emergency Operations Center is being activated. OMEB staff are not necessarily confident that "time" is the appropriate emphasis. For example, what may be more important, in this instance, is that the "right people" are available to address the problem.

4.1.3.2.5 Quality Assurance Efforts

PHEP has developed a guidance document for the six performance measures that details the following: the intent or rationale for each measure; the public health capability the measure reflects; the jurisdiction (e.g., state); the target for performance; specific definitions of terms; measurement specifications (e.g., numerator and denominator); data collection and submission methods (e.g., self-report semi-annually); and other comments or clarifications. Grantees are required as a condition of their funding award to report data for these six measures twice annually for distinct reporting cycles – three data submissions for the existing six measures have been received by CDC to date. Data are reported by grantees via the Preparedness Emergency Response System for Oversight

Reporting and Management Services (PERFORMS) – a web-based reporting effort that was originally designed as a grants management system in 2002.

OMEB has given significant attention to issues of data quality and validity. In part, the shear number of grantees along with variability in their capacity and resource levels, raises concerns about the quality and validity of the performance measurement data. PHEP contractors conducted a validation assessment in 2007 revealing a number of challenges to data quality including wide misunderstanding across grantees about the intent of the measures, uncertainty about how the measures were calculated and how measures should be reported, problems adhering to measurement protocols, misunderstanding of what data could be used to report on the measures, and others. The challenges to data quality and validity in such a decentralized system led PHEP to incorporate extensive guidance materials and technical assistance into the performance measurement system design. For instance, as part of the first two data submission cycles for the six performance measures, OMEB staff made individual phone calls with each grantee to review and verify that the data submitted were consistent with the requirements for each measure. This was a time-intensive exercise and OMEB does not have the capacity to continue the practice for every submission.

4.1.3.3 Use of the Performance Measurement System and Data

At this time, data from the six PHEP performance measures are used in a limited fashion given the recognition that the measures are still developmental and concerns about data validity. For the most part, data are used to address the demands for accountability and the frequent inquiries from CDC's federal partners, OMB, Congress, and others about whether or not "we are prepared." In February 2008, the first

Preparedness Report was issued by COTPER that incorporates a variety of data, including performance measurement data presented for each grantee. The report was released at a national preparedness conference during a plenary led by the CDC Director at that time, Dr. Julie Gerberding, and DSLR Director, Dr. Richard Besser. Within CDC, the compilation and release of the report was viewed as a significant accomplishment.

At this stage, PHEP is focused on the development of new measures and a data system to support their collection and analysis at CDC. There was little mention of specific efforts planned to build a broader system to ensure data use, although staff recognize a broader performance management system will be needed to facilitate data use in the future. To support data use, an effort will be needed to bring science and evaluation closer to program. However, as it stands, staff in other branches of DSLR have limited involvement and understanding of the performance measurement efforts. Most participants suggested it would be beneficial to have greater participation from program consultant staff in the Program Services Branch (PSB) – these staff oversee the PHEP cooperative agreements and have close working relationships with staff in the states, cities, and territories. Program consultants could be important change agents in the adoption and use of performance measurement. However, participants also acknowledged that the "brutal" pace and demands of the PHEP program limited program consultants' time to contribute to the effort.

## 4.2 Findings

### 4.2.1 Dependency on the PHEP program's network partners diminishes CDC and grantee control over performance.

As described in section 4.1.2, public health preparedness demands extensive collaboration with network partners at both vertical and horizontal levels. Ensuring an effective response to a natural disaster, bioterrorist attack, or nuclear event requires an integrated and seamless reaction across levels of government and across multiple sectors. The PHEP program's public health preparedness activities, such as developing plans for various scenarios (e.g., pandemic flu) and conducting exercises (e.g., bioterrorist attack), are dependent on the participation of others in state departments of health, education, and transportation as well as those in local departments of health, schools, first responder agencies, health care, and commerce among others.

The significance of intergovernmental relationships in a preparedness response presents a major challenge for CDC, given DSLR's limited control over the state and, especially, the local level, although that is where a significant proportion of the cooperative agreement funds end up. While the cooperative agreement provides CDC some authority over their grantees (e.g., defining allowable activities, including a requirement to report on performance measures), grantees' control over the local level may be limited by their state's model of decentralization and/or their choice of funding tool, both of which vary considerably across the country.

And while CDC and grantees may be able to exercise some authority through policy tools like cooperative agreements and contracts, formal command and control structures are virtually nonexistent across the horizontal domain, especially given the

informal nature of these relationships. So while CDC and its grantees are dependent upon

these partners to participate in public health preparedness efforts for which they lead and

for which they are accountable, both lack mechanisms of formal control to ensure their

participation or performance. One participant, speaking about the requirement for states

to develop a pandemic flu plan, said

> The [state] Health Department has the responsibility for coordinating the
> *statewide plan*, so the Health Department has to work with the Department
> of Transportation, the Department of Education, the Department of State,
> the Department of Highways, whatever they have in their state. And each
> of them [these partners] has to contribute to this overall plan. Well, you
> know, if they don't want to, or they don't get it, or they don't understand
> how, then it becomes the Health Department's job also to orient them, and
> educate them, and train them, and nag them to get the documents.

Participants described the challenges associated with previous performance measures that

required data from sources over which grantees had no formal authority or control.

> A number of the measures that we have collected historically were not
> really under the control of the grantee, which compromised the data. It
> was very difficult for them to get the data because it wasn't really under
> their purview. We would ask them, let's say, 'how fast are you in doing
> this lab test?' Well, not a lot of the states did those tests – they would have
> triaged those tests out to other labs. So their structure did not support the
> collection of that data or they were reporting on something that they were
> not responsible for.

Grantees contend that performance for some of the existing six measures is

dependent on the cooperation of informal, unfunded, horizontal partners. For instance,

grantees' ability to meet the target on measure 2A is dependent, in part, on receiving

viable samples from local laboratories that can, in fact, be tested. The team in OMEB

recognizes the difficulties this issue of compromised control poses for the development of

performance measures As part of the process to develop a revised set of performance

measures for the PHEP program, the Evaluation Workgroup, a team comprised primarily

of external stakeholders, established criteria for measure selection. One criterion is that

selected measures be "under the control" of public health – a recognition of the challenge

grantees face in affecting performance on measures that may be largely outside their

direct control. Grantees have pressured CDC to account for this dilemma, casting it as an

issue of "fairness." As example, in developing a future measure related to the PAHPA

requirement that all grantees develop comprehensive pandemic influenza plans, one

participant representing a grantee said,

> Like for Pan flu –a big piece of it is the school closure issue, and so we've
> brought Department of Education to the table. And I think it's been going
> okay, but I mean obviously it doesn't go as fast as if it's requiring the
> involvement of people who aren't within your purview. But I guess I
> would strongly argue for just holding us [grantees] responsible for health
> and medical indicators. You know, not that it's not important how other
> agencies do, but there has been a trend to hold us responsible for work that
> is being done by other agencies [that we don't control], and especially
> with this movement to tie funding to performance—it's really not fair.

**4.2.2 Network public management fragments the PHEP program's accountability
for results, creating challenges for performance measurement.**

As reflected in the case description, the PHEP program is under extraordinary

pressure to demonstrate accountability for results. The size of the program budget (over

$700 million annually), the political climate, and the importance of the issue all

contribute to demands on DSLR staff to demonstrate that "we're prepared."  However,

the challenge of defining preparedness complicates the identification of outcomes for

which grantees can be held accountable as reflected in this statement:  "Since there are no

shared-upon standards for things [preparedness], it's difficult for me to say, 'they

[grantees] have been accountable for what they're doing', you know, and really mean

something." In addition, as discussed above, important outcomes for PHEP, and

preparedness more generally, will likely result from the work of PHEP in coordination with many other network partners. So while it is understood that an effective response is contingent upon collaboration across sectors and levels of government, both the vertical and horizontal relationships muddy efforts aimed at assigning accountability.

This fragmentation of accountability has consequences for the design of PHEP's performance measurement system. For instance, measurement becomes more complex – one participant described the challenges presented by the inclusion of performance measures that capture collaborative performance:

> It's looking at those junctures between the systems [public health, transportation, emergency management] as indicators for performance. And that's where it gets tricky because those junctures are a shared goal [shared goal across agencies]…Let's just take it down to the local level— the emergency management, the fire, the police, the health department. How do we work together collectively? Who's responsible for that measurement? I mean, I'm going to boil this down to just the data itself, who's going to collect it?

This statement illustrates one problem for PHEP's performance measurement system generated by the joint production of outcomes inherent in the field of preparedness. That is, beyond working collaboratively to achieve outcomes, network partners must contend with pragmatic issues related to the complexities and costs of measurement. In the networked context, it is unclear who has responsibility for developing and managing the performance measurement system, including the data management systems, data collection, reporting, and analysis.

As already mentioned, the issue of control described in 4.2.1 inevitably converges with demands for accountability in the PHEP network. As you move along the results chain, accountability for outcomes further "out" will be difficult to assign. And as the

achievement of important performance outcomes becomes the providence of multiple

organizations, issues of fairness again arise when a single organization (the grantee) is

held accountable for them. Indications from this case study suggest that grantees will

continue to challenge proposed performance measures for which they will be held

accountable but that they cannot influence directly. Consequently, OMEB staff will be

pressured to select measures for which accountability is more straightforward (and for

which grantees can control performance). More often, such measures will be process and

output indicators that are more closely aligned with the work of the grantee.

**4.2.3 In the case of the PHEP program, performance measurement is a "political, social, and scientific" process.**

OMEB leadership referred to performance measurement as a "political, social,

and scientific" process – for the PHEP program, these three contextual factors interact to

shape the development effort and ultimate performance measurement design. The social

aspect reflects the need to continually interact with stakeholders at multiple levels in

building a performance measurement system. In the PHEP program's network context,

this translates into engaging an extensive group comprised of CDC leadership, the 62

grantees, federal partners (e.g., APHL, ASPR, DoD), and policy makers within both the

vertical and horizontal dimensions. One person commented,

> The socialization of the measures, it's really working with our partners,
> socializing them up [to federal levels above CDC], socializing them out
> [to horizontal partners], and socializing them down [to grantees], you
> know. It's about change management, buy-in, acceptability…It's the
> softer side of the work we do, but it's the most important…It's looking to
> the people who actually do the work, to say, 'does this make sense?' We
> want these measures to be relevant, we want them to be feasible, we want
> them to be reliable, and we want them to be valid. We can definitely look
> at the science-based side of things, the literature, as a guiding principle,

that's the other bucket. But that only takes you so far because you really
need to know, in the real world, how does this play out?

At the same time, OMEB staff face political demands and influence from these
network partners who shape the purpose and selection of performance measures. As
reflected in section 4.1, the PHEP program resides in a dynamic political context. The
program has swung from one focus to another as it has expanded in scope in response to
critical events including 9/11, the anthrax attacks, and hurricane Katrina. One person
described the increased attention given pandemic influenza,

> The same thing happened two years ago with pan flu. All of a sudden it
> was a 'presidential emergency.' So now we've infused another $600
> million into the [PHEP] cooperative agreement to say, 'while we are on an
> all-hazards approach, trajectory, we really need you [grantees] to
> concentrate on pan flu' which has redirected all of these resources to doing
> something that is against what we said we wanted to do.

This shifting political landscape has made it difficult for OMEB staff to build a
conceptual framework from which to identify potential performance measures – five
different sets of measures have been produced over the past six years. Most recently, an
evaluation workgroup comprised of external stakeholders identified five capability areas
and assigned individual subgroups of content experts to develop a new program
framework that can encompass priority areas as diverse as risk communications,
biosurveillance, and isolation and quarantine.

Participants spoke extensively about the implications of policy initiatives from
"above" including PAHPA and HSPD-21, federal level guidelines such as DHS' Target
Capabilities List and the HSEEP, and federal guidelines such as the National
Preparedness Guidelines for the PHEP program's performance measurement system.

Some of these policies impose specific measures, others set priorities and timelines – and most are viewed by participants as both unrealistic and reflecting little understanding of public health. One person said,

> The challenge is around the political agendas and the fact that Washington, ASPR, HHS, are completely unreasonable in what they expect from a measurement perspective. If you look at HSPD-21 that was just released in December [2007]. It gave us six months to develop measures for countermeasure distribution. The data, the first report has to be [submitted] within a year. I mean, it's just totally unreasonable.

Consequently, OMEB struggles to manage the incongruity between top-down, political imperatives and their feasibility for implementation in a network of 62 grantees representing varied contexts in regard to risk of hazard (e.g., hurricane, bioterrorist attack), geography (e.g., rural, urban), resources, and capacity.

One way OMEB staff manage the political demands is to apprise network partners in Washington D.C. of their work in regard to the performance measures as reflected in the following statement.

> The political element is How do we keep the powers that be satisfied with the information that we can provide them? And there is a social component to that—it means going up to Washington quite frequently, talking with them about the measurement strategy that we're proposing, recognizing we have these two tracks. 'You're going to get kind of the snapshot, quick and dirty data [six existing measures] and then we're also working this track to develop these performance-based measures [new measures under development]. I alluded to the idea that quality is harder to capture. That's going to take us some time to get there.

Finally, OMEB staff described performance measurement as a scientific process. For now, OMEB staff are faced with developing a performance measurement system for a relatively new and complex field with a minimal evidence-base. The complexity of

preparedness and the current vacuum in regard to science challenge OMEB staff to know *what* to measure. One person said,

> You know, for preparedness it's hard because it's not concrete, you're not delivering services to patients. It's not like a healthcare facility where you're measuring outcome in terms of morbidity or mortality -it's much more complex than that. I think that's why everyone's grappling with it -- because what does it actually mean to be prepared? There's no actual end point.

Some suggested that the lack of a strong science base for the preparedness program left DSLR more vulnerable to political influence. In speaking to the development of performance measures, one participant said, "I mean, it [the lack of a science base] compromises the ability to do it [develop performance measures] plus it leaves us wide open for political sway which is typically what's happened." OMEB staff are left to balance a science-based approach with political imperatives that seem indifferent to both the scientific complexities and intergovernmental implementation network within which CDC must operate.

In summary, performance measurement for the PHEP program reflects the intersection of political, social, and scientific processes. In developing performance measures for the PHEP program, OMEB staff must contend with a significant political context, including policy that directly affects the selection of performance measures and imposes demands on the development process. At the same time, staff must grapple with a complex program and a limited science-base while facilitating an inclusive and collaborative process with a large and varied network of partners, all of whom are important to ensuring a set of meaningful performance measures that can serve a program comprised of 62 diverse programs.

**4.2.4 The variability across the PHEP program's vertical network significantly shapes the design of its performance measurement system.**

Given that 62 grantees and thousands of local health agencies comprise the vertical network of the PHEP program, extensive variability exists in the risk for and potential scale of any given event; program priorities and activities; capacity and resources; availability of data sources and sophistication of data collection systems; and program context. This variability within the network introduces significant challenges for performance measurement and requires CDC to develop a system that is sensitive and responsive to the diversity among grantees. In many respects, the network demands that the development process be a negotiated process. One participant, representing a grantee, acknowledged the difficulty in defining a common set of performance measures appropriate for such a diverse set of grantees.

> We just have such dramatically different populations. We function very different as agencies. I just think there's great variation in a lot of our challenges, in a lot of our structures, so it just doesn't always work to use the same measure for everybody [all grantees].

Another design challenge imposed by the network structure involves developing measures that can be operationalized at the local level. Participants emphasized that preparedness starts at the local level. One person said,

> What's interesting to me is that the whole point [of the PHEP program] is to get that local response prepared. I mean, look at the county health department if somebody walks in with a potentially flu-like illness [pandemic flu], can they put everything in place, from that local health department to adjacent ones, all the way up, *literally* all the way up [to the Federal level], to make sure that the effect of an outbreak is minimized?

The extensive decentralization and subsequent variability across local level partners has proved a barrier to the identification of local-level performance measures. Consequently,

the six measures currently in place are state rather than local-level indicators. One

participant described the challenge as follows,

> I'll bet you've heard that all emergencies are local and so all response is
> local. If something happens in a county, it's not the state that comes in and
> responds, the county immediately responds…And so something about our
> measures needs to reflect how ready the locals are. And in any one state
> there may be 40, 60, 120 different local units. So trying to figure out what
> things they should all be able to do and how to measure those things and
> how to get them all to measure it in a way that rolls up [to the state level]
> is huge. That might be the biggest challenge for [performance
> measurement] in preparedness.

Concerns about capacity at both the state and local level, particularly related to

data collection and reporting, influences the choice of performance measures. The

implementation of performance measurement for the PHEP program requires that all 62

grantees have access to needed data, effective data collection and reporting systems, and

the capacity for data analysis. One participant said,

> There were other measures that we wanted to include but that, frankly,
> once we looked at how they [grantees and locals] would actually report the
> measures and the systems out there and the assumptions that would have
> to be in place…we couldn't come up with a way for the grantees to report
> them to us—that was a real eye opening experience for me.

OMEB has already experienced serious data quality problems – grantees have struggled

to understand the measures and to collect and report data. Consequently, OMEB has

conducted extensive quality assurance checks for data submissions and will have to build

in quality assurance mechanisms as they continue to develop the system. The variability

in capacity, resources, and context also limits CDC's ability to compare performance

across the PHEP program grantees. Most participants thought that such comparisons

192

would be looking at "apples and oranges" and saw little benefit in making such comparisons.

And finally, given the obstacles imposed by network variability along with other challenges related to program complexity and political pressure, the current set of six performance measures were often referred to by participants as "low hanging fruit" – common measures that all grantees could feasibly collect and report and are reasonably responsive to accountability demands. Despite the challenges posed by the network, staff in OMEB continue their efforts to collaborate with their partners to develop a conceptual framework for the PHEP program that will lead to more meaningful and useful performance measures for CDC and its grantees.

### 4.3 Summary

In summary, program results for the PHEP program ultimately depend on the coordinated and seamless response of network partners representing diverse sectors and disciplines. Without full control over the production of outcomes, grantees argue that it is unfair to hold them fully accountable for their achievement. In the case of the PHEP program, establishing a criterion for performance measurement selection related to "control" is one way to address dependencies on vertical and, particularly, horizontal network partners. But the decision to apply this criterion comes at the cost of limiting the choice of performance measures for the PHEP program. Specifically, important outcomes related to preparedness but reliant on the performance of the larger network are excluded. The choice may result in the inclusion of more measures reflecting outputs and shorter term outcomes directly tied to the work of the grantees.

The PHEP case also highlights performance measurement as an intersection of political, social, and scientific processes. Network stakeholders exert significant political influence throughout the development process, and OMEB staff must also contend with policy initiatives imposed from other federal agencies that directly affects the selection of performance measures and prescribes demands on the development process. Given the broad implementation network, OMEB must also negotiate an inclusive and collaborative (social) process to ensure the identification of measures that are meaningful and valid to 62 diverse grantees and their local level partners. The variability across these grantees represents a significant challenge to measurement development – in particular, to develop a common set of national measures reflecting local-level preparedness efforts that are sensitive to differentials in risk and scope of potential hazards, capacity, and resources.

And finally, staff must grapple with an extremely complex program which, at this time, is supported by a limited science-base. OMEB's most recent effort to engage experts from across the country to participate in constructing a programmatic framework around five central program capability areas (e.g., risk communications, biosurveillance) reflects a foundational step in furthering the development of their performance measurement system.

# CHAPTER 5

## COMPREHENSIVE STD PREVENTION SYSTEMS (CSPS)

### 5.1 CSPS Case Description

#### 5.1.1 The Program

##### 5.1.1.1 CDC Organizational Context

Alongside malaria and polio, CDC's STD program is one of its oldest, dating back to 1957. Today, the program resides within CDC's National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention (NCHHSTP), and is managed by the Division of STD Prevention (DSTDP). The mission of DSTDP is to provide national leadership, research, policy development, and scientific information to help people live safer, healthier lives by the prevention of STDs and their complications.

DSTDP has recently developed a five year strategic plan that identifies seven goals related to the prevention of STD-related infertility and HIV prevention, building STD prevention capacity and infrastructure, reducing STD-related health disparities, and addressing the costs of STDs and their associated sequelae. Specific objectives contained in the plan were developed in alignment with several other planning and priority setting efforts including the overall CDC goals, CDC's Coordinating Center for Infectious Diseases (CCID) performance goals, and the NCHHSTP's program imperatives.

The largest program within DSTDP is its CSPS program which funds 66 grantees – public health agencies in all 50 U.S. states, 7 cities (Baltimore, Chicago, District of Columbia, Los Angeles, New York City, Philadelphia, San Francisco), and 9 U.S. territories. The CSPS was funded through a grant up until the end of 2008. The program

now operates under a cooperative agreement which also includes funding components for the prevention of STD-related infertility (i.e., the Infertility Prevention Project or IPP) and syphilis elimination in high morbidity areas (i.e., the Syphilis Elimination Program or SE). At the time interviews were conducted for this study, the program was nearing the end of a five-year grant funding cycle initiated in 2004 and preparing to issue a new five-year beginning in January 2009.

In fiscal year 2008, DSTDP provided $104,263,809 in total funding for all three program components (CSPS, IPP, SE). Heretofore the overall program will simply be referred to as CSPS. Individual grant awards ranged from $43,609 – 6,711,083 with an average award of $1,585,575 (median award $1,137,423). CSPS is managed by the Program and Training Branch within DSTDP. A unique feature of the CSPS grant is its inclusion of "direct assistance," that is, the assignment of federal employees to some grantees to provide on-site managerial and technical support. While the placement of CDC "public health advisors" has been dramatically reduced over the past 15 years, CDC still supports over 175 federal positions assigned to state and city grantee programs.

The organizational context for CSPS can be characterized, in part, by the program's maturity and long-term tenure of its staff. Many of those interviewed for this study have worked in the STD program for over 15, 20, or 30 years. Staff in the program often initiate their careers as federal public health advisors (field assignees) conducting the fundamental "gum-shoe" detective work of STD case finding – interviewing patients diagnosed with STDs to identify others who were potentially infected and then contacting those persons to ensure appropriate testing or prophylactic treatment. Over time, these staff are often promoted into state-based management positions and eventually promoted

to CDC headquarters in Atlanta to serve as program consultants providing technical assistance to grantees or as managers supporting CSPS and other program efforts. Consequently, many staff, particularly those in the Program and Training Branch, bring extensive front-line experience to their positions.

5.1.1.2 Program Goals

The purpose of the CSPS grant is to support state and local STD prevention programs in designing, implementing, and evaluating their programs. The complexity of the program is reflected, in part, by the existence of over 25 STDs. DSTDP is responsible for all STDs other than HIV/AIDS which falls under the purview of the Division of HIV/AIDS Prevention. DSTDP places particular emphasis on preventing and controlling syphilis, gonorrhea, and chlamydia. The prevalence of these three STDs varies across the U.S. and its territories, and, given their infectious nature, related STD epidemics can shift over time both geographically and among sub-populations. For instance, the prevalence of syphilis was greatest among southern, heterosexual rural African Americans in the late 1990s, but a more recent epidemic has emerged among men who have sex with men, often those co-infected with HIV.

The activities supported by CDC's CSPS grant include providing community and individual behavior change interventions; providing medical and laboratory services; ensuring partner services; promoting leadership and program management; conducting surveillance and data management; providing or ensuring training and professional development, and ensuring a documented STD outbreak response plan. Grantees receiving funds for the IPP project component are responsible for ensuring the testing and treatment of young, sexually active women and their sex partners for chlamydia and

gonorrhea; supporting laboratory testing; developing surveillance systems; and ensuring provider training. As a requirement of the award, grantees receiving IPP funds must direct 50% of the award to Title X family planning programs for chlamydia and gonorrhea screening of women and their partners[29]. Grantees receiving SE funds are responsible for enhancing surveillance efforts, strengthening community partnerships; providing rapid outbreak responses; expanding clinical and laboratory services; and enhancing health promotion. Grantees eligible for SE funds are those with high syphilis morbidity areas as defined by CDC. Similar to the IPP funds, grantees receiving SE funds must award 15 percent of the funds to community based organizations (CBOs) that serve affected populations.

Overall, the vast majority of CSPS funds support non-clinical prevention services such as education, outreach testing, field investigation efforts, and surveillance activities. These activities are largely carried out at the local level with grantees typically providing a significant portion of their CDC funds to county health departments, but also Title X clinics (IPP funds) and CBOs (SE funds). CDC grantees and their sub-awardees (e.g., county-level departments of health) are encouraged to collaborate with public and private health care providers (including family planning clinics), community-based organizations, and others who serve persons at risk for STDs. Jails and juvenile detention facilities are also identified by CDC as important agencies with whom to partner, given a higher prevalence of STDs among incarcerated populations than the general public.

---

[29] Title X was funded in 1970 as part of the Public Health Service Act. Title X funds are administered by the Department of Health and Human Service's (HHS) Office of Family Planning in the Office of Population Affairs and support low cost, confidential family planning services for young and low income women.

Participants described a relationship with grantees that has changed over time. For many years – in the 50's, 60s, and 70s – the program was tightly managed by CDC. The placement of hundreds of public health advisors in grantee programs allowed CDC to exert more direct control and forcefully manage the STD program. During this time, extensive data were collected and reported to CDC about case surveillance and program activities (e.g., number of persons suspected to be infected with STDs that were interviewed within a certain number of days). But by the end of the 1980s, data requirements lessened and CDC required little other than surveillance case reports. Overall, grantees were given much greater latitude. Some participants suggested that, as a result, CDC has allowed each grantee to map their own course rather than provide the leadership to guide the national program in a clear direction.

One program strategy promoted by CDC over the past several years, however, is partner relationships. While STD programs have long collaborated with local partners, programs are being encouraged to engage public and private agencies, private health providers, health maintenance organizations, and others in order to address STDs from a broader, state-wide (or territorial-wide, city-wide) perspective. In some respects, the approach reflects a philosophical shift – rather than simply focusing on the clinics that the grantees fund, they are being asked to develop and leverage partner relationships in order to have broader state-wide effects.

Several factors contribute to the emphasis on what has been termed a "community approach." First, CDC's budget allocation has remained relatively stagnant for several years, and staff recognize the need to leverage partner resources in order to effectively address program goals. Second, patterns of health care access and delivery have changed.

Today, some patient populations are more likely to seek STD treatment services from private, primary care providers rather than public STD clinics. And third, there is the recognition that if CDC aims to have a population effect, they need other partners engaged in the effort. Consequently, grantees must involve and influence others who either interact with persons at high risk or provide health care services to affected populations in order to reduce and control STD infections in the broader community. As noted above, moving to an approach centered on community engagement and partnering does, to some extent, reflect a paradigm shift for grantees. And the changes require that programs recast their perspective to assume broader responsibility and accountability for STD prevention and control in their communities or state rather than for only in the publicly funded STD clinics.

One aspect of the community approach is to engage unfunded, informal partners (e.g., private medical providers, health maintenance organizations, corrections facilities) in the STD prevention and control effort along with those who receive program funding (e.g., community based organizations who receive SE dollars). There is an expectation that even without funding, STD programs should be able to leverage their status as "STD experts" in order to influence others to help address STD prevention and control.

### 5.1.1.3 Stage of Program Development

As noted above, the STD program is one of CDC's longest standing public health programs. But while the program is mature, changes in health care delivery and access, shifts in populations affected by STDs, differences in disease burden across grantees, and the stagnant funding context require that DSTDP and the CSPS program remain flexible in their ability to respond to fluctuations in the public health environment. Participants

suggested this was not necessarily easy. For while the longevity of the program and the experience of its staff reflects a well-instituted effort, its history also contributes to a culture where change and the adoption of new practices can be difficult. In addition, programs have experienced significant autonomy and relatively little oversight over the past 15 years.

### 5.1.1.4 Budget Stability

In general, the budget for the CSPS program has stayed relatively steady over the past several years. And given annual inflation and small salary increases for staff, flat funding results in fewer service delivery dollars. While some grantees receive support from their state or local coffers, many others rely solely on federal CDC dollars to fund their programs. Staff suggested that STDs are not particularly high profile or a big priority for federal support, especially in comparison to HIV/AIDS which receives greater attention and significantly more funding. As noted above, funding challenges have led, in part, to DSTDP encouraging grantees to leverage partner resources and assess how best to spend limited program dollars. In particular, rather than simply funding the same local agencies year after year, grantees are encouraged to assess where resources are most needed and move dollars if indicated.

### 5.1.1.5 Stakeholders

Data suggest that the grantees and the National Coalition of STD Directors (NCSD) represent two key stakeholder groups for the CSPS program. The NCSD is a national organization comprised of state, local, and territorial STD directors established to promote national awareness of policies that govern STDs. CDC consults regularly with

NCSD on a range of program concerns, and NCSD representatives are often invited to participate on DSTDP workgroups convened to address policy and other issues.

5.1.1.6 Political Context

At the federal level, participants viewed CSPS as having a relatively stable political environment. Among sexually transmitted diseases, HIV/AIDS gets the lion's share of political attention and federal resources. Individual grantees and NCSD are viewed as having a degree of political influence and power that can be exercised in their relationships with CDC. And at the grantee level, there are politics that come into play between the grantee and its subcontractors. For instance, grantees may face political opposition to move funds from one long-standing subcontractor to another.

**5.1.2 The Implementation Network**

5.1.2.1 Network Structure: Vertical Relationships

The vertical relationships involved in the CSPS program begin at the federal level with a budget appropriation from Congress and the President to HHS. Funds then come to CDC and, as described earlier, are distributed to 66 state, city, and territorial grantees through cooperative agreements. Given that CSPS program services are primarily delivered at the local level, grantees then use contracts, grants, or other funding mechanisms to support hundreds of local level implementers in the vertical chain, most of which are local public health agencies. The vertical structure within a state, city, and territory also varies. For instance, the California state program funds approximately 60 different local health agencies. Some states may fund at a county level, others at a city or

regional level. In frontier states, services may be provided directly by the state with a small staff addressing multiple public health areas (e.g., HIV/AIDS, STDs, tuberculosis).

As described in section 5.2.1.2 above, at least 50% of IPP project funds must be awarded to Title X family planning settings to support chlamydia and gonorrhea screening and the treatment of young, sexually active women and their sex partners. Recipients of Title X funds include both private and public agencies (e.g., nonprofit family planning clinics, hospitals, public health departments, university health centers, Planned Parenthood affiliates). Grantees must also award 15% of SE funds to CBOs that serve affected populations. Consequently, the vertical chain involves a network of funding relationships, primarily through intergovernmental channels (e.g., federal, state, local health agencies), but also with private and non-profit groups (e.g., non-profit Title X clinics, CBOs).

Most obvious from the program's vertical decentralization is that the CSPS program achieves its goals through service delivery that occurs several steps removed from CDC. As mentioned earlier, up until the late 1980s, the STD program was managed from a more strident command and control structure, aided by the extensive placement of field staff (public health advisors). Since that time, however, CDC has imposed less direct oversight over the grantees, and the grantees have grown accustomed to a greater level of autonomy in managing their programs.

More recently, CDC has attempted to restore some balance between providing grantees a degree of latitude that allows them to tailor programs appropriately to the disease burden, needs, and contexts of their communities while also imposing some expectations and means of accountability to ensure a strong national effort, especially in

regard to data collection and reporting. In theory, everyone is aiming for the same broad

goals – but the operationalization of a national program through multiple levels (i.e.,

national, state, local) introduces challenges related to the fidelity of implementation and

accountability.

5.1.2.2 Network Structure: Horizontal Relationships

Collaborative, horizontal relationships exist at the federal, state, and local level.

These relationships are primarily informal, voluntary, and intended to support either

program integration or improved access to populations at risk for STD infection. At the

federal level, DSTDP interacts with the other divisions in NCHHSTP (i.e., the Divisions

of HIV/AIDS, Viral Hepatitis, Tuberculosis) as well as with other Centers (i.e., Division

of Cancer Prevention and Control, National Center for Chronic Disease Prevention and

Health Promotion) largely to provide more integrated service delivery. For instance, at

the time interviews were conducted for this study, DSTDP was collaborating with the

Division of HIV/AIDS to develop programmatic guidelines for partner counseling. This

type of collaboration, aimed to provide more comprehensive and integrated services to

the public, is viewed as essential to achieving DSTDP's mission. As noted earlier,

DSTDP also collaborates regularly with the NCSD. While DSTDP works with other

federal partners such as the Health Resources and Services Administration (HRSA) and

the SAMHSA – both a part of HHS – collaboration could be improved.

At the state level, similar horizontal partnerships exist, especially in grantee

agencies that have stand-alone programs for STD, HIV/AIDS, and TB. Grantees with

smaller departments of health may have a more integrated program for these disease

areas. The 2009 funding announcement for grantees explicitly encouraged STD programs

to "engage in meaningful collaborations with appropriate public and private health care and other non-medical partners, particularly organizations serving at-risk populations." State-to-state collaboration is also occurring as state representatives convene, often on a regional basis (e.g., western states, northeastern states), to review surveillance data, discuss program activities, and provide peer-based technical assistance.

The "community perspective" described earlier is largely actualized at the local service delivery level where horizontal relationships between STD programs and private providers, laboratories, adult corrections, juvenile detention, family planning clinics, and CBOs are recognized as central to achieving program goals. Some of these relationships are formalized through funding arrangements while others are more informal.

## 5.1.2.3 Network Function: Authority and Control within the Network

The CSPS network structure diminishes CDC and grantees' formal authority and control over other network partners. Some control over network partners is facilitated by formal, funding relationships which provide a level of authority and subsequent accountability. But even with a funding relationship, CDC is struggling to impose authority and expectations, particularly around data collection and reporting. Several challenges contribute to the problem. As noted above, programs had been left to their own for a long time and are relatively resistant to the imposition of new requirements. Next, the strength of control is influenced, to some extent, by the funding mechanism used. In this case, the grant mechanism[30] used by CDC affords grantees extensive latitude in program implementation. And even with funding relationships, CDC's formal control is weakened with every step down the vertical implementation chain. For instance, while

---

30 DSTDP used grants as the funding tool for its CSPS awardees up until January 2009.

CDC may have some authority over the state level grantees, its power over the local level

is limited and primarily exercised through the state-local funding relationship. Most

grantees, which use a variety of funding mechanisms, also have limited authority over the

local level implementers.

More challenging than the vertical funding relationships, however, are the

voluntary, horizontal partnerships that are frequently informal in nature and unfunded.

Participants spoke of the challenges in enlisting correction facilities to conduct STD

screening among inmates and in engaging private physicians. Without mechanisms of

formal control, DSTDP encourages a model of "influence" as part of the community

approach – programs using their STD expertise as leverage to influence network partners

to participate in STD prevention and control activities. But influencing and persuading

the informal, horizontal partners is complex, challenging, and requires different

proficiencies than those traditionally held by front-line STD staff.

5.1.2.4 Network Function: Shared Organizational Goals and Priorities within the
Network

The decentralized nature of CSPS inherently results in differences around

organizational goals and priorities. While CDC and the grantees share some common,

broad public health goals related to STD prevention, grantees also hold other priorities

that may be shaped by their own context and political imperatives. The most significant

conflicts in mission and goals exist between STD programs and their horizontal network

partners that represent other sectors or disciplines. These differences challenge the ability

to forge alliances and engage partners in support of STD prevention activities. For

instance, while STD programs are encouraged to enlist adult corrections and juvenile

detention facilities in STD screening efforts, those institutions have very different missions. Consequently, engaging their involvement is difficult. Similarly, gaining the trust and respect of private physicians in a given community can be challenging. Participants emphasized the importance of developing personal relationships to overcome these differences. Good communication and negotiation skills among staff were cited as essential to working with partners, along with strong skills in STD prevention and control to secure their status as experts in the STD arena.

## 5.1.2.5 Network Function: Context, Resources, and Capacity

Grantees vary in regard to their context, resources, and capacity. In terms of context, grantees differ in their STD epidemiology and in the populations affected. Some grantees have a much greater incidence of syphilis, for instance, than others. The differences in epidemiology result in variability across grantees in regard to the program activities and partnerships emphasized. Even areas with similar disease burden may be dealing with very different populations. Context also varies in terms of institutional arrangements within a state or territory and has implications for both vertical and horizontal collaboration. For instance, some states may support collaboration with school districts and others restrict it.

Grantees also vary significantly in regard to their level of resources and capacity. The CDC award size varies across grantees; some have CDC direct assistance while others do not; and some receive state sources while others have none. Capacity also varies, affected to some extent by resource levels. As noted earlier, a small program may be dependent on two or three staff to manage several programs for an entire state (HIV, STD, TB) while another may be many disease specialists working all across the state.

**5.1.3 Performance Measurement System**

<u>5.1.3.1 Process to Develop the Performance Measurement System</u>

DSTDP staff in the Office of the Director initiated efforts to develop a performance measurement system in 1999 enlisting a consultant from Georgia State University. The measures have evolved over time – most recently, a revised set of 17 measures were included in a new five-year cooperative agreement awarded in January 2009. Table 17 summarizes the 2009 measures according to program area and notes the year each measure was originally introduced.

**Table 17. CSPS Performance Measures (2009)**

| Program Area | Performance Measure | Year Introduced |
|---|---|---|
| **Medical and laboratory services (MLS)** | MLS 1a. Proportion of female admittees to large juvenile detention facilities tested for chlamydia | 2004 |
| | MLS 1b. Proportion of female admittees to large juvenile detention facilities diagnosed with Chlamydia | 2009 |
| | MLS 2a. Among clients of IPP family planning clinics, the proportion of women with positive Chlamydia tests that are treated within 14 and 30 days of the date of specimen collection | 2004 |
| | MLS 2b. Among clients of IPP family planning clinics, the proportion of women with positive gonorrhea tests that are treated within 14 and 30 days of the date of specimen collection | 2004 |
| | MLS 3a. Among clients of STD clinics, the proportion of women with positive Chlamydia tests that are treated within 14 and 30 days of the date of specimen collection | 2007 |
| | MLS 3b. Among clients of STD clinics, the proportion of women with positive gonorrhea tests that are treated within 14 and 30 days of the date of specimen collection | 2007 |
| | MLS 4. Proportion of primary and secondary syphilis treated within 14 and 30 days of the date of specimen collection | 2009 |
| | MLS 5a. Proportion of female admittees entering selected project area adult city and county jails that were tested for syphilis | 2004 |
| | MLS 5b. Proportion of female admittees entering selected project area adult city and county jails tested newly diagnosed with syphilis (any stage). | 2009 |

| Program Area | Performance Measure | Year Introduced |
|---|---|---|
| | MLS 5c. Proportion of female admittees entering selected project area adult city and county jails diagnosed with syphilis (any stage) treated within 14 and 30 days of the date of specimen collection | 2009 |
| **Partner services (PS)** | PS 1. Proportion of primary and secondary syphilis cases interviewed within 7, 14, and 30 calendar days from the date of specimen collection, per primary and secondary syphilis case | 2007 |
| | PS 2a. Number of contacts prophylactically treated (disposition A) within 7, 14, and 30 calendar days from day of interview of index case, per case of primary and secondary syphilis | 2009 |
| | PS 2b. Number of contacts brought to treatment as new syphilis cases (disposition C) within 7, 14, and 30 calendar days from day of interview of index case, per case of primary and secondary syphilis | 2009 |
| | PS 3. Proportion of ALL gonorrhea cases interviewed within 7, 14, and 30 days from the date of specimen collection. [for non-HMAs only] | 2009 |
| **Surveillance and data management (SDM)** | SDM 1. Proportion of reported cases of gonorrhea, Chlamydia, primary and secondary syphilis, early latent syphilis, and congenital syphilis sent to CDC via the National Electronic Telecommunications System for Surveillance (NETSS) that have complete data for age, race, sex, county, and date of specimen collection | 2004 |
| | SDM 2. Proportion of reported cases of gonorrhea, Chlamydia, primary and secondary syphilis, early latent syphilis, and congenital syphilis sent to CDC via NETSS within 30 and 60 days from the date of specimen collection | 2004 |

| Program Area | Performance Measure | Year Introduced |
|---|---|---|
| | SDM 3. Proportion of reported cases of primary and secondary syphilis and early latent syphilis sent to CDC via NETSS where sex of the sex partners is known | 2009 |

The process of developing measures for CSPS has been an evolving and incremental one. From 1999 -2001 a multidisciplinary workgroup, including representatives from NCSD, developed a program logic model depicting CSPS activities, outputs, and outcomes (immediate, intermediate, and longer-term) to help guide measurement development. Criteria were defined to aid measure selection (e.g., measures should be meaningful, valid, reliable, timely, actionable). The workgroup identified roughly 60 candidate performance measures that were piloted by a group of seven grantees over a two year period (2001-2003). Based on the pilot and continued efforts of the workgroup, the first set of 12 CSPS performance measures was implemented in 2004 as part of a new five-year funding announcement and made a reporting requirement for the grant. Since that time, some measures have been dropped, others revised, and new ones added. In 2007, a revised set of 14 measures was introduced, and as data collection for this study ended, a revised set of 17 measures was finalized for inclusion in the new cooperative agreement awarded January 1, 2009. Grantees are required to report on the new 2009 measures twice annually. Table 18 summarizes some of the developmental milestones by fiscal year.

**Table 18.  CSPS Performance Measure System Development, 1999-2008**

| Fiscal Year | System Development |
|---|---|
| 1999 | Performance measurement effort initiated by DSTDP's Office of the Director; Interdisciplinary performance measurement workgroup convened |
| 2000 | Program logic model finalized specifying inputs, outputs, and outcomes |
| 2001 | Two year pilot project launched with 7 grantee sites and roughly 60 potential measures |
| 2003 | Pilot project ends and participating sites recommend measures to CDC |
| 2004 | 12 performance measures introduced for CSPS as part of 2004 five-year funding announcement |
| 2004-2007 | New division-wide performance measurement workgroup established, workgroup charter written, performance measurement guidance document developed, annual performance measurement consultations convened, performance measures revised, first performance measures data report issued, quality assurance learning tours initiated |
| 2008 | New performance measures proposed for fiscal year 2009 funding announcement, grantee feedback obtained, 17 measures finalized for introduction with new 2009 cooperative agreement in January 2009 |

The performance measurement development process can be characterized by extensive stakeholder involvement, increasing sophistication in the development process itself over time, and an intention to move toward population-based, short- or intermediate-level outcome measures reflecting the community approach endorsed by DSTDP. Stakeholders are an integral part of the development process helping to ensure that multiple perspectives are represented and that a variety of people contribute to developing and implementing the measures. Key stakeholders in the development process include the grantees, NCSD, and the staff and management of DSTDP. Grantee and NCSD representatives have been involved in developing and piloting the measures,

serving on performance measurement workgroups, participating in annual in-person consultations, and providing feedback on proposed measures. Stakeholder involvement is critical to the measurement development process, in particular, ensuring that the measures are meaningful, feasible in terms of data collection and reporting, and support the intended purpose of program improvement. Stakeholder involvement is also viewed as critical to securing the buy-in needed to effectively implement the measures and promote data use.

The involvement of stakeholders has not been without missteps, however. Most recently, proposed measures planned for inclusion in the 2009 CSPS funding announcement were not vetted with the grantees and NCSD. Upon realizing the error, grantees were given the opportunity to review the proposed measures and their feedback led to the exclusion of several of the proposed measures in the final set. In almost all respects, however, DSTDP conducts a thoughtful, systematic process in developing its measures.

The performance measurement development process has gained in sophistication over time. After the initial measurement workgroup dissipated, an interdisciplinary performance measurement workgroup was reconstituted in 2004. The workgroup includes representatives from DSTDP branches and management along with four members of NCSD's Program Operations Workgroup. A charter specifying the workgroup's function, membership, management, communication, authority, and coordination was drafted in 2004. In 2006, a detailed performance measures guide was disseminated to all grantees providing extensive information about each measure. Also in 2006, DSTDP introduced "performance measure learning tours," site visits to the

grantees to assess implementation of the measures and learn how the measures have affected project areas. DSTDP issued its first report of the performance measures in September 2007 based on data collected in 2005-2006. With NCSD and grantee agreement, state-specific data were included in the report.

In anticipation of the forthcoming 2009 funding announcement, the performance measurement workgroup led an effort to develop new or revised measures. Proposed measures were discussed at the annual performance measure consultation meeting held in Atlanta in the fall 2007. The workgroup hosts this consultation each year, inviting representatives from the workgroup along with other grantee and NCSD partners to provide advice on the development of the CSPS performance measures. Based on the consultation and other workgroup meetings, a proposed list of measures was drafted and, as noted above, was reviewed by grantees in early 2008. A final list of measures was approved by the DSTDP Division Director in March (table 17).

The brief history of the development process highlighted above reflects the hard work of some people in the Division who have championed the performance measure project. The leadership of these champions has been central in moving the project forward, implementing systems and structures to institutionalize the measures, and encouraging its adoption both within and outside the Division. Many on the performance measurement workgroup, including NCSD representatives, share the commitment and enthusiasm of these champions. At the same time, however, many participants discussed concerns that Division leadership and some of the DSTDP program consultants have not fully embraced the performance measures or recognized their importance.

Finally, the process to develop performance measures has reflected an attempt to develop more ambitious measures over time. The twelve measures originally introduced in 2004 were perceived as manageable, feasible, and achievable for the grantees, if not particularly challenging. The workgroup called them GRAM measures or "get right at 'em" because the measures were the easiest to implement at the time.

In developing the 2009 measures, the performance measurement workgroup proposed several outcome-level measures that were population-based focusing on the proportion of diagnosed STD cases that are treated (e.g., proportion of gonorrhea cases in men and women that were treated; proportion of diagnosed cases of chlamydia among women in juvenile detention facilities that were treated; proportion of diagnosed cases of chlamydia among men and women ages 15-25 that were treated). Grantees' comments about the proposed measures, collected through the vetting process, reflect a range of concerns including the feasibility of performing well on measures reflecting STD cases managed by private providers; the feasibility to perform well on population-level measures given limited resources and a lack of control over the broader health care system; the potential burden related to data collection and reporting; the capacity of grantees to address and achieve performance targets given limited resources; the consistency of the measures with program priorities; and the lack of access to data needed for some measures. Based on over 80 comments received from about 22 grantees, several of the proposed measures were excluded from the final set.

5.1.3.2 Performance Measurement System Design

   In this section, the design of the CSPS performance measurement system is

described based on several key aspects of its design. Table 19 summarizes these design

features.

**Table 19. Design Features of the CSPS Performance Measurement System**

| Design Feature | CSPS Performance Measurement System |
|---|---|
| Purpose of system | Program improvement and accountability |
| Level of measurement | Local and grantee level |
| Type(s) of performance measures | Process and immediate outcome measures |
| Use of targets or standards | Baseline and 3-year targets set by grantees |
| Quality assurance efforts | Yes |

5.1.3.2.1 Purpose

   Program improvement and accountability are the primary purposes for the CSPS

performance measures. In addition, the measures are an important means to help focus

program efforts or, at least, highlight some aspects of a program that all grantees should

be measuring in order to improve programs. In communications with grantees, the

purpose of program improvement is largely promoted by emphasizing the value of

performance measurement for program monitoring, to facilitate comparisons of program

performance over time, and to help identify best practices. While participants emphasized

that grantees' performance on the measures is *not* tied to funding, DSTDP has a clear

expectation that programs should be held *accountable*. Grantees are legally required, as a condition of funding, to report data for the measures semi-annually. Accountability was viewed as an important incentive for improving performance, even without the imposition of any sanctions or threat of punitive action.

Prior to the introduction of the measures, the use of performance data by grantees to support their day-to-day program management is unclear. From a national perspective, the introduction of the CDC performance measures in 2004 represented the first time since 1992 that grantees were required to submit program data other than case-base surveillance data. Allaying grantees' fears and nurturing acceptance of the measures has taken time. CDC's hope, however, is that the performance measures will encourage and help direct programs to explore underlying program issues that are behind the performance on any given indicator.

The program-level performance measures are aligned with the Division's program performance measures developed for GPRA and PART, efforts aimed to improve transparency and accountability in government. DSTDP has been PART-reviewed by OMB twice, most recently in 2007 when a review of the entire NCHHSTP was conducted. The STD program-level performance measures were important in securing an overall PART rating of "excellent," the highest qualitative score possible.

5.1.3.2.2 Level of Measurement

The 2009 performance measures relate to three program areas: medical and laboratory services (MLS), partner services (PS), and surveillance and data management (SDM). The measures primarily reflect data collected at the local level where program

implementation occurs, although the SDM measures that relate to timely and complete data are the responsibility of the grantee.

5.1.3.2.3 Types of Measures

All but one of the ten MLS measures focus on the testing, diagnosis, and treatment of gonorrhea, chlamydia, and syphilis in specific population groups and settings (i.e., female admittees to juvenile detention facilities, women in IPP family planning clinics, women in STD clinics, female admittees of adult city and county jails). One MSL measures is population-based focused on the treatment of all reported syphilis cases. Based on the CSPS logic model, the measures can be categorized as process (testing and diagnosis) and immediate outcome (treatment) measures.

Three of the four PS measures relate to syphilis and the fourth focuses on the proportion of gonorrhea cases interviewed in areas that are not categorized as high morbidity areas for syphilis. These measures are all population-based and categorized as process and immediate outcomes measures. The three SDM measures focus on complete and timely data submission and are process measures.

5.1.3.2.4 Use of Targets or Standards

Given the variability in disease burden across grantees, programs are encouraged to establish their own baseline figure for each indicator and set projected, three-year targets. Some suggested that DSTDP lacked the data to support target-setting at this point in time.

5.1.3.2.5 Quality Assurance Efforts

Several efforts have been instituted to support data quality. As mentioned, a guidance document was produced in November 2006 and has been updated to reflect the new 2009 performance measures. For each measure, the guidance document details its rationale, specification, and data sources; reporting criteria; and information about how to use the measure to improve performance. As example, for performance measure MLS 1a[31], the guidance document provides a rationale for the measure, acknowledging that while the STD programs do not have direct control over the juvenile facilities, grantees are expected to actively work with facility managers to increase their awareness about the importance of chlamydia screening. The guidance document also specifies the definition of "large juvenile detention facilities as those that book 500 or more adolescent females annually." In addition, the document suggests possible data sources for both the numerator and denominator. Consequently, the guidance document is an important quality assurance tool to make certain that all 65 grantees define terms similarly (e.g., "C" disposition), calculate the measures in the same way (i.e., numerator and denominator), and report the data in a consistent fashion.

Another important quality assurance initiative is the performance measures learning tours. The objective of the learning tours is to assess the implementation of the performance measures, and, in particular, to explore issues of data quality. A team comprised of the DSTDP program consultant and performance measurement workgroup members conduct the two and a half day visit with grantee staff. A detailed learning tour

---

[31] MLS 1a. The proportion of female admittees to large juvenile detention facilities that were tested for chlamydia.

guide has been developed to facilitate the site visit. The CDC team meets with a group

from the grantee agency and discusses each measure, looking at overall performance on

the measure over time, data sources for the measure, and efforts to assure data validity.

The group also discusses how data are being used to inform program management and

improve performance – this provides an opportunity for CDC to identify technical

assistance needs. Following the learning tour site visit, CDC develops a report of

recommendations and actions for the grantee, for which the program consultant has

responsibility for monitoring.

Even with these efforts, however, data quality for the performance measures

remains a major concern. Internal data analysis and the learning tours have identified

quality problems the performance measurement report issued in 2007 highlights data

quality as an important problem. Data problems identified include missing or unreported

data and the reporting of inaccurate data. Variability across grantees in the data systems

each uses may be a factor affecting quality. Data quality issues were also highlighted in

several abstracts accepted for poster presentations at a national STD conference. For

example, instances have been reported of grantees fabricating data or making rough

estimates to meet reporting requirements. And nearly all of those interviewed spoke

about data quality problems including misinterpretation of the measures, the challenge to

CDC of assessing data validity across so many varied grantees, the lack of buy-in for the

performance measures that may contribute to poor data quality, and the capacity of

grantees to collect and report quality data from the local level.

5.1.3.3 Use of the Performance Measurement System and Data

In regard to data use, some applications of the data have been already noted. For instance, the implementation of the program-level performance measures were highlighted in the most recent PART review and the first data report was produced in 2007. While that report provided grantee-specific data for each measure, performance comparisons across grantees were not particularly useful given the extensive variability in epidemiology, context, and resources.

Overall, use of the measures has been limited despite the mechanisms developed to support institutionalization of the performance measures (e.g., guide, learning tours, technical assistance, report). Few people interviewed believed the data were being used to support program management and many expressed disappointment and frustration about its limited use. The lack of buy-in for the performance measures by the CDC program consultants and grantees may be one factor affecting use at this time. First, while program consultants are viewed as the primary change agents in promoting the adoption and use of performance measurement, their endorsement of the measures is mixed and many are not yet "managing" with the data in their regular interactions with grantees. Program consultants face competing priorities (e.g., personnel issues related to field staff, evidence-based action plans) which may interfere with promoting the performance measures more ambitiously.

Second, buy-in on the part of the grantees has been slow to earn. Several interviewed said the grantees see the performance measurement effort as an "unfunded mandate," rather than part of their evaluation activities. Many grantees are simply collecting and reporting the data to meet the CDC reporting requirement. A few

participants described situations where program consultants sometimes enter the data into the web-based system for the grantees. Data revealed little, if any, engagement of the local level in use of the performance data.

However, some grantees are beginning to use the performance data for planning and improvement. A few states are developing state-specific measures and others are collaborating with neighboring states to review data and discuss opportunities for program improvement. Given that few grantee staff were included in this study, little is understood about their specific barriers to use (e.g., measures are not perceived as meaningful, time and resource constraints, lack of understanding, insufficient tools supporting use).

## 5.2 Findings

### 5.2.1 Dependencies and goal conflicts with CSPS' local-level, horizontal network partners compromises grantees' control over performance.

While CSPS' decentralized, networked approach is essential to provide STD prevention and control services at the local level, it introduces challenges in regard to grantees' ability to control performance on specified measures. As described in section 5.1.1.2, DSTDP encourages a "community perspective" to more effectively reach populations at risk for STD infections and contribute to broader, population effects. The community perspective is intended to expand grantees' responsibility and accountability for STD prevention beyond simply efforts within publicly-funded STD clinics. As one person asked,

> Is it 'STD control' or is it only dealing with the patients that come to the STD clinic, or only dealing with the patients that are reported? What we're trying to move people's thinking to is, 'you're responsible for controlling STD in your *community*.' Obviously there are a lot of things that you can't

do that go into contributing to STD control in your community, but if we narrow it too much to only the things that *you* can do, you're not really in 'STD control' anymore – you're in a 'clinic control' model.

The approach involves extending collaborative partnerships, especially with horizontal partners, to engage those who have access to high risk populations (e.g., primary care physicians, jails, juvenile detention centers). One participant said,

> The community perspective is not just what you have control over, but a good STD program should try to influence other parts that you don't have control over. For example, private providers, family planning clinics. Now this is different from the syphilis requirement where they actually give money to a CBO. This is, we don't give them any money, we don't do anything, this is just a good [grantee] program should be recognized as the STD expert in their county or in their state and therefore, should try to influence private providers to screen more, should educate them more about recognizing symptoms or how to conduct an interview with a teenager about their sexual activity.

But while influencing the practice of physicians, corrections, and others to participate in STD prevention is necessary to achieve longer term, population-level outcomes, mission and goal conflicts and the lack of formal authority over these horizontal partners challenge grantees' ability to affect performance on measures aligned with the approach.[32] Mission conflicts frustrate grantees' efforts to engage horizontal partners and the lack of formal authority compromises their influence – both limit grantees' ability to affect performance on identified measures. As one participant stated, "The dollar bill usually influences pretty reliably, but the smile and handshake are not quite as reliable." Another person explained,

> I think it's pretty clear why grantees don't like that – you're calling it performance, which is a reflection on them, and they're not performing because

---

[32] For instance, performance measure MLS 4, the proportion of female admittees entering selected project area adult city and county jails that were tested for syphilis.

someone in their community isn't supporting them or there are things that are beyond their control. And other people just have the luck of being in a place where they have a cooperative prison system or something.

Grantees exerted their political muscle to successfully reject some of the population-based, performance measures proposed for 2009 (e.g., proportion of gonorrhea cases in men and women that were treated) that were viewed as too far outside their control. DSTDP has been more successful in incorporating process-oriented, venue specific performance measures. The Division's dilemma around performance measurement given the network structure is that while networks are necessary to achieve critical outcomes, grantees' have minimal control over affecting performance on measures dependent on informal, horizontal partners. Consequently, DSTDP has been limited in the choice and types of measures included in their performance measurement set.

### 5.2.2 Given the CSPS network context, performance measurement is a negotiated and incremental process.

Efforts to develop a performance measurement system for CSPS were started in 1999 and have evolved significantly over the past decade. The size of the CSPS implementation network, its culture, and network members' political influence have all shaped what has been an incremental development process based on negotiation with grantee and NCSD stakeholders.

The mere size of the network requires that performance measurement be understood and adopted by hundreds of network partners. CSPS' vertical network includes 66 state, city, and territorial grantees along with hundreds of local level partners. The horizontal network extends that vertical chain at all levels (federal, state, and local).

224

Consequently, the development and implementation of the CSPS performance measures has taken time. Piloting measures is necessary to assess their feasibility and utility across a spectrum of grantees; vetting of measures is needed to further assess feasibility and build a stake in the system; and technical assistance and training is required to help grantees to understand the measures and build necessary data collection and reporting systems.

In developing the CSPS performance measures, DSTDP has faced a stubborn institutional culture resistant to CDC oversight and data reporting requirements. Grantees had enjoyed extensive autonomy since the late 1980s with few stipulations from CDC. And DSTDP's use of a grant mechanism, a policy tool offering grantees a good deal of independence, limited the Division's authority. In discussing the challenges to developing and implementing performance measurement in the CSPS network, one person said,

> I would say the biggest challenge has simply been to change the culture or
> the mindset [of grantees] to view these [measures] as valuable tools to
> manage and run programs. I don't think that it [performance
> measurement] was part of the culture necessarily. To get people
> comfortable with the idea that you can actually evaluate how your
> program is doing, at least to a certain extent, by coming up with well
> thought out measures that can be quantified and looked at over time – that
> this [performance measurement] is a valuable tool. To me, that's been the
> number one challenge is to get people into that mindset and it's still an
> ongoing process.

Champions of the performance measures have also confronted a lack of commitment from management and staff within DSTDP that has slowed the effort. Leadership from Division management as well as program consultants who work most closely with the grantees is essential to promoting performance measurement to the larger network. A participant commented,

The other challenge internally is just to get everybody in the Division on the same page as far as the importance of the performance measures and [the idea that] everybody needs to be working towards the same goal and that there should be minimum standards and expectations around what you do with regards to performance measures in your day-to-day job. And that hasn't happened yet.

Finally, given CSPS stakeholders' political influence, the development process is, to some extent, a negotiated one. While DSTDP is inclusive in its development process, in both good faith and with appropriate intentions (assuring feasibility and acceptance of the performance measures), the Division also recognizes that the network commands political influence that affects the choice and type of measures selected and requirements related to their reporting and use. As noted above, stakeholders rejected several population-based measures proposed for 2009 that were excluded from the final set.

Consequently, the CSPS network has important implications for the development process of its performance measurement system. DSTDP has long recognized the necessity of involving network stakeholders in the development process. Indeed, participants suggested stakeholder participation is essential to assure a feasible and meaningful set of performance measures and to build the buy-in needed to facilitate their adoption. However, the size of the CSPS network along with its cultural norms and political sway all prescribe a more incremental and negotiated development process – one that has frustrated DSTDP given what one person called its "glacial progress."

### 5.2.3 Extensive variability across grantees influences the design of the CSPS' performance measurement system.

The 66 grantees that comprise the CSPS implementation network represent states, cities, and territories that differ substantially in regard to disease burden, populations

affected, geographic context, resources, and capacity. The prevalence of STDs varies across the U.S. and its territories – while some states face epidemic levels of syphilis infection, others experience few cases in a given year. Grantees also vary in terms of populations affected by STDs – states face differing prevalence of disease among African Americans, Latinos, Whites, heterosexuals, and homosexual men. Some states have few staff and are responsible for expansive geographic areas; others support hundreds of staff across rural and urban contexts. Resources range dramatically from just over $46,500 for a U.S. territory to over $6 million for a heavily populated state.

All of these factors affect an individual grantee's priorities, program activities, and ability to achieve results. In designing the CSPS performance measures, then, DSTDP must accommodate the variability imposed by the network structure. First, DSTDP faces challenges in identifying a common set of measures that has relevance for all 66 grantees. One person said,

> I guess that's probably *the biggest challenge* [variability across grantees]
> because even being able to come up with the 12 or 14 [measures],
> whatever they have now, and getting everybody to agree. I mean, that was
> a tremendous accomplishment in and of itself.

The syphilis measures present a good example – there are some states with low syphilis morbidity for which the measures have little relevance. Consequently, the performance measurement workgroup added a partner notification measure related to gonorrhea cases for low morbidity areas, a concession that was not necessarily supported by all workgroup members. One participant noted,

> I think our challenge has been to identify a set of measures that is *useful*
> for the *majority* of project areas, both small and large and that's been an
> internal debate in the performance measure work group. There are a
> number of participants on the work group who will come out and say that

227

small states don't matter, because the real outcome measure is reducing morbidity. And reducing syphilis morbidity in New England where all together they have 90 cases of P&S [primary and secondary] syphilis a year compared to Texas that has 90 in a month – there are advocates for not worrying about the small states.

Network variability influenced the original selection of the "get right at 'em" or GRAM measures as the first set of performance measures. Given the diversity in grantees, the set of twelve measures was seen as manageable, feasible, and achievable for grantees, if not particularly challenging. One person said,

You know, these measures were not implemented because they were necessarily the best measures. They were implemented based upon the pilot results that suggested they're ones we can get. In fact, they were called GRAM measures, 'get right at them' measures because they were the easiest, to be quite frank.

Although DSTDP has worked to introduce more challenging measures over time, the network variability continues to pose a quandary. For instance, grantees persist in voicing concerns around different priorities for grantees based on disease burden and populations affected; the capacity of all programs to access, collect, and report data, especially at the local level; and the feasibility of grantees with limited resources to affect performance on this measure or that. One participant spoke to resource issues for some grantees,

Some project areas see it [performance measures] more as a burden. I would say some of the more rural project areas, the frontier states like North and South Dakota, Wyoming, who receive such limited funding to offer an STD program. They see an additional mandate from CDC around performance measures as a little burdensome because they just don't have the manpower and the resources to focus on it – they may not have the staff to do data entry or data collection or analysis to provide the data back [to CDC]. So they struggle with that. Even some of the larger program areas who have [their own] specific priorities, may not take it seriously.

Consequently DSTPD has needed to design a performance measurement system with enough flexibility to accommodate the variability. For instance, grantees have the option to simply "opt-out" on reporting on a particular measure if they can provide an adequate justification. And, as mentioned, DSTDP includes some "alternate measures" for those grantees that are not high morbidity areas for syphilis. Grantees also have the flexibility to set their own baseline and three-year targets for each measure. And finally, DSTDP recognizes that comparing grantee performance on the measures is not particularly useful given their variability.

Aside from flexibility, DSTDP has also needed to attend to data quality issues. A participant said,

> We knew from the pilot that we did back in the early 2000, we knew that everybody's [grantees] interpreting things in their own way. But I think that it was a real eye opener for folks when they started going out on the learning tours to find out that, oh my God – even though it's written very clearly in the guidance document, [we] thought it was clear as day. Even with all of that, people don't read it and still just interpret it their own way.

The vertical network alone challenges the ability of DSTDP to assure the collection and reliable data from hundreds of local level sites – the introduction of horizontal partners only compounds the difficulty. One person said,

> Now I feel relatively good about the quality of data coming out of an STD clinic but all of a sudden you open it up where 66% of those cases are going to be coming from outside the STD clinic?  Who *knows* what we're going to get.  You know, but again, if the goal is to really prevent infertility and measure how well a program is doing, then you've got to look at the big picture from a national level.

## 5.3 Summary

As DSTDP has assumed a more comprehensive, community perspective aimed at achieving broader effects, program outcomes often reflect the efforts of local-level, funded STD programs and informal network partners. Dependencies are created that compromise grantees' control over performance, introduce difficulties in acquiring or collecting data, and diminish data quality. In this context, CSPS grantees are more amenable to performance measures reflecting process, outputs, and short-term outcomes than longer term or population-based measures. Next, the experience of DSTDP suggests that the network context also shapes the development process itself. For CSPS, the development of performance measures can be characterized as a negotiated and incremental process where adoption of the measures is slowed by the breadth and depth of the network as well as CSPS culture. And finally, DSTDP faces extensive variability across the CSPS grantees in terms of epidemiology, geographic context, program priorities and activities, data availability and sophistication of data management systems, and capacity and resources. This variability has challenged DSTDP to identify a common set of challenging performance measures equally meaningful and relevant to the 65 state, city, and territorial grantees. In addition, differences have required that DSTDP shape a flexible performance measurement system and that staff consider individual grantee context in data interpretation.

# CHAPTER 6

# NATIONAL BREAST AND CERVICAL CANCER EARLY DETECTION PROGRAM (NBCCEDP)

## 6.1 NBCCEDP Case Description

### 6.1.1 The Program

6.1.1.1 CDC Organizational Context

The NBCCEDP was established based on the Breast and Cervical Cancer Mortality Prevention Act of 1990 (PL 101-354). The program resides within CDC's National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) and is managed by the Division of Cancer Prevention and Control's (DCPC) Program Services Branch. DCPC develops, implements, and promotes effective strategies for preventing and controlling cancer, working with partners at the national, state, and local level.

Congress began funding the NBCCEDP in 1991 providing $30 million which DCPC awarded to eight states. Over time, the appropriation has increased and by 1997, all 50 states, the District of Columbia, five U.S. territories, and 12 tribes were funded. Today the NBCCEDP represents the Division's largest program. In fiscal year 2008, DCPC awarded a total of $157,226,794 to 68 state, tribal, and territorial grantees through cooperative agreement awards. The current five-year cooperative agreement (program announcement #703) has been in place since June 2007. For the 12-month budget period in fiscal year 2008, grantee awards ranged from $75,000 to $8,821,221 with an average award of $2,312,159 (median award $2,064,252). In 2007, The National Breast and Cervical Cancer Early Detection Program Reauthorization Act of 2007 was signed into

law by the President, authorizing the program for another five years. The legislation establishing the NBCCEDP proscribes several requirements related to its implementation including:

- Grantees provide matching funds of $1 for every $3 of federal funds.

- Of federal funds received by a grantee, 60% must be used to directly support screening services and no more than 10% may be used for administrative expenses.

- Program dollars may *not* be used to support costs for the treatment of cancer or for research.

Program consultants in the Program Services Branch of DCPC are responsible for providing programmatic oversight and technical assistance to the grantees. Up until 2000, the Program Services Branch provided "direct federal assistance" to grantees, that is, CDC staff were assigned to work on-site to provide administrative and programmatic support. While some program consultants have been with the program for many years, staff turnover is not uncommon after three or four years. Physicians, epidemiologists, economists, and behavioral scientists from the Division's other three branches collaborate to conduct evaluation and research of the NBCCEDP program.

Overall, the NBCCEDP is a high profile program for the Division and the Center (NCCDPHP) given both the size of its budget and the political environment surrounding breast cancer. The CDC Director is well aware of the program. As part the Director's quarterly performance reporting to Health and Human Services (HHS), data on a number of indicators related to CDC programs, including a DCPC GPRA measure for mammography, are presented.

<u>6.1.1.2 Program Goals</u>

The NBCCEDP provides breast and cervical cancer screening to low-income women (typically at or below 250% of the federal poverty guidelines) who are un- or under-insured. Since the NBCCEDP's inception in 1991, grantees have served more than 3.2 million women, provided more than 7.8 million screening examinations, and diagnosed over 35,000 breast cancers and 2,100 invasive cervical cancers. Up until 2000, grantees were challenged to identify resources (e.g., charity care) to assure treatment for women diagnosed with cancer through the NBCCEDP given that the program serves un- or under-insured populations. But in 2000, Congress passed the Breast and Cervical Prevention and Treatment Act of 2000 (PL 106-354) after extensive advocacy on the part of the American Cancer Society (ACS), Susan G. Komen for the Cure (heretofore referred to as the Komen Foundation), and others. This law provides a waiver to women screened and diagnosed with cancer through the NBCCEDP so that they may qualify for full Medicaid benefits through the end of their cancer treatment[33]. The law has been instrumental in assuring that women diagnosed through the NBCCEDP receive the medical care and treatment needed.

The NBCCEDP is comprised of eight unique program components including program management, client recruitment, screening, quality assurance, professional development, data management, partnerships, and evaluation. The screening component includes five different program activities: breast and cervical cancer screening, tracking, diagnostic follow-up, case management, and re-screening. The NBCCEDP has developed a conceptual model representing its comprehensive approach to breast and cervical

---

[33] The extension of Medicaid coverage is limited to women served by the NBCCEDP who are U.S. citizens.

cancer control. There is strong scientific evidence supporting the effectiveness of breast and cervical cancer screening in the early detection of disease and reductions in related morbidity and mortality, however, less of a science base is in place to inform approaches for the other supporting services such as public education and outreach, professional education, and case management. Finally, while the focus of the NBCCEDP is on screening for medically underserved women, its educational activities, public and private partnerships, and quality assurance standards are intended to benefit all women.

All grantees report a subset of program data to DCPC called the minimum data elements (MDEs) as a condition of their funding award. The MDEs are a set of approximately 100 standardized data elements considered to be minimally necessary for grantees and CDC to monitor client demographics and clinical outcomes of women screened with NBCCEDP funds. The MDEs also are used to inform NBCCEDP policies and practices, assess the national program's screening outcomes, and respond to the information needs of CDC stakeholders and partners. The nature of the NBCCEDP as a clinically-based, service delivery program is relatively unusual for public health. Contrary to many other public health programs, the NBCCEDP is more easily monitored and evaluated given the ability to measure clinical service provision and related outcomes.

Those involved with the NBCCEDP express a clear and consistent understanding of the program and its goals, a program driven by its authorizing legislation and one focused, for the most part, on reducing the morbidity and mortality related to breast and cervical cancer among low-income, under-insured women served by the program. DCPC

staff and grantees express a sincere appreciation for the gravity of breast and cervical cancer and their responsibility to women in achieving NBCCEDP goals and outcomes.

### 6.1.1.3 Stage of Program Development

Although programs were incrementally funded over time, the program is perceived as relatively mature. A level of capacity for program implementation has been achieved across grantees – both the capacity for service delivery through the establishment of a network of screening and diagnostic providers and the capacity to collect and report the required MDE data. Program policies and procedures have long been in place and stakeholder relationships are well established. The strength of the national data system that has been established over time has rendered a program perceived as data-driven; MDE data are used extensively to monitor program implementation, service quality, and screening outcomes.

### 6.1.1.4 Budget Stability

Funding for the NBCCEDP has remained relatively stable for several years with no significant increases in the federal appropriation. DCPC economists collaborated with researchers at the U.S. Census Bureau to develop estimates of the NBCCEDP-eligible population size by state. Based on existing federal resources, results suggest that the NBCCEDP is currently able to screen approximately 14% of the eligible population[34] or about 1% of the total women in the U.S. (Tangka et al. 2006). Consequently, DCPC encourages the program to achieve efficiencies wherever possible. One repercussion of

---

[34] Women at 250% or less of the federal poverty guidelines and who are un- or under-insured.

level funding has been limited resources to recruit many *new* women to the program given that re-screening is an important goal of the program.

6.1.1.5 Stakeholders

A strong group of stakeholders orbit the NBCCEDP including its grantees and a formal council comprised of the grantee program directors. The council is a committee of the National Association of Chronic Disease Directors (NACDD), which also stays closely informed of NBCCEDP-related issues. A formal federal advisory committee[35] is in place for the NBCCEDP and is comprised of national experts and leaders in the field of breast and cervical cancer screening. The federal advisory group meets annually to review program progress and discuss related policy.

Other stakeholders include the National Association of County and City Health Officials (NACCHO) and cancer-related national organizations such as ACS and the Komen Foundation, both of which have state, regional, and/or local affiliates. In addition, other CDC-funded programs that have a close relationship to the NBCCEDP include DCPC's National Comprehensive Cancer Control program and the WISEWOMAN™ program (Well Integrated Screening and Evaluation for Women Across the Nation). And, of course, Congress and the Office of Management and Budget (OMB) also have an important stake in the program.

6.1.1.6 Political Context

A significant political context encompasses the NBCCEDP given that it is a legislated program, has a relatively large budget compared to other chronic disease

---

[35] Based on the Federal Advisory Committee Act of 1972

programs at CDC, and shares company with a group of strong national advocates for breast cancer (e.g., ACS, Komen Foundation, National Breast Cancer Coalition). DCPC frequently responds to Congressional inquiries, the CDC Director reports quarterly to HHS on its performance, and DCPC staff are periodically required to provide Congressional testimony about the NBCCEDP.[36]

Changes in medical technology and the health care reform movement are having direct effects on the program. For instance, while scientific evidence does not support a significant clinical benefit to the use of digital mammography over traditional film mammography, the technology companies are effectively promoting digital to radiologists. Because the cost of digital is significantly higher, grantees are struggling to identify mammography facilities willing to provide traditional film mammography or those willing to accept the Medicaid reimbursement rate[37] for film mammography when digital is provided. Consequently, to assure access to screening, CDC recently changed its policy to allow reimbursement for digital mammography at the Medicare rate, a procedure that currently has no scientifically determined increased benefit. The policy change has implications for the program overall – fewer women will be screened with the same dollars. The potential for universal health care – whether realized through a national plan or state-by-state adoption (e.g., Massachusetts) – also has implications for the NBCCEDP although these are less clear at this time.

---

[36] Most recently, CDC provided Congressional testimony on January 29, 2008 to the House Oversight and Government Reform Committee chaired by Representative Henry Waxman.

[37] NBCCEDP policy requires that reimbursement for screening, diagnostic, and other clinical services provided through the program is capped at the Medicare rate.

Advocacy groups were important supporters of the legislation that established the

NBCCEDP as well as the 2000 Treatment Act, and they continue to be important partners

to DCPC. However, CDC policies are sometimes in conflict with those of the advocate's,

which has led to tension for both CDC and the grantees. Most notable is CDC's policy

requiring that 75% of women receiving NBCCEDP-funded mammograms be aged 50 and

older given greater effectiveness of mammography in detecting breast cancer among

older women. In contrast, ACS and other groups have issued guidelines that promote

screening for women 40 and older. Grantees face substantial pressure from the advocates

to screen this younger age group.

**6.1.2 The Implementation Network**

The NBCCEDP is carried out through a network of agencies including the 68

grantees and over 22,000 local-level health care providers. The implementation network

for the NBCCEDP can be described based on five dimensions. The first two relate to the

structure of the network: 1) vertical relationships and 2) horizontal relationships. The

other three relate to the function of the network: 3) authority and control within the

network; 4) shared organizational goals and priorities within the network; and 5)

variability in context, capacity, and resources. Each of these is discussed below.

6.1.2.1 Network Structure: Vertical Relationships

The vertical chain for the NBCCEDP begins in Congress with its annual budget

appropriation for the program. Those funds come to CDC (DCPC) and are distributed

through cooperative agreements to the 68 NBCCEDP grantees. A myriad of decentralized

configurations characterize the dissemination of funds by grantees to local level providers

conducting the breast and cervical cancer screening and diagnostic tests. Some grantees contract directly with providers, radiological centers, and laboratories while using their own staff to conduct client recruitment, professional education, case management, quality assurance, and data management. Other grantees fund regional or local-level agencies to implement the program; in fact, some states *require* that programs be implemented through an existing intergovernmental network. In these cases, those regional or local-level public health agencies will typically initiate contracts with health providers to conduct the screening and diagnostics in their service area. The regional or local-level coordinating agency could be a local health department, a community-based organization (CBO), or hospital and might be responsible for carrying out some program activities such as client recruitment, public education, case management, and data management. In a few cases, a funded tribe or territory has its own health center and provides the screening and diagnostic services directly.

Three points should be highlighted about the vertical dimension of the NBCCEDP network. First, the most significant consequence of the vertical network is that service delivery for the NBCCEDP is typically delivered several steps removed from CDC and at least one or two steps removed from the grantee. So while DCPC is fiscally and programmatically accountable for the NBCCEDP to Congress, OMB, and others, service delivery is actually provided by local-level providers at the other end of the vertical chain. In addition, the service delivery network is extremely large with over 22,000 providers involved in delivering screening and diagnostic services for the NBCCEDP.

Second, the vertical network is comprised not only of intergovernmental relationships (e.g., CDC – state health department – county health department), but also

relationships with private non-governmental and public agencies (e.g., CBOs, private health care providers, public hospitals, cancer centers, laboratories). These agencies may have more or less experience participating in public health efforts.

And third, the NBCCEDP goals (i.e., early detection of cancer among the priority population for the program) are primarily achieved through the actions of the vertical network with little dependency on horizontal network relationships. As discussed below, this fact has important implications for program oversight and accountability, even when the network is as extensive as this.

6.1.2.2 Network Structure: Horizontal Relationships

Horizontal relationships are in place at all levels of the vertical structure. CDC collaborates with other federal agencies (e.g., National Cancer Institute or NCI), national partners (e.g., ACS, NACDD, NACCHO), and the federal advisory committee. Within DCPC, the Program Services Branch collaborates with its sister branches including the Cancer Surveillance Branch that oversees CDC's National Program of Cancer Registries and with other CDC programs within the NCCDPHP such as WISEWOMAN™.

Likewise, grantees have established horizontal relationships with local chapters of ACS, Komen Foundation, and regional offices of NCI's Cancer Information Services. Grantees also collaborate with state-based cancer centers, academic institutions, and private industry. Within the grantee institution, staff leading the NBCCEDP program collaborate with other cancer screening programs within their institution (e.g., prostate cancer, colorectal cancer), the cancer registries program, the CDC-funded comprehensive cancer control program, and WISEWOMAN™ in order to maximize opportunities for program integration. If a county or regional system is part of the vertical chain, they may

have horizontal relationships with regional or local ACS, Komen Foundation, and CIS programs. And at the service delivery level, horizontal partnerships may be established between the provider sites and client referral sites (e.g., Title X family planning clinics, CBOs, churches), local ACS chapters, and Medicaid enrollment offices.

For the most part, the horizontal partnerships (at all levels) are voluntary, informal, and un-funded, although grantees may provide funding to some local level agencies (e.g., churches, CBOs) to assist with public education and client recruitment. As already noted, some of the horizontal relationships help promote service integration so that women might receive more comprehensive screening (e.g., breast, cervical, colorectal, cardiovascular). But in general, these partnerships expand the reach of the NBCCEDP, extending its resources and advancing the broader goals of the NBCCEDP such as enhanced access and quality care for *all* women. Some partners like Komen Foundation, the Avon Foundation, and ACS provide additional in-kind and financial resources to grantees.

6.1.2.3 Network Function: Authority and Control within the Network

Authority and control over service delivery within the NBCCEDP are compromised given the decentralized program structure. And, as just discussed, the structure varies – some grantees directly fund screening providers, others involve local or regional-level agencies which then secure providers. In general, as the number of vertical links in the NBCCEDP implementation chain increases, DCPC and grantee's influence over program implementation decreases.

For the NBCCEDP, authority and control over local-level service delivery is strengthened with the presence of a funding relationship. The specific funding tool used

will affect the level of authority and determine any sanctions that can be exercised. As described, DCPC uses a cooperative agreement as the policy tool to fund the NBCCEDP grantees. This funding mechanism supports a collaborative (i.e., "cooperative") relationship between DCPC and the grantees – cooperative agreements offer CDC the opportunity to include programmatic and reporting requirements and allow for substantial DCPC involvement in programmatic decision making. At the same time, cooperative agreements are not as restrictive as a contract and provide NBCCEDP grantees flexibility in managing and implementing the program.

Grantees apply varied funding mechanisms (e.g., contracts, memoranda of understanding, grants)[38] either directly with providers or with regional or local-level agencies. Some NBCCEDP grantees build requirements into the funding arrangements that afford them greater control over program implementation and performance. For instance, some grantees hold back reimbursement to providers or regional/local agencies until those organizations submit required data and invoices. Other grantees use performance-based contracts, incorporating screening requirements such as the DCPC-established performance measures. Funding tools used by DCPC and the grantees may also allow for the imposition of sanctions to influence program implementation. For instance, in the past, when serious concerns about the quality of screening were identified, DCPC has restricted cooperative agreement funds,[39] forcing programs to halt screening women until problems were resolved. For their part, grantees have cancelled

---

[38] A quantitative assessment of the types of policy tools used by grantees or local/regional agencies is not available

[39] Restricting cooperative agreement funds typically involves "freezing" a portion of funds so that the grantee cannot draw down reimbursement until corrective action has been taken.

contracts with providers that do not meet quality requirements or withheld provider reimbursement if screening projections are not met or if data submissions are delayed.

In contrast, little control is possible over unfunded, informal partners, although these relationships are typically less significant in achieving program outcomes. For instance, programs sometimes struggle to control how quickly family planning programs refer women with abnormal Pap tests into the NBCCEDP for diagnostic testing.

With or without funding, participants discussed the value of building collaborative relationships with agencies to increase DCPC's and grantees' influence with important partners and, ultimately, to yield a greater impact. DCPC program consultants who have effective relationships with their grantees may have greater opportunity to encourage specific behavior rather than resorting to threats of punitive action. And grantees that establish strong working and reciprocal relationships with their providers or regional/local partners may be more successful in affecting implementation.

6.1.2.4 Network Function: Shared Organizational Goals and Priorities within the Network

Partners within the NBCCEDP network share similar goals, particularly as it pertains to the value of breast and cervical cancer screening and its intended outcomes (e.g., early detection and treatment). Some differences in priorities are evident, especially related to policy issues such as DCPC's emphasis on screening women age 50 and older for breast cancer. As previously mentioned, ACS and others recommend screening for women 40 and older. But overall, there is strong consensus on the program purpose and goals.

<u>6.1.2.5 Network Function: Variability in Context, Resources, and Capacity</u>

The NBCCEDP grantees are tremendously variable in regard to their context, capacity, and resources. Grantees' context differs in terms of priority populations, demographics, geography, culture, and political environment. While California's eligible population for the NBCCEDP may be ethnically diverse, West Virginia's is primarily white, rural, and poor. The context of tribal and territorial programs is typically in stark contrast to a state program. For instance, DCPC funds a tribal program in Alaska that is responsible for small pockets of Native Americans scattered over an area 40,000 square miles in size. Strategies used to reach women in this context are quite different from those used to recruit women for screening in a dense urban population like Chicago. Different cultural norms across grantees also have implications for program implementation. For instance, women in rural Appalachia have unique barriers to screening that likely differ from the barriers facing women in an urban minority population. Grantees also vary in their capacity and level of resources. As noted earlier, award sizes for the NBCCEDP grantees range from as little as $75,000 to over $8 million. Grantees funded at smaller amounts and with smaller programs overall likely struggle more in terms of capacity, particularly in terms of data management requirements for the program.

**6.1.3 Performance Measurement System**

<u>6.1.3.1 Process to Develop the Performance Measurement System</u>

The NBCCEDP performance measures are based on the MDE data which have been collected on all women served through the program since its inception in 1991. The

NBCCEDP performance measures were formalized in 2004 as a unique set of priority

indicators and are summarized in table 20 according to indicator category and include the

CDC standard or target.

**Table 20. NBCCEDP Performance Measures (2008)**

| Indicator Category | Performance Measure | CDC Standard |
|---|---|---|
| Screening Priority Population Measures | Percentage of initial program Pap tests that are conducted among rarely or never screened women | ≥ 20% |
| | Percentage of screening mammograms provided to women ≥ 50 years of age | ≥ 75% |
| Cervical Cancer Diagnostic Measures | Percentage of abnormal screening results with complete diagnostic follow-up | ≥ 90% |
| | Percentage of abnormal screening results with time from screening test result to final diagnosis > 60 days | ≤ 25% |
| | Percentage of women diagnosed with HSIL, CIN2 , CIN3, CIS[40], Invasive with treatment started | ≥ 90% |
| | Percentage of women diagnosed with HSIL, CIN2, CIN3, CIS with time from date of diagnosis to treatment started > 90 days | ≤ 20% |
| | Percentage of women diagnosed with invasive carcinoma with time from date of diagnosis to treatment started > 60 days | ≤ 20% |
| Breast Cancer Diagnostic Measures | Percentage of abnormal screening results with complete diagnostic follow-up | ≥ 90% |
| | Percentage of abnormal screening results with time from screening test result to final diagnosis > 60 days | ≤ 25% |
| | Percentage of women diagnosed with breast cancer with treatment started | ≥ 90% |
| | Percentage of women diagnosed with breast cancer with time from date of diagnosis to treatment started > 60 days | ≤ 20% |

---

[40] HSIL (High-grade squamous intraepithelial lesion); CIN (cervical intraepithelial neoplasia); CIS (Carcinoma in situ)

The NBCCEDP performance measures are fairly well institutionalized and

accepted by grantees as an important management tool for the program. Table 21

summarizes some key developmental markers related to the measures.

**Table 21. NBCCEDP Performance Measure System Development, 1991-2008**

| Fiscal Year | System Development |
|---|---|
| 1991 | NBCCEDP program implemented in 8 states and the MDE data set implemented, reporting required |
| 1993 | Data Quality Indicator Guide (DQIG) report developed; Data contractor funded to manage MDE data and provide support to NBCCEDP grantees |
| 1997 | All 50 states, Washington D.C., and a group of U.S. territories funded |
| 2004 | 11 priority performance measures with standards formalized for NBCCEDP |
| 2005 | Coding algorithms for performance measures provided to grantees; Subset of measures used as part of broader, performance-based budgeting formula |
| 2006 | Software developed and distributed to grantees to calculate their own performance reports at any time |
| On-going | Revisions to measures made as needed |

As noted, the current performance measures are derived from the MDE data

which have been collected since the program's inception in 1991. The collection and

reporting of the MDE data have always been a program requirement for grantees and

significant resources support the activity. In 1993, a data contractor was secured by

DCPC to manage the MDE data for the NBCCEDP and provide grantees technical

support on data management-related issues. All grantees are required to employ a data

manager using CDC funds for the program and support a data management system for the collection and reporting of the MDE data. The data contractor has developed a data management software program for the MDE data that is available free of cost to any NBCCEDP grantee. An MDE Data Users Manual is provided to all grantees; the manual includes detailed definitions of all MDE variables. The data contractor employs a team of technical consultants with data management expertise and, in particular, an in-depth understanding of the MDEs. Each technical consultant is assigned a group of NBCCEDP grantees to whom he or she provides technical support. The technical consultant works closely with the DCPC program consultant, participating in regular conference calls, all site visits, and semi-annual data calls to discuss the MDE data submissions.

An MDE committee comprised of multi-disciplinary staff from DCPC (e.g., program consultants, epidemiologists, medical advisors) and the data contractor meets regularly to discuss issues related to the collection and use of the MDE data. In the early 1990s, the MDE committee developed a report based on the MDE data called the Data Quality Indicator Guide or DQIG. The DQIG includes 27 indicators with benchmarks to evaluate both data quality and quality of care. In 2004, the "core" performance measures, a subset of clinical measures from the DQIG, were identified as a unique, stand-alone set of indicators and are considered the most important indicators of NBCCEDP program performance. Therefore, the performance measures represent part of a larger set of monitoring data that support program improvement. So while the NBCCEDP grantees were already familiar with these indicators, the eleven selected measures were given new prominence as part of the core performance measurement set. Heretofore, the NBCCEDP core indicators will be referred to as the performance measures.

Stakeholders have had a voice in reviewing and refining the performance measures over time. A workgroup of grantee staff has been convened in the past to review the DQIG. The Council of NBCCEDP Program Directors (heretofore referred to as the NBCCEDP Council), a committee of the NACDD, reviews proposed changes for the larger MDE data set and has vetted the performance measures in the past. The NBCCEDP Council worked closely with CDC in developing a performance-based budgeting formula that incorporates seven of the eleven NBCCEDP performance measures. In addition, an annual meeting of the grantee data managers provides a regular opportunity to discuss the MDE data and performance measures.

The NBCCEDP performance measures have been adjusted slightly over time. For instance, based on discussion with grantees, the standard for one measure related to follow-up of abnormal Pap screening results was revised based on practice-base realities. More recently, the calculation for the measures on follow-up of abnormal screening results was revised given complaints from grantees that they did not have any control over the timeliness of women who were screened elsewhere and referred to the NBCCEDP for diagnostic services. In addition, a statistical test (z-test) is now applied to more accurately determine whether a grantee meets the DCPC standard on a particular measure – this has been important for performance measures reflecting relatively rare events (e.g., diagnosis of invasive cervical cancer), especially for smaller grantees that screen smaller numbers of women overall. Similarly, CDC invoked a policy whereby a measure is not calculated if the denominator data is ten or less. Both adjustments account for variability across grantees and ensure a more equitable application of the measures.

In 2006, the data contractor developed an edit program specifically for the performance measures that was disseminated to all grantees. The computer program permits grantees to produce a performance indicator report at any given time using the precise algorithms to calculate the measures. Using the edit program, grantees can produce the performance report for individual regions within their state, tribe, or territory and for individual providers to better monitor program performance. The edit report is believed to be a factor contributing to improved performance on the measures during the last two to three years.

At this time, many of the NBCCEDP grantees are meeting the established standards for the performance measures. While the standards could be increased for the existing measures, NBCCEDP staff suggest that stakeholders would oppose such an adjustment. Participants suggested that performance measures will continue to evolve as grantees meet existing standards, new priorities emerge, and improvements in measurement develop. Currently, there are measurement challenges that limit the choice of indicators. For instance, while there is an interest in measuring adherence to recommended screening intervals for cervical or breast cancer, women tend to go in and out of eligibility for the NBCCEDP complicating the ability to identify denominator data for that particular calculation. There is also an interest to expand the measures beyond only clinical ones and incorporate indicators reflecting measures of efficiency or other program areas (e.g., public education) that are more difficult to measure.

6.1.3.2 Performance Measurement System Design

In this section, the design of the NBCCEDP performance measurement system is described highlighting the following: purpose of system; level of measurement; types of measures; use of targets or standards; and quality assurance efforts. Table 22 summarizes these design features.

**Table 22. Design Features of the NBCCEDP Performance Measurement System**

| Design Feature | NBCCEDP Performance Measurement System |
|---|---|
| Purpose of system | Accountability, program improvement, budgeting |
| Level of measurement | Local level (patient-level clinical data) |
| Type(s) of performance measures | Process measures (Priority population measures) Short-term and intermediate outcome measures (Diagnostic measures) |
| Use of targets | Yes – Common standards set by CDC |
| Quality assurance efforts | Yes, extensive |

6.1.3.2.1 Purpose

The NBCCEDP performance measures serve several purposes including accountability, program improvement, and budgeting. The NBCCEDP is accountable to three main groups: Congress, the public, and to the women served through the program. DCPC staff view the performance measurement data as invaluable evidence that the program is serving women of highest need and that timely and quality clinical screening and diagnostic services are provided. As noted earlier, both DCPC staff and the grantees

express a sincere responsibility to the women served through the NBCCEDP and view the performance measures as a means to demonstrate accountability in meeting that obligation.

The performance measures also support program improvement at all levels. DCPC, working with its data management contractor, monitors all MDE data, working closely with grantees to provide technical assistance when potential issues are identified. Many of the grantees have developed the capacity to use their performance measurement data in their own jurisdictions to help identify problems at the local level (e.g., a particular provider who might not be conducting diagnostic follow-up in a timely fashion). Using these data allows grantees to quickly make programmatic adjustments. Interviewees suggested that grantees are more likely to focus strictly on the performance measures with their local level agencies or providers rather than the broader set of MDE data.

In 2005, DCPC began using a subset of the NBCCEDP performance measures as part of a performance based budgeting formula. The formula also includes a measure of fiscal responsibility (i.e., annual spending rate) and a score of the grantee's annual application calculated by the program consultant. Many of those interviewed believe that the introduction of performance-based budgeting has led to improved performance by grantees. Data comparisons (i.e., pre- versus post- the introduction of performance-based budgeting) reflect significant increases in performance although other factors (e.g., providing computational tools) have likely contributed to these improvements as well.

6.1.3.2.2 Level of Measurement

As described, the NBCCEDP performance measures are used for multiple

purposes and reflect patient-level, clinical data collected at the local level.

6.1.3.2.3 Types of Measures

The NBCCEDP performance measures include three types: process, short-term

outcome, and intermediate outcome. The measures of the priority populations (i.e.,

women over 50 for breast, rarely and never screened for cervical) are process-level

measures while the diagnostic measures reflect short- and intermediate-level outcome

measures (e.g., diagnosis and treatment). The goals of the NBCCEDP are reflected in

both the choice of measures and the level of measurement. That is, the performance

measures are aligned with the program goal of early detection and treatment for women

served by the program. Consequently, the performance measures reflect clinical

indicators of follow-up for abnormal screening results and initiation of treatment for

women diagnosed with cancer.

6.1.3.2.4 Use of Targets or Standards

DCPC has set a standard for each of the eleven performance measures – standards

are based on varied sources: available clinical evidence, DCPC policy, cumulative MDE

data, and expert opinion. The standards allow grantees some flexibility – for instance,

while DCPC policy encourages that programs screen women age 50 and older for breast

cancer, the standard allows for 25% to be under age 50. Similarly, CDC has established a

standard that 90% of women diagnosed with breast cancer should initiate treatment,

allowing for those women who refuse treatment or who may be lost to follow-up. While DCPC has been fairly transparent over the years in the development of its MDEs and performance measures, participants suggested that there was less explicit rationale about some of the standards.

6.1.3.2.5 Quality Assurance Efforts

Like the service delivery models, data systems vary across the 68 grantees. Some use more decentralized models of data collection in which MDE data is managed at the local or regional level and then submitted to state at required intervals. Others use paper-based systems in which forms are submitted to the state and data entry takes place in a more centralized fashion. The development of web-based systems is on the rise, allowing providers or local managers to enter data into a system where the grantee then has immediate access to it. As mentioned, DCPC's data contractor developed a data management system for grantees in the mid-1990s – about half of the programs use it while others rely on data bases they have developed or modified. One challenge of the decentralized system is that so many different people in different roles conduct data entry – and with staff turnover, training and re-training are a constant issue.

DCPC requires, as a condition of funding, that a de-identified, data file be submitted to the data contractor by all grantees semi-annually. The data contractor then cleans the data, conducts edit checks to assess data quality, develops a set of standardized reports, and produces an aggregate data set for DCPC. The standardized reports produced by the data contractor were developed in collaboration with DCPC's MDE committee and include detailed edit reports, frequency plots, the DQIG, and a summary report of the 11 performance measures. The full set of reports is distributed to the grantees, DCPC

254

program consultants, and DCPC's NBCCEDP data manager. For each MDE submission, a conference call is scheduled between the grantee, their technical consultant (i.e., data contractor), and the DCPC program consultant to review the data reports. In advance of the call, a detailed set of "data notes" is developed and distributed by the technical consultant to the grantee and DCPC staff – these notes reflect any issue (data quality and clinical quality) identified by the technical consultant and are used to guide the discussion during the call. Following the call, the technical consultant prepares a list of action items that require a narrative response by the grantee. This data review protocol has been in place for years and is a fundamental component of DCPC's technical assistance and quality assurance process. The approach is soundly geared toward program improvement. The technical consultant and the DCPC program consultant work as a team and are both considered important resources for on-going support for the grantees. In addition to the semi-annual data review, grantees are responsible for data quality assurance efforts. Edit checks are regularly conducted by grantees along with periodic chart audits to assess the validity of the MDE data against medical records.

DCPC conducted a study from 2003-2007 to assess the validity of the MDE data. Over 5,000 total medical records were extracted and compared against MDE data records in six states representing the largest NBCCEDP programs (i.e., numbers of women screened). Results reflected high quality MDE data for the national program.

6.1.3.3 Use of the Performance Measurement System and Data

As already apparent, the NBCCEDP performance measures are extensively used in ways consistent with their intended purposes of accountability, program improvement, and budgeting. Data suggest that the collection, reporting, and use of the performance

measures (and MDE data, overall) are deeply rooted within the NBCCEDP program
culture. Several factors contribute to their integration in NBCCEDP management
practice. First, the MDEs have been collected since the program's inception, hence, data
collection and reporting have always been an important and valued program activity.
Second, the data component is well funded – grantees are provided resources through
their cooperative agreement award to support data collection and reporting (e.g., staff
salary for a data manager, data system support), DCPC funds a data contractor, and
NBCCEDP staff include a senior data manager and programmer to work with the data
contractor and manage the CDC MDE dataset. And third, the program is "data-driven" –
all levels of DCPC management, including the program consultants, use these data for
decision making and program improvement. In fact, there is an *expectation* that these
data are an integral component of program operations and management.

The performance measures are used by DCPC to hold grantees accountability.
Interviews with two DCPC program consultants and two grantee representatives suggest
that some grantees are also using the data to hold local, regional, or providers
accountable. In addition, the performance measures are used for political purposes.
DCPC and NBCCEDP grantees use the measures to support requests for increased
funding or defend against funding cuts. DCPC regularly uses performance data as part of
Congressional testimony or other reports. Next, the performance management system,
with its regular data reviews, follow-up action reports, and on-going quality assurance all
support the use of the performance measurement data for program improvement. And
finally, the use of the performance measures as part of a larger performance-based budget
formula beginning in 2005 has given the measures even greater attention. Those

interviewed said there was *not* much pushback from grantees in reaction to use of the

performance measures in budgeting, perhaps because of the wide acceptability of the

measures and confidence in the data. There are, however, some unintentional problems

and "gaming" created by tying the measures to funding, including some manipulation of

performance to meet CDC standards and changes in program practice that allow grantees

to more effectively perform on the measures.


## 6.2 Findings

### 6.2.1 Although the NBCCEDP network compromises control over program implementation, DCPC has designed its performance measurement system in ways that support CDC's and grantees' control over performances.

In contrast to most public health programs, the NBCCEDP is a clinical, service

delivery program which offers DCPC benefits in terms of its evaluation. In addition,

DCPC has defined the NBCCEDP goals narrowly to focus on the women served through

the program rather than aiming for broader, population-level effects. Consequently,

grantees are accountable only for the women screened through the NBCCEDP. This is

contrast to the Division's GPRA and PART measures that are population-based and

largely viewed as unreasonable standards for which DCPC should be held accountable.

As noted in section 6.1.1.4, a DCPC economic study conducted in collaboration with the

U.S. Census Bureau showed that, given resource levels, the NBCCEDP is able to screen

approximately 16% of the eligible population[41] or about 1% of the total women in the

U.S. for breast cancer. Consequently, achieving population-level effects in regard to

mortality are viewed as not particularly realistic. As one participant remarked,

---

[41] Women at 250% or less of the federal poverty guidelines and who are un- or under-insured.

The first thing I do is look at it [GPRA/PART measures] and I say, 'this is ridiculous'. You know, these are really bad ways to evaluate this program…I always think it's better to say that this can't be done or we can't answer this than pretend we can. They [OMB] want to see an impact on mortality and they just are not going to see an impact on mortality. There are 15 other things they [OMB] could ask us to show to put some credibility in the program and to evaluate it. It just tells me that the people who are asking the question don't understand mortality.

With program outcomes based on the women served through the NBCCEDP and with the availability of patient-level clinical data (i.e., MDEs), the NBCCEDP has selected 11 performance measures that are closely tied to the work of the health care providers within the vertical network of the program. Because the NBCCEDP represents a service delivery program, DCPC was able to identify a common set of performance measures appropriate and relevant for all even though it is comprised of a vast implementation network with considerable variability – 68 grantees, 22,000 local providers, and horizontal partners at every level.

The types of measures include process-level measures (e.g., who is screened, completeness of data), short-term outcome measures (e.g., timeliness of diagnostic follow-up for women with abnormal screening results), and intermediate outcome measures (e.g., timeliness from the date of diagnosis to the initiation of cancer treatment for women diagnosed with cancer). In this case, the types of measures selected are those over which health care providers have the ability to effect. In addition, grantees are able to maintain some control over their vertical network partners based on their funding relationships and CDC has structured its cooperative agreements to support greater authority over grantees. Therefore, accountability within the NBCCEDP is less fragmented than in networks heavily reliant on unfunded, horizontal partners.

In regard to the 11 performance measures, grantee staff have complained when specific measures are outside what they view as their direct control. Grantees are especially sensitive to this issue given that DCPC is using seven of the 11 measures as part of a performance-based budgeting formula. In response to grantees' concerns, DCPC has made revisions to its performance measures over time to strengthen grantees' control over them. For instance, one performance measure assesses the timeliness from an abnormal Pap screening result to diagnostic follow-up and includes a standard of 60 days. Grantees argued it was not "fair" to hold them accountable for the timeliness of women referred to the NBCCEDP for diagnostic follow-up who had been screened outside the program. That is, grantees argued that they do not have control over the efficiency in which horizontal partners refer women to the program for diagnostics or over how long women wait to come in for diagnostic follow-up. One person explained,

> A good number of these women are referred in [to the NBCCEDP] for diagnostic care because they were screened and identified as having an abnormal Pap somewhere else. It might be that the Title X family planning program screened a woman, they found an abnormal Pap, but it's not within their purview to provide diagnostic testing, so they refer her to us. What the programs don't have control over is how long it takes that woman to show up.

Consequently, DCPC recently changed how the measure is calculated for the purpose of performance based funding to adjust for women referred for diagnostic follow-up. The adjustment not only addresses grantees' issues with the measures, but should  resolve unintended consequences DCPC had documented stemming from how the measure was traditionally calculated – most notably, at least one grantee, worried that referral delays were affecting their performance on the measure, had stopped accepting diagnostic referrals into their program.

**6.2.2 Policy tools, management practice, and network partnerships enhance CDC and grantee control over the NBCCEDP performance measures.**

Given that the decentralized network structure of the NBCCEDP compromises

control within the network, DCPC and grantees apply policy tools and management

practices to help preserve a level of authority over vertical partners on whom

performance is dependent. For instance, DCPC uses a cooperative agreement as the

funding mechanism which affords CDC come control over grantees. One person said,

> The concept [of a cooperative agreement] is that it's a *collaborative*
> *process*. We [DCPC and grantees] set goals together, we monitor progress,
> we work together for improvement, and we make changes as we go
> forward, as needed. So, the whole idea is that the Federal government is
> playing a substantial role in the activities, and that's what the name of it
> means. A lot of people don't understand it that way. Most people in the
> health departments in the higher up levels understand it, but it takes some
> reminding sometimes because they'd like to just receive the money and
> check back with you when it's time to ask for more.

While grantees choose mechanisms to fund local level partners at their discretion,

this study suggests that some are using performance-based contracting, reimbursement

policies, or other means to help preserve control over performance that is otherwise

dependent on local level partners. A participant representing a grantee explained,

"They've [providers] got case load requirements, so if they contract [with us] for a

thousand women [to screen for the year], they've got to serve at least 97% of them in

order to get paid for their administrative part." Another grantee representative suggested

that, over the years, poor performers had simply been replaced. Other grantees withhold

reimbursement until providers or local level agencies submit required MDE data or tie

reimbursement to providers' performance on the NBCCEDP measures.

DCPC also has a powerful monitoring tool in the MDE data set – all grantees are

required, as a condition of their award, to submit MDE data to DCPC semi-annually.

Grantees have developed extensive data management systems that allow them to collect

data from the local level. Consequently, DCPC is able to carefully monitor local-level

service delivery by continually assessing patient-level screening data and intervene as

necessary if and when implementation problems are identified. The same is true for

grantees – for instance, the MDE data can be used to identify specific providers who are

performing poorly facilitating quick intervention. Therefore, the availability of the

comprehensive MDE data enhances control within the vertical network even when

implementation may be several steps removed.

Finally, developing effective relationships between vertical network partners also

facilitates control and allows CDC and grantees to better influence performance. A

representative of a grantee agency explained their success in consistently meeting the

performance measurement standards this way,

> I think it's because we have developed really good relationships with our
> contractors [that they have control]. They're there for us when we need
> something, or we need something done differently, or we've identified an
> issue. And we're there for them, so we're reciprocating.

### 6.2.3 The NBCCEDP network requires that DCPC make significant resource investments to build a comprehensive performance management system in order to ensure data quality and the use of performance measurement data at multiple levels.

NBCCEDP's network structure requires that DCPC maintain a performance

management system that supports data quality and the use of the data for accountability,

program improvement, and budgeting. DCPC has committed extensive resources to this

effort. The provision of resources (e.g., MDE data users manual, data management software), technical assistance (e.g., data contractor's technical consultants, CDC program consultants), training (e.g., on-site, annual data managers meeting), extensive quality assurance efforts, and the semi-annual data review process are all part of a comprehensive NBCCEDP performance management system. These resources represent millions of federal dollars and underscore the investment needed to support performance measurement in a network context.

Since the NBCCEDP's inception, DCPC has nurtured a data-driven culture. MDE collection and reporting were a requirement of the first eight cooperative agreements in 1991. The emphasis on data has been strengthened over time and serious attention to the MDE data and performance measures is given by every level of staffing at CDC from program consultants, Branch management, and Division and Center leadership. One person said, "I think the bottom line is that the programs had a voice in designing the MDEs, they grew up with the MDEs, and for the [new] programs that were added on later, the data were a given. It's just part of the programmatic requirement that starts on day one."

As described in section 6.2.3.1, DCPC hired a data contractor in 1993 to manage the MDE data submitted by grantees and provide technical support to NBCCEDP grantees. The data contractor has worked with DCPC over time to develop a data management software program (that about half the grantees use), the MDE Data Users Manual, and provide technical assistance to grantees. Each NBCCEDP grantee is assigned a technical consultant (from the data contractor agency) and a DCPC program consultant – together this team provides on-going data management and programmatic

support through monthly conference calls, site visits, and annual meetings of grantee

program directors and data managers. Semi-annual data reviews,[42] based on each MDE

submission, represent an important component of DCPC's performance management

system. One person said,

> It's [the data review process] just a cycle that continues forever basically.
> We look at a fairly current period of time and look at performance,
> identify any problems, [including] trend-type problems, and have the
> grantee address it either as a program issue that the CDC program
> consultant would deal with or as a data collection/reporting problem that
> our data contractors could give advice on.

These efforts along with others also help ensure data quality for all MDEs and,

consequently, for the performance measures. Again, the network implementation

structure of the NBCCEDP demands the inclusion of extensive quality assurance

mechanisms at all levels. Edit checks have been built into data management systems,

grantees perform local level chart reviews, and the data contractor conducts edit checks.

One participant suggested that data quality is enhanced when all levels of the vertical

chain are aware of and understand the NBCCEDP performance measures.

> As far as data quality, no matter where the data collection is happening, if
> the grantee relays those [performance] indicators that are expected of them
> down the chain to the providers, or the regional folks, I think that's when
> we see a good quality data set coming in. Everybody's on board of what
> the expectations are from the CDC. So the local folks know that they need
> to get in timely and accurate data to their state so that those indicators are
> met. I've definitely seen programs that relay that information down,
> especially in the same type of feedback reports that the CDC uses, down to
> their local levels. I think there tends to be cleaner data and more accurate
> data if everybody's on the same level.

---

[42] The semi-annual data review process is detailed in section 6.2.3.1

Finally, NBCCEDP's data-driven culture and DCPC's investment in its

performance management system contribute to greater utility of the performance data at

all levels. One person said,

> I've been extraordinarily impressed with how its [data system] become a
> very sophisticated MDE system, number one. And I've been
> extraordinarily impressed with the fact that not only do we collect the data
> but we actually *use it*, which, in a lot of places, even at CDC, either it's
> not collected, or it's not collected well, or if it's collected, nobody ever
> looks at it or uses it! I think that we [NBCCEDP] have the full spectrum.
> Our Division has a lot to be proud in that regard – we manage with data.

Utility has been enhanced by grantees' use of the data with their local level partners. A

participant remarked,

> Some of the states have broken this [performance indicator report] apart
> by their units, their districts, or whatever they have, and say 'Here's how
> we performed as the state, but look at your district, it's much better or
> worse.' We found that the old saying, 'what gets measured gets done'
> works. And people are becoming more accountable by having these
> [performance] reports and being able to see it. That gives them something
> to shoot for, an opportunity to improve.

The most important factor influencing the use of the NBCCEDP performance data

may be that DCPC and grantees alike view the measures as fair, meaningful, and relevant

to the program goals. The choice of measures and the types of measures (i.e., process,

short and intermediate-level outcome) assure that they are tied closely enough to the

program's work so that they are, indeed, actionable. Universally, participants suggested

that the performance measures represented the most important priorities for the program

– primarily, high quality care for women screened through the NBCCEDP. A participant

remarked,

> If you did a random survey of the grantees – or asked all of them – I
> would not be surprised if all of them would say they do not want to get rid

of it [performance measures]. This is a security blanket for them. We're a program that screens women for cancer and this is a security blanket for them to know that women are followed up, and they're not going to find out 3 to 6 months down the pike that a woman had cancer that they missed because they screened her and didn't follow up. So I think it serves a win-win purpose there, and I think the programs clearly see that.

## 6.3 Summary

The NBCCEDP represents a clinical, service delivery system. Given the NBCCEDP's decentralized program structure, DCPC has designed their performance measurement system in ways that increase grantees' control over program outcomes and enhance data utilization. First, rather than assuming a broad population-based perspective, DCPC identified program goals that are consistent with the program authorizing language and emphasize accountability for only those women screened through the program. Second, DCPC selected performance measures that are closely tied to the work of providers, network partners in the vertical chain over which DCPC and grantees have some degree of control given the existence of funding mechanisms and management practices. Third, DCPC has made revisions to the way it calculates some measures to reduce grantees' dependence on factors outside the control of providers. Finally, DPCP has made a significant investment in structuring a comprehensive data management system that has aided the adoption of the performance measures and enhanced utilization of data. In the end, the NBCCEDP performance measures are overwhelmingly viewed as meaningful, valuable, and fair by both DCPC staff and grantees. This may account for the ability of DCPC to use the measures for strengthening accountability, program improvement, and budgeting. One person said, "No matter what the programs say, their data basically speak for them. I think it just brings this common

265

denominator to every program. Despite how they would describe themselves, it's like, well, your data describe you."

# CHAPTER 7

## NATIONAL TOBACCO CONTROL PROGRAM (NTCP)

### 7.1 NTCP Case Description

#### 7.1.1 The Program

##### 7.1.1.1 CDC Organizational Context

In the 1990s, federally-funded tobacco control efforts were supported through two different initiatives: The National Cancer Institute's (NCI) American Stop Smoking Intervention Study for Cancer Prevention (ASSIST) Program, a demonstration program, and CDC's Initiatives to Mobilize for the Prevention and Control of Tobacco Use (IMPACT). In 1999, these two efforts were combined to form the NTCP, and, under CDC's management, all 50 U.S. states were funded. The purpose of the NTCP is to reduce tobacco-related disease, disability, and death. The program is administered by CDC's Office on Smoking and Health (OSH) in the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP).

Today the NTCP represents OSH's largest program. In fiscal year 2007, CDC awarded a total of $66.1 million to 58 grantees (all 50 states, D.C., and 7 U.S. territories). Individual awards ranged from $172,516 to $2,059,294 with an average award of $1,140,588 (median award $1,253,543). The current five-year cooperative agreement (PA 03-022) has been in place since 2003. The agreement was extended for one year in 2008, and a new competitive funding announcement (DP09-901) was issued in November 2008 with an expected award date of March 2009. While some staff in OSH have worked there for many years, there has also been turnover in key management positions. The Division

Director position was vacant for several years, and the leader of the Evaluation Team within the Epidemiology Branch has had four acting branch chiefs over the course of the last 2 to 3 years – the position was permanently filled during the summer of 2008.

7.1.1.2 Program Goals

According to the American Cancer Society (ACS),[43] lung cancer is the leading cause of cancer-related death for both men and women; over 215,000 new cases and nearly 162,000 deaths from lung cancer were estimated for 2008. Smoking tobacco is the cause of over 80% of lung cancer cases. Tobacco control and prevention is complex – multiple factors at the individual, community, and environmental level contribute to tobacco use and the consequent risks posed by secondhand smoke.

OSH established the NTCP to encourage coordinated, nationwide tobacco control activities. From the NTCP's start, the program has articulated four program goals:

1. Prevent the initiation of tobacco use among young people.

2. Eliminate nonsmokers' exposure to secondhand smoke.

3. Promote quitting among adults and young people.

4. Identify and eliminate tobacco-related disparities.

Given the complexity of tobacco control, OSH and its partners encourage a comprehensive approach to tobacco control. In a 2007 report, CDC states, "A comprehensive approach one that optimizes synergy from applying a mix of educational, clinical, regulatory, economic, and social strategies- has been established as the guiding principle for eliminating the health and economic burden of tobacco use."[44] And

---

[43] http://www.cancer.org/downloads/PRO/LungCancer.pdf, accessed January 21, 2009.
[44] CDC, HHS. *Best Practices for Comprehensive Tobacco Control Programs*, 2007.

recognizing that a comprehensive approach cannot be accomplished by any single agency or organization, OSH also emphasizes that tobacco control "requires coordination and collaboration across the federal government, across the nation, and within each state."[45]

OSH encourages the achievement of the four national goals through community interventions and mobilization; counter-marketing; policy development and implementation; and surveillance. The NTCP cooperative agreement supports the following activities: program and fiscal management, strategic planning, collaboration and communication with partners, surveillance and evaluation, training and technical assistance, information exchange, and local grants programs to support community based coalition building, planning, policy development and implementation, and local-level surveillance and evaluation. With the exception of the Quitlines, OSH funds may *not* be used for research or to provide direct services such as individual and group cessation services, patient care, personal health services medications, patient rehabilitation, or other costs associated with the treatment of diseases caused by tobacco use.

### 7.1.1.3 Stage of Program Development

The NTCP can be described as a mature program – it has been in place in one form or another since the mid-1990s and its goals are clear and well articulated. In fact, partners at all levels have adopted the four goal areas that comprise the NTCP framework and also promote them. Over time, the NTCP has developed logic models for three of the four goal areas to model the causal linkages that lead to reductions in tobacco-related

---

[45] CDC, HHS. *Best Practices for Comprehensive Tobacco Control Programs*, 2007.

morbidity and mortality[46] (see figures 1-3 in section 7.1.3.1). The tobacco control field

benefits from a strong science base, the result of many years of well-funded research

efforts. Robust evidence supports a number of effective interventions. Consequently,

OSH has constructed evidence-based logic models that highlight, from left to right, the

inputs, activities, outputs, and outcomes at three levels – short-term, intermediate, and

long-term.

The linkage between behavioral outcomes, such as reductions in tobacco

consumption, and the distal outcomes (i.e., morbidity and mortality) is well established.

Various reports including those from the Institute of Medicine (IOM) along with CDC's

*Guide to Community Preventive Services: Systematic Review and Evidence-based*

*Recommendations*, provide a valuable synthesis of evidence related to tobacco control

activities. In 1999, OSH published a guide titled, *Best Practices for Comprehensive*

*Tobacco Control Programs*, to encourage evidence-based practice in the field.[47] The

guide describes evidence-based, best practices for each of five components of a

comprehensive tobacco control program – state and community interventions, health

communication interventions, cessation interventions, surveillance and evaluation, and

administration and management. The guide also provides recommended program

intervention budgets for each state based on population, prevalence of tobacco use,

infrastructure costs, the number of local health units, the proportion of the population that

is uninsured, and other factors. For instance, OSH recommends an annual investment of

$254.3 million for tobacco control intervention budgets in New York.

---

[46] There is not a strong science-base related to the health disparity goal, therefore, a
related logic model has not been developed.
[47] The guide was updated in 2007.

As a condition of their cooperative agreement awards, NTCP grantees report on their progress semi-annually through a web-based system managed by OSH called the Chronicle. The system captures both narrative and quantitative data. The Chronicle was under a significant redesign effort at the time this study was conducted. Grantees are only required to report on activities funded through OSH dollars, not those reflecting the totality of their tobacco control funds.

### 7.1.1.4 Budget Stability

Like other programs at CDC, the NTCP has been relatively flat-funded for the past six to seven years. The total anticipated award for the fiscal year 2009 program is estimated at $63 million[48]. While the NTCP awards from OSH may reflect the sole resources for some grantees, many others receive additional sources of funds, sometimes many times the size of their CDC award. Individual settlements with the tobacco industry and the 1998 Master Settlement Agreement (MSA) with the tobacco industry provided an influx of funding to states in the late 1990s and beyond. Today, roughly 90% of funds for tobacco control efforts are provided by state legislatures through excise tax revenues and settlement funds[49]. Consequently, grantees often integrate multiple funding streams to support the implementation of their comprehensive tobacco control programs. Unfortunately, in many states, tax and settlement fund resources have been diverted to address other priorities as the country's economic situation has worsened. Massachusetts lost 92% of their state funding in 2002 when the state experienced an economic crisis.

---

[48] Fiscal year 2009 federal budget had not been finalized at the time of this writing.
[49] CDC, HHS. *Best Practices for Comprehensive Tobacco Control Programs*, 2007.

### 7.1.1.5 Stakeholders

The goals of the NTCP can only be achieved in concert with its many stakeholders. Externally these include CDC's sister agencies such as the NCI and the Substance Abuse and Mental Health Services Administration (SAMHSA). National foundations are also critical partners and include the Robert Woods Johnson Foundation, the American Legacy Foundation, ACS, American Lung Association, American Heart Association, Campaign for Tobacco Free Kids (CTFK), and Americans for Non-Smokers' Rights, among others. The National Association of Chronic Disease Directors (NACDD) is an organization comprised of state and territorial chronic disease directors and represents another important stakeholder group. Overall, NTCP enjoys a strong commitment across partner agencies to work together in order to achieve the four goals highlighted earlier.

### 7.1.1.6 Political Context

The broader political climate around tobacco control extends to the NTCP. OSH representatives work closely with Health and Human Services (HHS) offices in Washington D.C. to coordinate tobacco control efforts across federal agencies. CDC also provides staffing to the U.S. Surgeon General's office, meeting with the Surgeon General several times throughout the year and providing support in developing Congressional testimony and other presentations. Over time, the NTCP has benefited from political victories related to tobacco control that have resulted in periods of high visibility and legislative opportunity. During late 1990s, the first states resolved legal battles with the tobacco industry and the MSA with other states was completed. Of course, the tobacco

industry presents a powerful and well-funded political opponent to the entire tobacco control community, including the NTCP. Participants spoke about how the political context within an individual state may influence the areas of focus for a NTCP grantee program. For instance, a particular state may have a strong restaurant and bar lobby that makes it difficult to pursue policy related to second-hand smoke.

## 7.1.2 The Implementation Network

7.1.2.1 Network Structure: Vertical Relationships

NTCP's vertical chain begins in Washington D.C. with Congress, HHS, and others. Federal funds are appropriated to CDC, and staff in OSH's Program Services Branch oversee the 58 NTCP cooperative agreements. As mentioned, these dollars are often complimented by other, more significant resources at the state level, although some grantees have experienced reductions in state contributions over the past several years. For most grantees, a large portion of their NTCP funds are retained by the grantee (state or territory) to support staff and other infrastructure costs that allow them to manage a comprehensive tobacco control program. Some grantees also use a portion of their NTCP funds to support Quitlines[50] and grants to local-level partners. Typically, non-CDC resources (state and private contributions) are used to fund local-level agencies, mass media campaigns, universities, and others. Regardless of how CDC funds are used, however, OSH encourages grantees to involve local-level partners.

In summary, the comprehensive tobacco control approach endorsed by the NTCP involves an extensive vertical network of relationships with between CDC, its grantees,

---

[50] Quitline services are available in every state. The program provides telephone counseling and support for tobacco cessation.

and their partners at the local level, often city or county health agencies or community based organizations (CBOs). Therefore, the vertical chain is not strictly intergovernmental. In addition, the vertical structure within individual states and territories varies – for instance, as a home rule state, Massachusetts must collaborate with 351 towns, each with its own government, while California works with 61 different local lead agencies.

## 7.1.2.2 Network Structure: Horizontal Relationships

Extensive horizontal partner involvement is a prerequisite in achieving a comprehensive tobacco control strategy targeting four different goals (i.e., smoking initiation, cessation, second-hand smoke, and disparities) through a complement of interventions (e.g., educational, clinical, regulatory, economic, and social strategies). The NTCP network includes horizontal relationships at all levels of government, including federal, state, and local. OSH collaborates externally with federal partners such as NCI, IOM, SAMHSA, and the Agency for Healthcare Research and Quality as well as with nongovernmental partners including ACS, CTFK, American Lung Association, American Heart Association, American Legacy Foundation, Robert Woods Johnson Foundation, Americans for Nonsmokers' Rights, and others. OSH policy staff meet quarterly with many of these partners to ensure coordination of efforts. OSH also works closely with national organizations that represent the interests of state and local-level public health groups such as the National Association of City and County Health Officials (NACCHO), the Association of State and Territorial Health Officers (ASTHO), and the NACCD. Given that tobacco-use is an important risk factor for other chronic diseases

(e.g., cancer, heart disease, diabetes), OSH staff work with other CDC divisions within the NCCDPHP.

At the state level, extensive collaboration occurs in support of comprehensive tobacco control. CDC requires that grantees participate in state-wide tobacco control planning and coalition building. States typically collaborate with others in state government (e.g., education, environmental health, other public health programs), state or regionally-based advocacy organizations (e.g., ACS, CTFK), universities, and foundations (e.g., American Legacy Foundation). As a member of these coalitions, the state health department is typically viewed as an influential partner, and depending on resources, may play the leading role in state tobacco-control efforts. The planning efforts help prioritize intervention efforts and coordinate funding support. For instance, the state health department may offer a number of local-level grants for second-hand smoke reduction, American Legacy Foundation funds may support an intervention for youth, and ACS may commit resources to work on state legislative policy initiatives.

At the local level, horizontal partners are also engaged. Local-level coalition building is encouraged and often funded by NTCP grantees to bring together varied partners from public health, education, health care, business, and local advocacy groups to plan for coordinated tobacco control efforts.

### 7.1.2.3 Network Function: Authority and Control within the Network

As noted, OSH uses a cooperative agreement as the funding mechanism to support its NTCP grantees. Participants viewed cooperative agreements as facilitating "shared ownership," but not providing them much control or authority over grantees. Requirements specified in the program announcement (e.g., reporting requirements) are

relatively minimal and, as already noted, OSH can only require reporting on CDC-funded activities. In contrast, grantees more often use contracts to fund local-level partners or memoranda of agreement with other partner agencies.

## 7.1.2.4 Network Function: Shared Organizational Goals and Priorities within the Network

Network partners, both vertical and horizontal, share a strong commitment to the four program goals discussed earlier. At the same time, however, each organization involved in the tobacco control community has their unique agenda and philosophy which sometimes introduces relationship challenges. In addition, states and territories must contend with the tobacco industry which represents views diametrically opposed to their cause. CDC's endorsement of strategic planning, coalition building, and sustainability reflect pointed attempts to bring partners together to build a shared vision and priorities.

## 7.1.2.5 Network Function: Variability in Context, Resources, and Capacity

The NTCP grantees vary tremendously in their context, resources, and capacity. The individual context of a state or territory may be an important factor driving the emphasis of program efforts; one state may focus on second-hand smoke while another gives priority to youth access. Resources also vary and often limit the ability of grantees to address all four program goals. Across the NTCP grantees, program budgets may reflect extreme differentials when all funding sources are considered. And the funding context can change quickly and radically as indicated by the Massachusetts experience

when 92% of their state funding was eliminated in a single year. Capacity varies across grantees as well, influenced by factors such as staff resources and expertise.

### 7.1.3 Performance Measurement System

7.1.3.1 Process to Develop the Performance Measurement System

In 2005, OSH published an evaluation guide titled, *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs* intended for planners, managers, and evaluators of state and territorial tobacco control programs and for CDC's national partners. Over a four-year process, an evaluation team led by OSH's Epidemiology Branch developed 120 evidence-based key outcome indicators (KOIs) that are included in the guide. Using the three logic models developed for the NTCP program goals as a framework, the CDC evaluation team worked to refine the logic models and identify relevant short-term, intermediate, and long-term outcomes. The guide does not address the "left half" of the logic models, that is, the process measures related to inputs, activities, or outputs. Crafted in a "*Consumer Reports*" format, each of the 120 indicators is rated on a number of criteria. Reporting on the KOIs is not required of NTCP grantees, but they are encouraged to use the guide and indicators for program planning and evaluation and to report measures of their choice. Nearly all grantees report on some measures as part of their semi-annual progress reporting through OSH's Chronicle.

Soon after the key outcome indicators (KOIs) guide was published in 2005, OSH embarked on two other related efforts – to develop process-level indicators and to identify a subset of the KOIs called "core indicators" in order to provide a "national snapshot" of the program. The 32 core indicators (and the data sources for each

indicator[51]) are summarized in table 23 and are organized according to the goal area and indicator type as defined by the key outcome indicator guide[52]. Three of the four logic models for NTCP goals are included as figures 1-3. The indicator number in the table below corresponds to the numbered box in the logic model. For instance, two of the core indicators (1.6.3 and 1.6.5) correspond to outcome #6 in the logic model for goal one (i.e., figure 1, outcome #6 – "increased knowledge of, improved anti-tobacco attitudes toward, and increased support for policies to reduce youth initiation"). Likewise, core indicator 1.14.1 corresponds to outcome #14 in the goal one logic model (i.e., figure 1, outcome #14, "reduced tobacco-use prevalence among young people").

---

[51] Addressing Tobacco in Managed Care (ATMC), Survey of Health Plans 1997-1998; National Adult Tobacco Survey (NATS); Behavioral Risk Factor Surveillance System (BRFSS); California Adult Tobacco Survey (CATS); Campaign for Tobacco-Free Kids (CTFK); CDC Health Profiles (Profiles); CDC State Tobacco Activities Tracking System (STATE); CDC Youth Risk Behavior Surveillance System (YRBSS); Current Population Survey: Tobacco Use Supplement (CPS TUS); Youth Tobacco Survey (YTS)

[52] Two of the 32 indicators are included in two goal areas, therefore, there are a total of 30 unique core indicators.

**Table 23. NTCP Core Outcome Indicators**

| Goal Area | Core Indicator and Data Source(s) | Indicator Type |
|---|---|---|
| **Goal One: Preventing Initiation of Tobacco Use Among Young People** | 1.6.3 Proportion of students who would ever wear or use something with a tobacco company name or picture (YTS) | Short-term outcome |
| | 1.6.5 Level of support for increasing excise tax on tobacco products (NATS) | Short-term outcome |
| | 1.7.1 Proportion of schools or school districts reporting the implementation of 100% tobacco-free policies (Profiles) | Short-term outcome |
| | 1.8.2 Proportion of jurisdictions (State and Local) with policies that require retail licenses to sell tobacco products (measurement protocol under development) | Short-term outcome |
| | 1.8.7 Changes in state tobacco control laws that preempt stronger local tobacco control laws (STATE) | Short-term outcome |
| | 1.9.1 Extent and type of retail tobacco advertising and promotions (measurement protocol under development) | Short-term outcome |
| | 1.9.10 Number and type of Master Settlement Agreement violations by tobacco companies (measurement protocol under development) | Short-term outcome |
| | 1.10.5 Proportion of young people who are susceptible never-smokers (YTS) | Intermediate outcome |
| | 1.11.1 Proportion of successful attempts to purchase tobacco products by young people (YTS, YRBSS) | Intermediate outcome |
| | 1.12.1 Amount of tobacco product excise tax (STATE, CTFK, State departments of revenue) | Intermediate outcome |
| | 1.13.2 Proportion of young people who report never having tried a cigarette (YTS, YRBSS) | Long-term outcome |
| | 1.14.1 Prevalence of tobacco use among young people (YTS, YRBSS) | Long-term outcome |
| | | |
| | 2.3.7 Level of support for creating tobacco-free policies in public places and workplaces (NATS, BRFSS) | Short-term outcome |

| Goal Area | Core Indicator and Data Source(s) | Indicator Type |
|---|---|---|
| **Goal Two: Eliminating Exposure to Secondhand Smoke** | 2.4.1 Proportion of jurisdictions [State and Local] with public policies for tobacco-free workplaces and other indoor and outdoor places (measurement protocol under development) | Short- term outcome |
| | 2.4.3 Proportion of the population that works in environments with tobacco-free policies (NATS, CPS TUS) | Short-term outcome |
| | 2.4.4 Proportion of the population reporting voluntary tobacco-free home or vehicle policies (NATS) | Short-term outcome |
| | 2.4.6 Changes in state tobacco control laws that preempt stronger local tobacco control [clean indoor air] laws (STATE) | Short-term outcome |
| | 2.6.1 Perceived compliance with tobacco-free policies in workplaces (NATS) | Intermediate outcome |
| | 2.7.1 Proportion of the population reporting exposure to secondhand smoke in the workplace (CATS) | Long-term outcome |
| | 2.7.3 Proportion of the population reporting exposure to secondhand smoke at home or in vehicles (NATS, YTS) | Long-term outcome |
| | 2.7.5 Proportion of nonsmokers reporting overall exposure to secondhand smoke (YTS; California Independent Evaluation) | Long-term outcome |
| | 2.8.1 Per capita consumption of tobacco products (STATE, State departments of revenue) | Long-term outcome |
| | | |
| **Goal Three: Promoting Quitting Among Adults and Young People** | 3.7.1 Number of callers to telephone Quitlines (Quitline call monitoring) | Short-term outcome |
| | 3.8.5 Level of support for increasing excise tax on tobacco products (NATS) | Short-term outcome |
| | 3.10.1 Proportion of insurance purchasers and payers that reimburse for tobacco cessation services (measurement protocol under development) | Short-term outcome |

| Goal Area | Core Indicator and Data Source(s) | Indicator Type |
|---|---|---|
| | 3.11.1 Proportion of adult smokers who have made a quit attempt (NATS, BRFSS, CPS TUS) | Intermediate outcome |
| | 3.11.2 Proportion of young smokers who have made a quit attempt (YTS, YRBSS) | Intermediate outcome |
| | 3.11.3 Proportion of adult and young smokers who have made a quit attempt using proven cessation methods [measure adults only] (NATS) | Intermediate outcome |
| | 3.12.1 Amount of tobacco product excise tax (STATE, CTK, State departments of revenue) | Intermediate outcome |
| | 3.13.1 Proportion of smokers who have sustained abstinence from tobacco use (NATS, BRFSS, YTS) | Long-term outcome |
| | 3.14.1 Smoking prevalence [same as 2.8.3] (NATS, BRFSS, YTS, YRBSS) | Long-term outcome |
| | 3.14.4 Per capita consumption of tobacco products (STATE, State departments of revenue) | Long-term outcome |
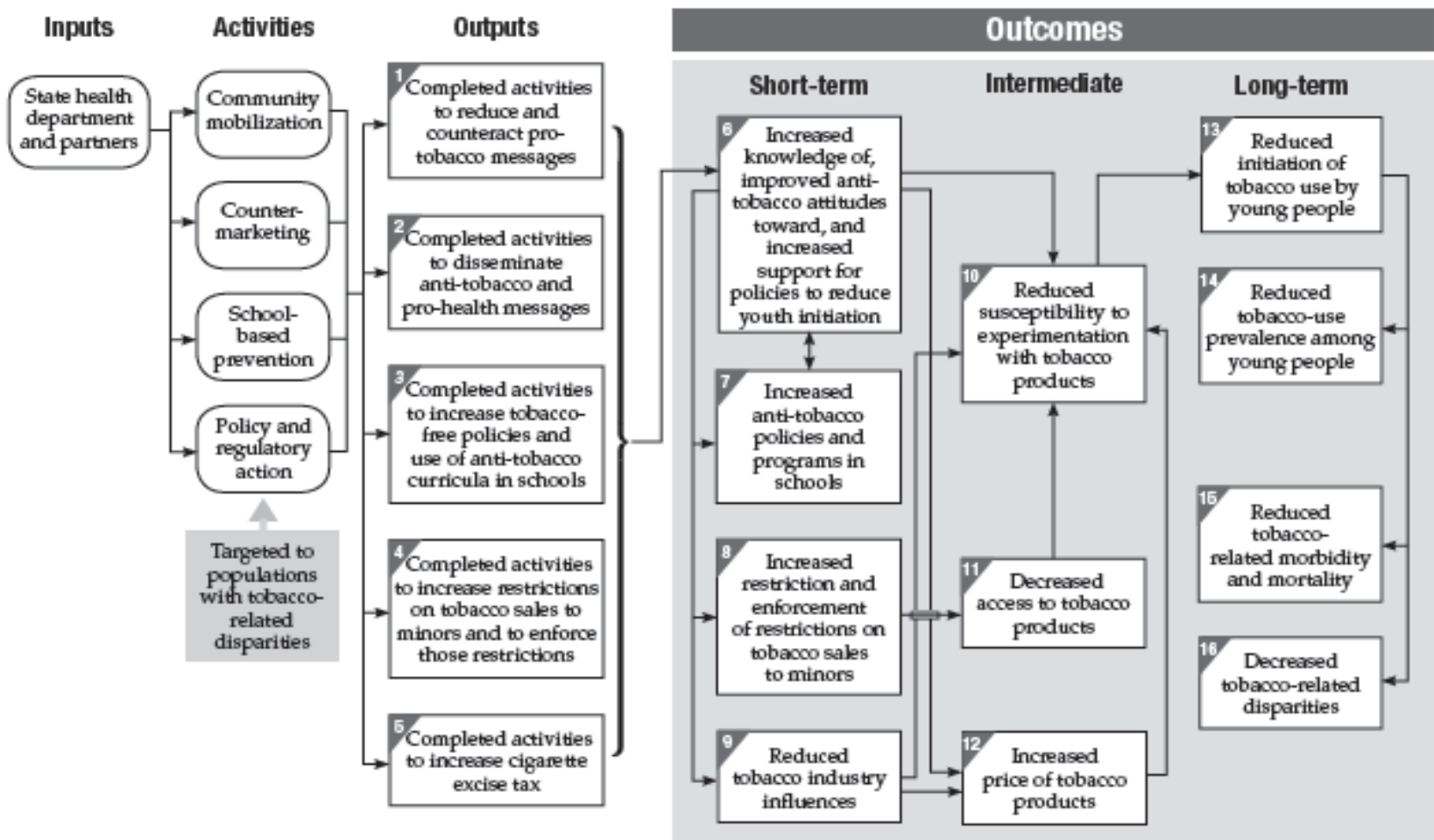
**Figure 1. NTCP Goal One Logic Model: Preventing Initiation of Tobacco Use Among Young People**
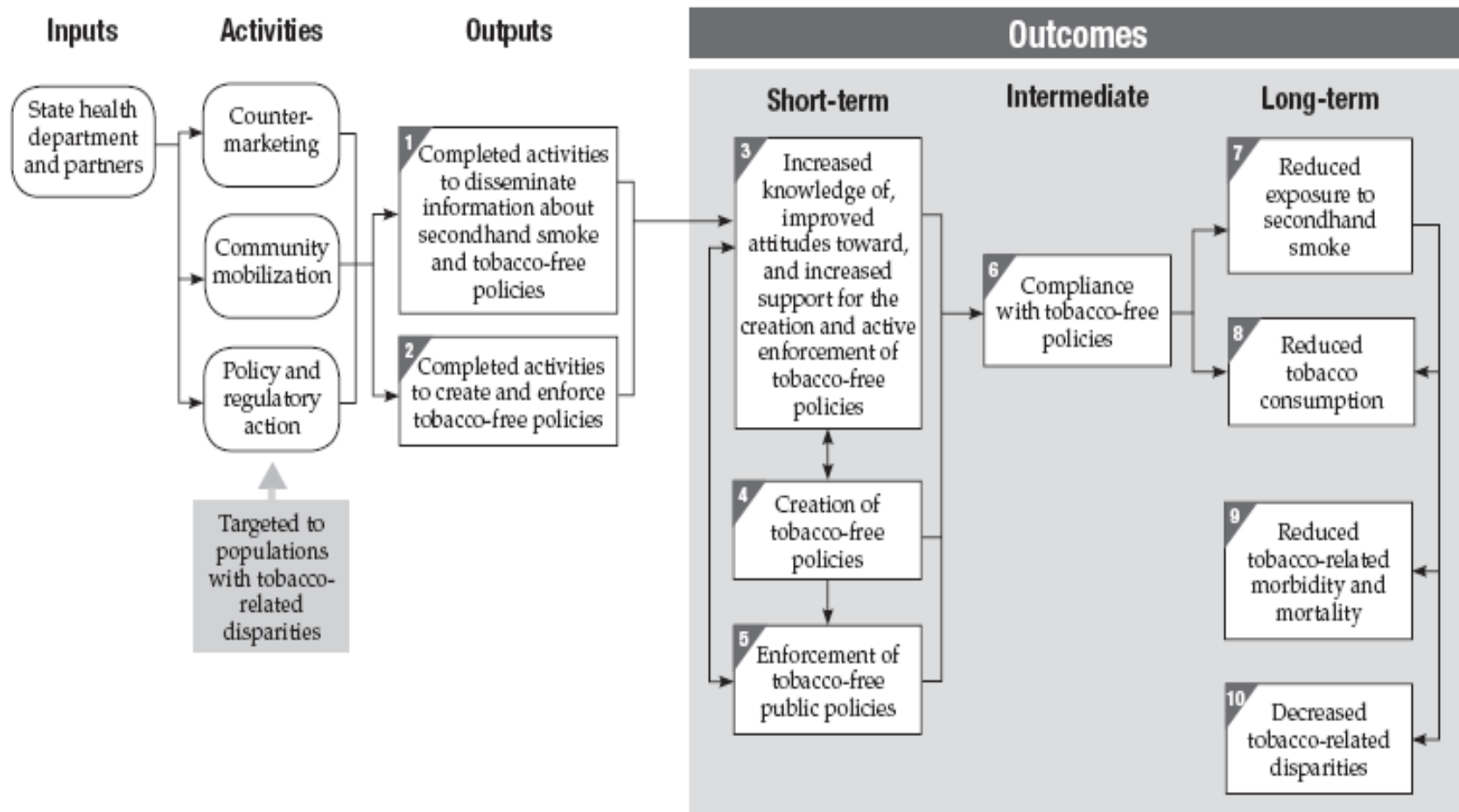
**Figure 2. NTCP Goal Two Logic Model: Eliminating Exposure to Secondhand Smoke**
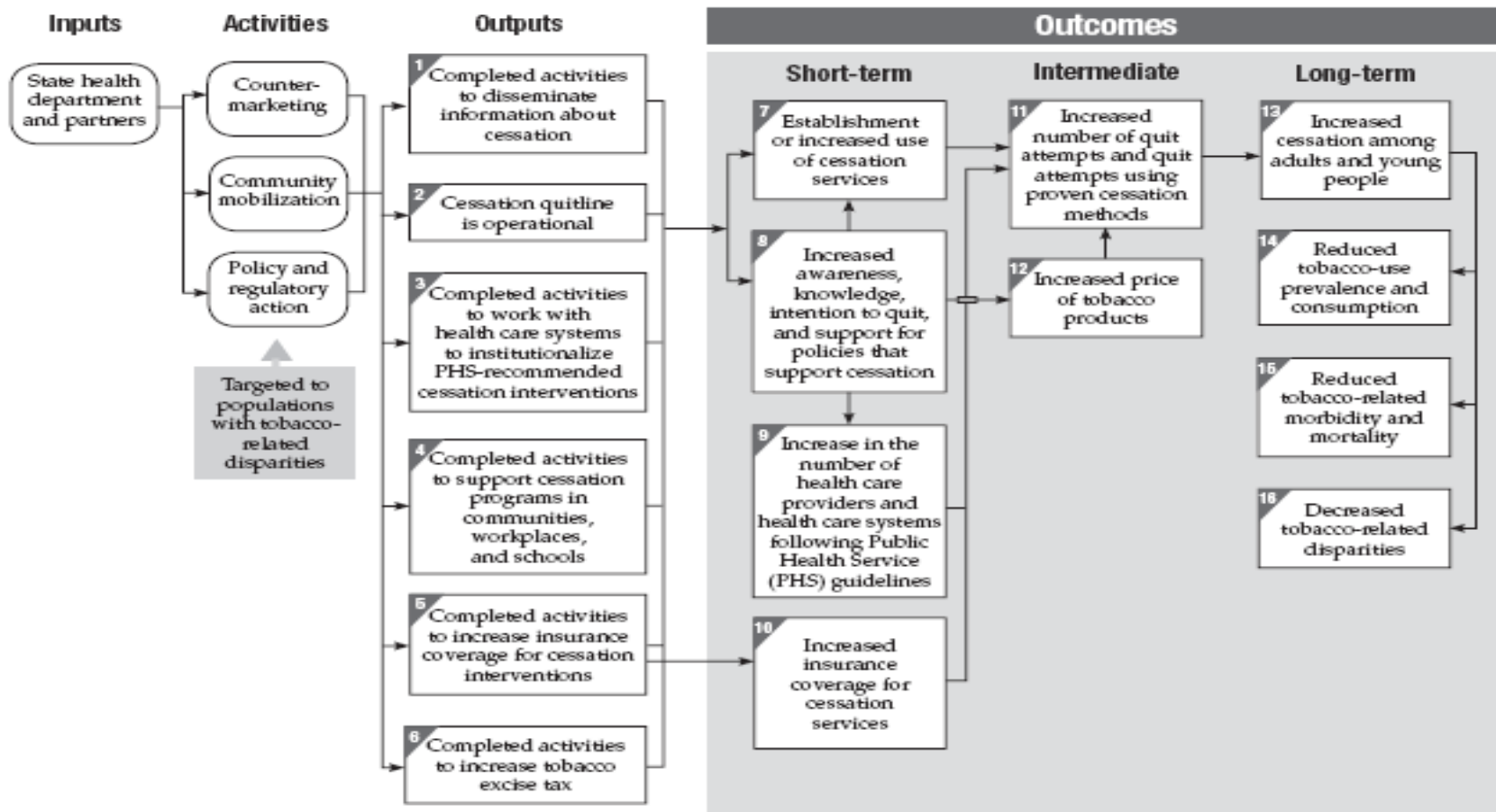
**Figure 3. NTCP Goal Three Logic Model: Promoting Quitting Among Adults and Young People**

The selection of the core indicators was dependent on the development of the KOIs. Consequently, the overall effort can be viewed as one development process and is described below. Table 24 summarizes some of the key developmental milestones.

**Table 24. NTCP Indicator Development, 1999 2008**

| Fiscal Year | System Development |
| --- | --- |
| 2001 | Initiated effort to develop key outcome indicators |
| 2002 | Convened expert panel to assist in KOI development |
| 2003 | KOIs disseminated to grantees and entered into Chronicle |
| 2005 | KOI guide published and disseminated to grantees |
| 2006 | Core indicator project launched |
| 2009 | Estimated completion of core indicators |

The development of the KOIs was led by an evaluation team within OSH's Epidemiology Branch. The team, working with contractors, began the project in 2001 initially thinking they would identify a set of 10-20 outcomes. In brief, the evaluation team worked across the three NTCP goal-area logic models, developing potential indicators to populate the boxes for short-term, intermediate, and long-term outcomes. Although the team had aimed to develop a small number of indicators, the range of the three goals areas and the variation in activities implemented by the grantees resulted in the development of a much larger pool of KOIs. Once the evaluation team identified potential indicators, they conducted a literature review to determine whether an association existed between the proposed indicator and the outcome component in the logic model. Next, potential data sources for each indicator were identified; if a data

source did not exist, the group developed example questions that could be added to existing surveys to collect the data.

OSH then convened a panel of 16 experts who represented state tobacco control programs, universities, cancer or research institutes, and national partners. The panel met in Atlanta and reviewed a total of 164 candidate performance measures, rating each on the following criteria: strength of evaluation evidence, resources needed for data collection and analysis, utility, face validity, conformity with accepted practice, uniqueness, and overall quality. After all measures were rated by the panel, the evaluation team made a number of revisions, merging some indicators, eliminating others. Next, based on advice from the panel, the evaluation team hired an independent contractor to conduct a literature review to more thoroughly assess the strength of the evidence for each indicator. The contractor reviewed nearly 850 published and unpublished articles as part of the comprehensive review process.

In 2003, OSH distributed a total of 120 KOIs to grantees, inviting them to begin using the KOIs and also to provide CDC relevant feedback on the indicators. Staff in the Program Services Branch entered all indicators into the Chronicle system. As discussed, reporting on the KOIs is not a stipulation of the NTCP funding award, but grantees are encouraged to include relevant measures as part of their semi-annual progress reporting requirement. The OSH evaluation team continued to develop the KOI guide – it was eventually published in May 2005 and includes the 120 key outcome indicators. The Guide is organized according to the three goal areas and provides a detailed profile for each indicator including an indicator rating table with assessments of the overall quality, resources needed, strength of evaluation evidence, utility, face validity, and accepted

practice – all in a "*Consumers Report*" format mentioned earlier. The profile for each measure also includes more detailed information such as a rationale for the measure, examples of data sources, other comments, and references.

Even before the KOI guide was published, the evaluation team had recognized the need for a set of core indicators that could be assessed and monitored at the national level. OSH staff realized that given the range of KOIs (120 unique indicators) and differences in resource levels and areas of programmatic emphasis across grantees, reporting on particular KOIs via the Chronicle would likely vary. Using the logic models as a guiding framework and the KOIs as the primary pool of indicators, the evaluation team initiated an effort to identify a set of core indicators. During the process, new indicators that were not part of the KOIs were also introduced for consideration. Several criteria informed the selection process. First, the group wanted to select indicators for each of the three goal areas and ones that represented each level of outcome (short-term, intermediate, long-term) in order to emphasize the causal linkages. Second, the team aimed to choose indicators that were most closely aligned with OSH priorities in order to convey national priorities and influence practice. Next, the team looked to identify indicators associated with outcomes viewed as the strongest levers of change. And fourth, selection was influenced, to some degree, by what data were available easily and quickly.

The first set of 32 core indicators was approved by Division leadership in October 2005. Since that time, the evaluation group has suffered from staff absences and some turnover. More recently, the evaluation team has focused considerable effort to finalize the core set. As reflected in table 23, most of the final 30 core indicators can be assessed using national-level survey data. Grantees can use those same data at a state-level. At the

time of data collection for this study, the evaluation team was collaborating with others in OSH to finalize a new National Adult Tobacco Survey (NATS) that will be conducted in 2010 during the second year of the new 5-year NTCP funding cycle. OSH evaluators were working to ensure that the NATS include survey questions relevant to several of the core indicators. At the same time, the team has been working to develop data collection protocols for five core measures that otherwise lack a data source (i.e., 1.8.2; 1.9.1; 1.9.10; 2.4.1; 3.10.1). Some of these indicators are particularly challenging, such as measuring the extent and type of retail tobacco advertising and promotion (1.9.1) and the number and type of MSA violations (1.9.10).

Of interest, indicators selected as core are fairly consistent with the most frequent KOIs voluntarily reported by grantees in the Chronicle. Although such a correlation was not a criterion for selecting the core indicators, OSH staff are relieved that the core indicators will not represent a major departure from what grantees are already reporting.

7.1.3.2 Performance Measurement System Design

In this section, the design of the NTCP performance measurement system is described highlighting the following: purpose of system; level of measurement; types of measures; use of targets or standards; and quality assurance efforts. Table 25 summarizes these design features.

**Table 25. Design Features of the NTCP Core Outcome Indicator System**

| Design Feature | NTCP Core Indicators |
|---|---|
| Purpose of system | Accountability |
| Level of measurement | Primarily state but some local level |
| Type(s) of performance measures | Short, Intermediate, and Long-term Outcomes |
| Use of targets | Set by grantees |
| Quality assurance efforts | Use of national survey data, provision of data collection protocols for some measures |
| Use of performance data | Core indicators not yet implemented |

7.1.3.2.1 Purpose

With the increase in tobacco control resources in the late 1990s and the resulting influx of settlement funds, OSH evaluators recognized the importance of developing outcome measures to address accountability demands at both the federal and state levels. More recent cuts in state tobacco control resources have underscored the importance of outcome measurement as a means to effectively defend resource allocations. At the federal level, OSH wants to provide a "national picture" of the NTCP for HHS, the Office of Management and Budget (OMB), and others "above" CDC. Participants also suggested that the core indicators will be used for monitoring purposes to assess the NTCP at both the state level and national level.

7.1.3.2.2 Level of Measurement

Most of the core indicators reflect state-level measures and are derived from survey data including CDC's youth tobacco survey, CDC's behavioral risk factor surveillance survey, the current population survey, and the forthcoming NATS. National monitoring systems such as one maintained by the Campaign for Tobacco-Free Kids and CDC's State Tobacco Activities Tracking System also serve as data sources for the core indicators. As discussed, a small subset of measures reflects more complex requirements for local-level data collection (e.g., extent and type of retail tobacco advertising and promotions).

7.1.3.2.3 Types of Measures

The choice by OSH evaluators to focus on outcomes was influenced, in large part, by political demands for accountability that emphasized outcomes rather than process-level measures. The core indicators reflect three levels of outcomes: short-term, intermediate, and long-term. In particular, the evaluators wanted to identify outcomes where change could be detected within five years or less, rather than longer-term outcomes that might take ten or more years to achieve. Given the focus on outcomes, OSH recognizes that results reflect the combined work and resources of many agencies and organizations involved in the comprehensive tobacco control effort. In other words, the outcomes reflect effects of the overall comprehensive tobacco control programs, not simply the CDC-funded efforts.

7.1.3.2.4 Use of Targets or Standards

The variability in grantee resources and capacity make it difficult to establish national targets. In addition, national targets for specific measures such as smoking prevalence are challenged by variable prevalence rates across states. Currently, grantees specify the goal area, the objective, a baseline, proposed target, and target date when reporting on a KOI in the Chronicle. Evaluation team members anticipate that grantees will do the same for the core indicators once the measures are finalized, although revisions to the Chronicle may have implications for reporting.

7.1.3.2.5 Quality Assurance Efforts

A majority of the core indicators will be calculated based on national survey data (or, for grantees, calculated based on state survey data). Consequently, the validity and reliability of those data are well established. Appropriate sample sizes are determined for both national and state-level assessments ensuring a representative sample size. But while there is confidence in the quality of the data, problems have persisted in its reporting, at least as it pertains to the KOIs. In particular, OSH has been challenged by grantees' use of "customized" measures rather than the KOIs as defined in the guide. The Chronicle allows for grantees to submit customized measures, but the feature has been misused to an extent. Grantees have tended to create custom indicators that are similar but slightly different from the KOI as defined in the guide. OSH conducted an assessment of the KOIs reported in the Chronicle for the period 2007-2008 and found that 38% of the indicators reported were categorized as "customized," and, of these, 56% were similar or identical to the evidence-based indicator defined in the guide. Consequently, OSH has

been limited in its use of the KOI indicator data that has been reported through the Chronicle, especially in aggregating data across states.

As discussed, the evaluation team is working to define data collection and measurement specifications for five of the core indicators that are more challenging to measure (e.g., extent and type of retail tobacco advertising and promotions). The team is developing strict data collection and reporting protocols for grantees to follow (e.g., sampling guidelines, observation instructions). But these measures are likely to be especially vulnerable to data quality issues given that the data will be collected at the local level by staff or volunteers of varied expertise.

7.1.3.3 Use of the Performance Measurement System and Data

The KOIs have been used by the grantees, even if in a "customized" fashion. The assessment recently conducted found that all 50 states and D.C. reported on at least one KOI and most (n=44) are using at least one KOI in all three goal areas. Of the 120 total indicators, 106 had been used by at least one state since they were introduced. Other than the *reporting* of the KOIs to CDC, grantees' use of the indicators is not known. At the time of data collection for this study, the evaluation team was planning a qualitative assessment of the use of the KOIs with a sample of nine states. Many participants spoke of the value of the KOI guide and its use by grantees for program planning, in particular. Program consultants are especially enthusiastic about the guide's value and promote it extensively among their grantees. In particular, the logic models are viewed as useful in helping grantees understand the causal pathways necessary to achieving longer-term outcomes. Dog-eared copies of the guide were evident in several offices.

In regard to the core indicators, CDC anticipates using these to characterize the national program and help defend resource allocations for the national program. Beyond this, participants expressed differing views about their use. Some interviewed suggested the core indicators are better defined as part of an evaluation system while others suggested they were part of a performance monitoring system. Reporting the core indicators will remain voluntary under the forthcoming five-year program announcement. The infrequency of data collection inherent in many of the surveys challenges the core indicators' use for on-going data monitoring, and OSH has not developed a performance management system (other than functions within the Chronicle) to support data use for on-going performance measurement. At the state level, grantees may not have the resources to participate in some of the surveys that provide needed indicator data.

## 7.2 Findings

### 7.2.1 Network consensus on goals, a strong evidence base, and extensive survey data facilitate identification of outcome measures for comprehensive tobacco control.

Tobacco control efforts in the U.S. involve a network of agencies and organizations that implement a mix of educational, clinical, regulatory, economic, and social strategies. These efforts are supported by diverse sources of funding from federal and state government as well as business and private foundations. State excise taxes on tobacco products and settlement funds from legal victories over the tobacco industry represent a significant investment in tobacco control efforts in some states, dwarfing federal funds provided by CDC.

The NTCP was established to encourage coordinated, nationwide tobacco control activities, and, as noted in the case description, CDC funds are primarily used by grantees

to support program infrastructure costs, coalition building, statewide planning, and

surveillance and evaluation activities. One person said,

> For most states, we're [CDC] the backbone of their tobacco program and
> we're okay with the fact that the whole program is funded by *much* more
> than us. I mean we're just that much [indicating a small amount with her
> fingers] of a lot of states in dollars, but they couldn't get those other funds
> without ours.

CDC funds are often used by the NTCP grantees to leverage other resources that are

sometimes more restricted in their use (e.g., required to be distributed to local levels).

Consequently, resources for tobacco control are frequently integrated at the state health

department to support a range of activities and often complimented by initiatives

supported by other organizations.

In 2001, OSH embarked on their effort to develop program performance measures

so that they could better characterize the NTCP nationally for accountability purposes.

The political climate and influx of funding in the late 1990s proved an incentive for OSH

to identify indicators of performance that would help defend resource allocations at both

the federal and state levels. Given the emphasis on accountability for results, OSH

wanted to identify outcome measures in particular. One person said about the focus on

outcomes,

> That was a conscientious decision and we recognized, because
> accountability was driving it, because outcomes are what people want
> from an accountability standpoint, that that was the most important area.
> So we focused on the outcomes piece and we did not focus on the process
> piece initially.

Another participant commented,

> The rationale [to focus on outcomes] was a very deliberate decision.
> …And the reason the focus was on outcomes was because OSH felt that
> states needed to demonstrate progress. They needed to, for political

purposes, for using MSA dollars and other tax dollars – we really wanted to invest in keeping an eye on the outcomes.

The process to develop outcome measures for the NTCP, led by evaluators in OSH, has been facilitated by three factors: network consensus around program goals; a strong evidence base; and the availability of extensive survey data.

As noted above and detailed in the case description, a network of agencies and organizations are involved in planning, developing, and implementing tobacco control efforts at the federal, state, and local levels. Leaders in the field of tobacco control at all levels have adopted four program goals to direct tobacco control efforts: prevent the initiation of tobacco use among young people, eliminate nonsmokers' exposure to secondhand smoke, promote quitting among adults and young people, and identify and eliminate tobacco-related disparities. The four goals provide an organizing framework from which to structure coordinated and complimentary interventions by varied network members in individual states and territories. The network consensus around these four goals has aided OSH in identifying program outcomes for which all network members are committed and has supported the development of related performance measures.

Building on the four goals for tobacco control, OSH evaluators, in collaboration with network stakeholders, have developed detailed logic models for three of the four program goals that explicate relationships between program inputs, activities, and outcomes (short-term, intermediate, and long-term). Construction of the logic models has been aided by a strong science base, the result of a significant investment in research and evaluation efforts over the past decades. In particular, OSH has developed causal pathways for each goal and identified 120 key outcome indicators along those pathways

that can be measured to help assess progress in ultimately achieving population-level

affects on tobacco-related morbidity and mortality. One participant said,

> We know what works in tobacco control, and that's where the key
> outcome indicators are so helpful. It [the KOI guide] focuses on what data
> sources you need and which of the key outcomes are a little bit more
> effective. So states can look through [the guide] and decide which
> outcomes best meet their needs for what they're doing. But then it
> provides, like I said, some consistency and some congruency with what
> folks are doing.

Lastly, the development of outcome indicators for the NTCP has been supported

by the availability of extensive survey data. In many ways, OSH is "data rich" compared

to other public health programs. As indicated in table 23, survey data is used to calculate

most of the proposed core indicators at both the state and national levels (e.g., YTS,

YRBSS, BRFSS, NATS, and STATE). Consequently, OSH can calculate many of these

measures at the national level independent of grantees' reporting of data. Not all grantees

have the resources to support state-level data collection for every survey. One person

working on the core indicators said,

> In fact, we created a table that took the core indicators and allocated them
> to states and to OSH [in regard to data collection and analysis]. There's a
> bunch of indicators that OSH can gather, even at the state-specific level,
> without asking the states to do anything. And with the measurement
> surveillance systems that are in existence, even more so now that NATS is
> coming down, the National Adult Tobacco Survey. They've asked me to
> cross check NATS drafts with the core indicators specifically. And there's
> a large percentage of them that will be captured by NATS that will be a
> method to capture a large percentage of core indicators.

**7.2.2 The network implementation structure of the NTCP leads to a joint production of outcomes and shared accountability.**

As described, tobacco control is implemented by a network of agencies at multiple levels. Vertical relationships, primarily supported by funding arrangements, assure the involvement of federal, state, and local health agencies as well as other local-level organizations. Populating each level are horizontal partnerships representing government, nonprofits, academic institutions, health care, and commerce. Working with a wide range of these partners, NTCP grantees frequently lead or participate in coalition efforts to develop statewide plans for tobacco control that are implemented by a host of agencies. For instance, in addressing the goal related to the prevention of tobacco use among youth, the National Legacy Foundation might contract with a marketing firm to develop a large scale, state-wide, counter marketing campaign. The state department of education might endorse an anti-smoking school-based curricula or tobacco-free policies for schools. The state health department might fund CBOs at local levels across the state to implement anti-smoking interventions for youth. And advocacy groups might introduce legislation to strengthen tobacco control laws in ways that protect young people. One participant provided this example,

> TFL [Tobacco Free Living] has the money for paid media – the state
> doesn't, because basically their only funding is from CDC. So in the
> implementation of the smoke free air law education campaign, TFL took
> the lead. They had the money. They had the expertise on site. The state
> would compliment that by looking at activating local mini-grants to get
> earned media at the local level to support the paid media that was going
> on. So again, that's leveraging, you're linking.

This comprehensive approach is intended to optimize synergies across network contributors in order to maximize program effects. Ultimately, program outcomes reflect

297

the contributions of the many. In other words, program outcomes are jointly produced by network partners. This poses a potential concern for OSH's venture to develop a set of core outcome indicators. Specifically, accountability is less easily appropriated to one agency or another – instead, accountability for outcomes in comprehensive tobacco control efforts may best be characterized as "shared accountability." One participant described earlier attempts to try and disentangle each funder's contributions,

> We tried that back in the early '90s when there weren't as many players in the game and you still couldn't do it. I mean, you had NCI who was doing the $17 million dollars with the ASSIST program, you had our $5 million, and Robert Wood Johnson's smokeless states that were probably in there for about $3 or $4 million. Even with that small amount of resources, you still couldn't ferret out the who's who, and what money did what, and it started driving the states crazy. And it's like, we can play this game or we can actually do public health and have something happen. So we all just said look, we're going to step back and evaluate these things in terms of 'what are effective interventions in the direction we have to go as a *tobacco control community'*.  All the major funders came to that agreement, because we were driving ourselves crazy, our grantees crazy, and we were detracting time from actually doing the real work.

Therefore, for their measurement system, OSH has accepted what some view as a limitation and others recognize as an unavoidable consequence of the networked environment. That is, the inherent difficulty of assigning accountability for results when outcomes follow from multiple and diverse interventions funded and implemented by a network of collaborating institutions and agencies. OSH's answer to this dilemma of outcomes, accountability, and networks is to openly acknowledge that the NTCP core outcome indicators more accurately represent outcomes for comprehensive tobacco control, not simply the CDC-funded program.

**7.2.3 Network variability limits types and choice of NTCP core measures and their use.**

NTCP's extensive network implementation structure has important implications for the design of the core indicators. In particular, grantee variability in terms of state or territorial context, tobacco control priorities, implementation activities, resources, and capacity limits the types, choice, and use of core measures. One person described the diversity across grantees as follows,

> You have California that was already allocating $100 million [of state funds] back in the '90s, so you have that wide range [of resources across grantees]. And there's a level of sophistication in the programs – you have some programs that, as I like to put it, are really just doing the basics. They're trying to change social norms, they're really trying to get the word out about why tobacco is harmful, about how the [tobacco] industry manipulates you, trying to get local ordinances passed, and that's where they are. Every few years or so, when there's a perfect storm of opportunity that brews up, they'll try to get an increase in their excise tax on tobacco products. But for the most part, they're just chugging along trying to get all their school districts to be tobacco-free, trying to get local communities to have secondhand smoke ordinances, and those types of things. Then you go and you find other states that are really progressive and they're out there working with the tobacco advocates, beginning to work with the Medicaid folks, and are really zeroing in on which [segment] of the low income population is still using tobacco during pregnancy and developing special interventions and programs for them. So you have other states that are really cutting edge and looking at how they can use internet blogging and those types of communication avenues for cessation programs. So it runs the gambit.

That kind of variability, as well as the breadth of four NTCP goal areas, is what led to the initial development of 164 proposed key outcome indicators. Although OSH evaluators had originally intended to identify a small set of outcome indicators for accountability purposes, the number of indicators quickly expanded as reflected in the statement below.

It was [person's name] that had actually come up with the idea of 'let's come up with a list of 20 or 30 indicators that we think are important for states to measure.' And that grew into, like 200 indicators, because they're all important, depending on what the state is doing. And that's where you get into, you have 50 different states and 50 different programs, and depending on what the state's working on, they're going to want to measure different things.

"What the state's working on" often depends on resources and the selection of priorities consistent with the organizational and political context of a given grantee. While one grantee may prioritize policy efforts to increase smoking bans in bars and restaurants, another grantee may face a powerful and well-funded lobby that dissuades it from addressing the issue.

It was also variability that tabled the effort to develop a similar set of process measures to accompany the KOIs. OSH evaluators found it impossible to develop a common set of process measures given that grantees conducted activities in so many different ways. One participant said,

I went in and told them [Division and Center leadership] that basically, if you want indicators I'll give you indicators. But if you want *useful* indicators, I can't give you useful indicators even after 3 years [of trying to develop process measures]…I mean you've got *Best Practices*, you've got the *Community Guide*,[53] but you've got all that stuff being implemented 14 thousand different ways. And so to try to come up with useful [process] indicators that would actually contribute to program improvement, or give a better understanding of what contributed to the outcomes, we just basically said 'no, we can't do that.'

Therefore, the *type* of measures to include in the core indicators was limited to outcome indicators, although, as discussed earlier, OSH evaluators wanted to identify outcomes to address accountability. The paradox is, however, that outcome-level

---

[53] *Best Practices* and *the Community Guide* both provide listings and descriptions of evidence-based interventions.

measures typically represent the work of many, not simply CDC. In other words, accountability becomes fragmented in the network context and outcomes represent results of multiple network partners.

The *choice* of measures was also affected by network variability. While variability in program activity is what led to developing 120 key outcome indicators, the evaluation team has the difficult task of narrowing these to a smaller set of core measures appropriate for 58 grantees. As one person said, "We've come full circle back to the core indicator idea!" Evaluators also considered data availability in their choice of core measures. Resource constraints limit some grantees from participating in some surveys – especially those that require funds to support specific tobacco-related modules. One person said,

> I think in large part it's about resources. You're at the state and you can collect data, and you know you need your prevalence data – you need it for youth, you need it for adults, you need this, and you need that. All these things cost money. How much of your budget can go into collecting these data and how many of these data? You have to kind of pick and choose. It's a resource issue – you have to hire people who are doing the data collection, analyzing the data, and how many people can you hire?

Consequently, data availability was a consideration in selecting core measures. A participant described it like this,

> Even if you just take the low hanging fruit, the 'what's available' approach. We at CDC are already going to have data on a number of them (indicators). So we can think, 'okay well, there's a core set' – just low hanging fruit, what's available.

With so many grantees, OSH evaluators also considered the effects that the core measures might have across the national programs and their acceptability for implementation.

With the exception of the kind of whacky core indicators, the ones that we're having measurement issues with and that we have to define, it's not as if they're [grantees] going to have a big program shift and do anything differently than they're already doing. For the most part, they just have to keep doing what they're doing. Which is great because we were worried about that – we thought 'we don't want to do anything radically different, and say okay, we're going to do *this* in tobacco control now instead of this' We're adding a couple of new things like Master Settlement violations and the retail observations that are going to be new and different for most states, but other than adding those couple of things, everything else is pretty standard.

Finally, the selection of outcome-level indicators that are primarily derived from periodic survey data may limit their utility as part of a performance measurement system. Much of the needed survey data to support the core measures are collected once every two to three years – at this time, the NATS is funded for only one year of data collection. Consequently, the core indicators may better be used for program evaluation than as part of a performance management system that supports on-going program monitoring.

### 7.3 Summary

At the time of data collection for this study, OSH was finalizing its development of a set of core outcome indicators for its NTCP. Consequently, findings for this case focus on OSH's development process rather than implementation of the measures. Despite the NTCP's vast network that includes extensive variation across grantees in their level of resources, goal priorities, and specific intervention efforts, OSH has been able to identify a set of core outcome indicators for NTCP. The consensus on and commitment to four program goals by the broad group of stakeholders within the network has facilitated this effort as has the availability of a rich evidence-base and extensive survey data. And while the network structure leads to the joint production of many

program outcomes, OSH has determined that it would be too costly and burdensome to conduct the level of evaluation needed to decipher the unique contribution of individual network members in 58 different grantee states and territories. Consequently, OSH and its partners have accepted that outcome-based performance measures will typically reflect the overall efforts of the tobacco control *community* as well as other factors.

In addition, the variability across grantees in context, priorities, intervention activities, resources, and capacity has important design consequences for selecting NTCP's core indicators. Type of measure was limited to outcomes given that process measures proved impossible to identify at a national level. Choice of measures was impacted by resources, data availability, and an interest to ensure their acceptability across grantees. And use of core indicator data may more accurately support program evaluation efforts than a performance measurement system.

# CHAPTER 8

## CROSS-CASE RESULTS

### 8.1 Introduction

While the previous four chapters summarized findings from the within-case analysis for each program, this chapter presents results from the cross-case analysis. As detailed in chapter 3, this study applies an instrumental case study approach. In contrast to intrinsic case study that strives to understand a particular case, instrumental case study is meant to achieve greater comprehension of a particular phenomenon – the case is used to gain broader insight (Stake 2006). By including multiple cases in an instrumental case study, the phenomenon is examined in several contexts to achieve a more nuanced understanding. For the present study, the researcher has examined four unique cases to investigate the implications of networked public management on the design, implementation, and utilization of federal-level performance measurement systems. Both the unique situational analysis of each case presented in chapters 4-7, and the cross-case analysis presented here contribute to a deeper understanding of performance measurement as applied in public health programs implemented in networked contexts. In particular, greater insight is achieved by studying the similarities and differences observed across the four cases in how networks affect the use, design, and impact of performance measurement.

This chapter begins with a descriptive summary of the four cases using the typology that was applied for the case descriptions in the individual case chapters (i.e., the program, network, and performance measurement system). Next, the cross-case results are presented, organized around the three research questions. As detailed in

Chapter 1, Jennings and Haist (2004) propose a set of 25 hypotheses intended to explicate

assumptions related to the impact of performance measurement (appendix A). More

specifically, Jennings and Haist organize their hypotheses according to five unique

groups including incentives and accountability (nine hypotheses), organizational

characteristics (eight hypotheses), political context (three hypotheses), agency type (three

hypotheses), and leadership (two hypotheses).

The three research questions guiding this study were derived from two of the

three hypotheses related to agency type. Jennings and Haist built these hypotheses from

James Q. Wilson's (1989) typology of four agency types: production, procedural, craft,

and coping agencies. Wilson's typology incorporates two dimensions, one related to

whether outputs are observable and another to whether outcomes are observable.

Jennings and Haist define observability in two ways that relate to performance

measurement – the ability to *measure* outputs and outcomes and the ability to *attribute*

outcomes to the program of interest (see Chapter 1, section 1.4). The two hypotheses of

Jennings and Haist and the related research questions for this study are summarized in

table 26 below.

**Table 26. Study Research Questions Related to Jennings and Haist's Hypotheses of Agency Type**

| Jennings and Haist Hypothesis | Study Research Questions |
|---|---|
| Hypothesis #13<br><br>"The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed." (p.185) | Research Question #1.<br><br>How does networked public management affect the observability of CDC program outputs and outcomes? |
| | Research Question #2.<br><br>How does networked public management influence CDC's use of performance measurement and the types of performance measures used? |
| Hypothesis #14<br><br>"Measurement will be more common and will have greater impact when agencies have greater control over outcomes." (p.185) | Research Question #3.<br><br>How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? |

As detailed in Chapter 3, section 3.8.5, the researcher has used various data displays to facilitate analysis and the development of individual case findings and cross-case findings. First, a set of matrices was created summarizing the descriptive characteristics of each case using a standard typology that was devised based on the literature review and inductive case analysis. These matrices include one for program characteristics (appendix I), network characteristics (appendix J), and the performance measurement systems (appendix K). Next, tentative findings were developed for each case based on the individual research questions (appendix L). From these tentative, case-specific results, findings were constructed for each case – these are summarized in appendix M.

Finally, to aid the development of cross-case results, a matrix was developed to relate the hypotheses of Jennings and Haist with relevant research questions and individual case findings. This matrix includes information about the effect of each finding on the performance measurement system for the specific program, the broader implications of each finding for performance measurement, evidence for each finding from the specific case, information about the case context that is relevant to the finding, and details from the other cases that support or counter the finding. This matrix is included in appendix N.

To aid the reader, acronyms are only used when referring to CDC and GPRA/PART. Given that all four cases are discussed in this chapter, Division and other organization names, laws and policies, and other terms previously referenced with abbreviations are spelled out in full to limit any confusion. The exception to this is in table 27 where a list of stakeholders for each case is detailed – for these, please refer to the list of acronyms provided at the beginning of the dissertation before the start of chapter 1.  In table 27, below, a key is provided clarifying how the four cases will be identified throughout chapters 8 and 9. Finally, the terms "program" and "case" are used interchangeably and represent the national programs included in the study.

**Table 27. Key to Case Names Used in Chapters 8 & 9**

| Full Case Name | CDC Division and Office/Center | Acronyms Used in Chapters 4-7 | Abbreviated Name for Chapters 8 & 9 |
|---|---|---|---|
| Public Health Emergency Preparedness | Division of State and Local Readiness, Coordinating Office for Terrorism Preparedness and Emergency Response | PHEP | Preparedness Program |
| Comprehensive STD Prevention Services | Division of STD Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention | CSPS | STD Program |
| National Breast and Cervical Cancer Early Detection Program | Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion | NBCCEDP | Cancer Detection Program |
| National Tobacco Control Program | Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion | NTCP | Tobacco Program |

## 8.2 Descriptive Summary of Four Cases

### 8.2.1 Program Characteristics

Several program characteristics were identified from inductive analysis that describe the four cases including their organizational context, program goals, stage of program development, budget, stakeholders, and political context. Table 28 summarizes each of these characteristics; appendix I provides a more detailed synthesis.

**Table 28. Summary of Key Program Characteristics by Case**

| Preparedness Program | STD Program | Cancer Detection Program | Tobacco Control Program |
|---|---|---|---|
| **Organizational Context** | | | |
| 62 grantees – states, cities, territories | 65 grantees – states, cities, territories | 68 grantees – states, D.C., tribes, territories | 58 grantees – states, D.C., territories |
| Politically-driven culture | Change difficult | Data-driven culture | Science-based |
| **Program Goals** | | | |
| Evolving goals over time; developing program framework | Lacking clear goals from CDC | Goal consensus among CDC and grantees | Goal consensus among CDC and grantees |
| Outcomes difficult to define ("preparing," "preventing") | Focus on 3 of 25 STDs: syphilis, gonorrhea, and chlamydia | Focus on clinical outcomes for women screened | Focus on evidence-based outcomes for three of four goal areas |
| **Stage of Program Development** | | | |
| New | Mature | Mature | Mature |
| **FY 2008 Budget** | | | |
| $700 million | $104 million | $157 million | $66 million |
| Median award/grantee: $8.9 million | Median award/grantee: $1.1 million | Median award/grantee: $2.1 million | Median award/grantee: $1.2 million |
| **Stakeholders** | | | |
| DHS, HHS, ASPR, FEMA, ASTHO's Directors of Public Health Preparedness, NACCHO, CSTE, APHL, grantees, | NCSD, grantees, other CDC programs, OMB, Congress | National organizations and advocates, NACDD, NACCHO, NBCCEDP Council, grantees, other CDC programs, OMB, | NCI, SAMHSA, NIH, National organizations (ACS, CTFK), NACDD, grantees, other CDC programs |

| CDC Director, OMB, Congress | | Congress | |
|---|---|---|---|
| Political Context | | | |
| Politically visible and volatile | Stable | Politically visible given strong advocacy community | Politically visible given strong advocacy community and adversary |

## 8.2.1.1 Organizational Context

Each case has a unique organizational context that affects the character of the program. The Preparedness Program is situated in a highly politicized context influenced by public concern about terrorist threats and a steady influx of policy initiatives and directives from federal partners in Washington D.C. – Health and Human Services, Assistant Secretary for Preparedness and Response, Department of Defense, Department of Homeland Security, and others. Since its inception in the late 1999, the program has expanded in scope from a focus on terrorism to "all hazards," responsible for the public health in the case of events as diverse as hurricanes and anthrax attacks. The program exploded in size after September 11th and the 2001 anthrax attacks from a $45 million dollar program in fiscal year 2001 to almost $1 billion in fiscal year 2002. The Preparedness Program remains the largest funded single program at CDC. Several participants referred to Preparedness Program's organizational culture as the "wild, wild west" reflecting its chaotic, dynamic nature, but also recognizing that preparedness is an emerging field in public health that is blazing new paths.

In stark contrast to the Preparedness Program, the STD Program is largely defined by its status as one of the longest-standing programs at CDC dating back to 1957. CDC staff working with the STD Program often initiated their careers as federal assignees to state-based programs conducting disease investigation work. Over time, these staff typically advanced from front-line positions to state-based management positions and eventually to CDC headquarters in Atlanta. Consequently, many of the CDC staff working with the STD Program have decades of experience with the program. While the program benefits from well-tenured and experienced staff, the organizational culture is viewed as somewhat entrenched and hard to change, making the adoption of new ideas difficult.

The Cancer Detection Program was established by law in 1991 and is unique in the fact that it is a service delivery program. The organizational context for the Cancer Detection Program can best be described as data-driven. The program benefits from a long-standing data management system that supports a large set of clinical and data quality variables. Data are used extensively to manage the Cancer Detection Program and ensure that women receive quality care.

The Tobacco Control Program was also initiated in the early 1990s, although under a different name. The program benefits from a robust evidence-base reflecting years of well-supported research efforts. The organizational culture for the Tobacco Control Program can be characterized both by its strong science-base and also its collaborative nature – Tobacco Control Program staff at CDC work closely with many other national partners on comprehensive tobacco control efforts.

Each of the four cases included in the study support a large number of grantees, typically representing government health agencies in states, cities, territories, and tribes. The Tobacco Control Program has the smallest number of grantees with 58 while the Cancer Detection Program supports the largest with 68; all represent broadly decentralized programs managed by CDC, a point discussed further in section 8.2.2 below.

8.2.1.2 Program Goals

Preparedness, STD prevention, breast and cervical cancer screening, and tobacco control – four very different programs representing unique sub-fields in public health: public health preparedness, infectious diseases, and chronic diseases. As noted above, public health preparedness represents a complex area addressing hazards as diverse as hurricanes, bioterrorist attacks, and pandemic influenza. States vary in their risk for a given event and in the potential scale of an event. As detailed in chapter 4, the Division of State and Local Readiness has struggled to define what "preparedness" means and how it can be represented in terms of public health outcomes for the Preparedness Program given the infancy of the field, the science base for public health preparedness is thin. In collaboration with experts from across the country, the Division of State and Local Readiness is working to develop a programmatic framework for the Preparedness Program based on five key capabilities (i.e., incident management, risk communications, biosurveillance, countermeasure distribution, isolation and quarantine/community containment). Program goals for the Preparedness Program are further shaped by its dominant political context. Top-down policy initiatives from the Assistant Secretary for

Preparedness and Response, Department of Homeland Security, and others dictate requirements that may not be consistent with expert opinion or programmatic realities.

The Division of STD Prevention has struggled to define and communicate national-level goals for the STD Program even though it is one of the oldest programs at CDC. There are over 25 different sexually transmitted diseases – the Division of STD Prevention focuses on three of them: gonorrhea, chlamydia, and syphilis. In 2004, the Division of STD Prevention began promoting a broader "community perspective" for the STD Program that emphasizes the need to engage and influence community-level partners in order to expand the program's reach, leverage resources, and extend program effects. In part, this transformation has been in response to stagnant resource levels, changes in health care access and delivery, and a desire to more effectively leverage population-level impact. The "community perspective" requires a paradigm shift from a more narrow focus on the sexually transmitted diseases clinics, traditionally funded through the STD Program, to one that assumes broader responsibility for sexually transmitted disease control and prevention in a community or state. The change in perspective represents an adjustment in program scope that has implications for program priorities and goals. However, as noted above, the organizational culture is somewhat resistant to change and the STD Program currently suffers from a lack of consensus about the program's national priorities.

The Cancer Detection Program is guided by its authorizing legislation. At a patient-level, clinical outcomes are straight-forward and can be easily assessed given an adequate data collection and reporting system. These outcomes (e.g., timely diagnostic follow-up, timely initiation of cancer treatment) are consistent with the overall program

goals which are widely accepted across CDC, the grantees, and individual providers within the implementation network.

As described in chapter 7, the Tobacco Control Program is strongly rooted around its four national goals (e.g., prevent the initiation of tobacco use among young people, eliminate nonsmokers' exposure to secondhand smoke). These goals are broad in scope and intended to direct efforts of the larger community of stakeholders involved in comprehensive tobacco control. The Office on Smoking and Health emphasizes that achieving the four Tobacco Control Program goals requires collaboration "across the federal government, across the nation, and within each state." Consensus around program goals and the strong science base has enabled the development of detailed logic models for three of its four goals specifying outcomes at multiple levels.

8.2.1.3 Stage of Program Development and Budget

Three of the four programs can be characterized as relatively mature; the Preparedness Program, however, has rapidly expanded in scope since its inception in 1999 and can therefore be considered less established. For fiscal year 2008, program budgets ranged from roughly $66 million for the Tobacco Control Program to over $700 million for the Preparedness Program whose budget is several times the size of the other three cases. Similarly, median grantee awards range from $1.1 million for the STD Program to $8.9 million for the Preparedness Program. All four programs have faced relatively flat or decreasing program funding over the past several years, and CDC staff uniformly expressed concern about the need to defend current allocations given the challenging economic climate. Many interviewed described a political climate that

stresses accountability, and they view performance measurement as an important means to provide data needed to justify their programs.

8.2.1.4 Stakeholders and Political Context

All four cases involve an extended network of stakeholders at the federal, state, and local-levels. Stakeholders typically include other federal agencies (e.g., Substance Abuse and Mental Health Services Administration, Health Resources and Services Administration, Federal Emergency Management Agency), national-level coalitions representing state or local health agencies (e.g., National Association of Chronic Disease Directors, National Association of City and County Health Officials, National Coalition of STD Directors), national non-profit organizations (e.g., American Cancer Society, Campaign for Tobacco Free Kids), other programs at CDC, the program grantees, and local-level agencies primarily responsible for program implementation. These stakeholders, especially some of the national organizations and grantees, leverage their political power to influence CDC decision making, particularly around issues of program policy. Of note is the large *number* of stakeholders involved in each of these four cases – important federal level partners, 58-68 grantees, and hundreds to thousands of local-level partners.

Aside from the political influence of stakeholders, some programs confront a more intensive political climate than others. Of the four cases, the Preparedness Program clearly contends with the fiercest political climate that imposes significant demands on the program. The Cancer Detection Program and Tobacco Control Program also reside in a political context. Breast cancer, in particular, has a strong advocacy community with whom the Division of Cancer Prevention and Control must collaborate. In contrast, the Tobacco Control Program faces a unique political "adversary" in the tobacco industry.

**8.2.2 Network Characteristics**

This section summarizes the network characteristics of the four cases including aspects of network structure and function. Table 29 summarizes some of the network characteristics according to each of the four cases; appendix J offers a more detailed listing.

**Table 29. Summary of Network Characteristics by Case**

| Preparedness Program | STD Program | Cancer Detection Program | Tobacco Control Program |
|---|---|---|---|
| Network Structure: Vertical Relationships | | | |
| 62 grantees; thousands of local-level partners | 65 grantees; hundreds of local-level partners | 68 grantees; 22,000 local-level providers | 58 grantees; some local-level partners |
| Most vertical relationships formalized via funding | Most vertical relationships formalized via funding; dominant dimension | Most vertical relationships formalized via funding; dominant dimension | Most vertical relationships formalized via funding |
| Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory |
| Primarily intergovernmental throughout vertical chain | Local-level partners include: local public health agencies, CBOs, family planning clinics | Regional and local-level partners include: public health agencies, CBOs, private providers, community health centers, family planning clinics | Local-level partners include local health agencies, CBOs (usually funded with non-CDC funds) |
| Program activities at state and local levels | Most program activities at local level | Service delivery at local level | Program activities at state and local levels |
| Network Structure: Horizontal Relationships | | | |
| Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local |
| Typically informal and unfunded | Typically informal and unfunded | Some funded, others informal and unfunded | Typically informal and unfunded |

| Preparedness Program | STD Program | Cancer Detection Program | Tobacco Control Program |
|---|---|---|---|
| Essential to achieving program goals – program outcomes dependent on network efforts | Increasing importance of partnering with horizontal partners to access priority populations | Support referral of priority populations to program services; program integration; provide advocacy | Essential to achieving program goals – program outcomes dependent on network efforts |
| **Network Function: Control and Authority Within the Network** | | | |
| Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation |
| No authority over unfunded, horizontal partners on which grantees are dependent | No authority over unfunded, horizontal partners | In vertical chain, control and authority facilitated by funding and mgmt tools, network relations | No authority over unfunded, horizontal partners on which grantees are dependent |
| Cooperative agreement provides some authority over grantees | Grant offered minimal authority over grantees; recent shift to cooperative agreement | Cooperative agreement provides some authority over grantees | Cooperative agreement provides some authority over grantees, but not exercised |
| **Network Function: Shared Organizational Goals and Priorities Within the Network** | | | |
| Differing priorities across grantees; goal and mission conflicts with some horizontal partners | Differing priorities across grantees; goal and mission conflicts with some horizontal partners | Shared goals among network partners | Shared goals among network partners |
| **Network Function: Context, Capacity, and Resources** | | | |

| Preparedness Program | STD Program | Cancer Detection Program | Tobacco Control Program |
|---|---|---|---|
| Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources |
| Variability in level of risk for and type of "hazard" | Variability in STD epidemiology | Variability in demographic profile of priority population; cultural barriers | Variability in emphasis around goal areas |

8.2.2.1 Network Structure: Vertical Relationships

The network implementation structure for the four cases is comprised of both vertical and horizontal dimensions. As indicated, all four programs support a large number of grantees representing states, cities, territories, and tribal organizations. Those grantees, in turn, support hundreds and up to thousands of local-level partners primarily responsible for program implementation, although the grantee agency may also conduct some program activities (e.g., state-wide planning, surveillance). The Cancer Detection Program's vertical network is the largest with 68 grantees and over 22,000 local-level screening and diagnostic service providers. As depicted in figure 1, the vertical network structure of the Cancer Detection Program can include up to four levels.
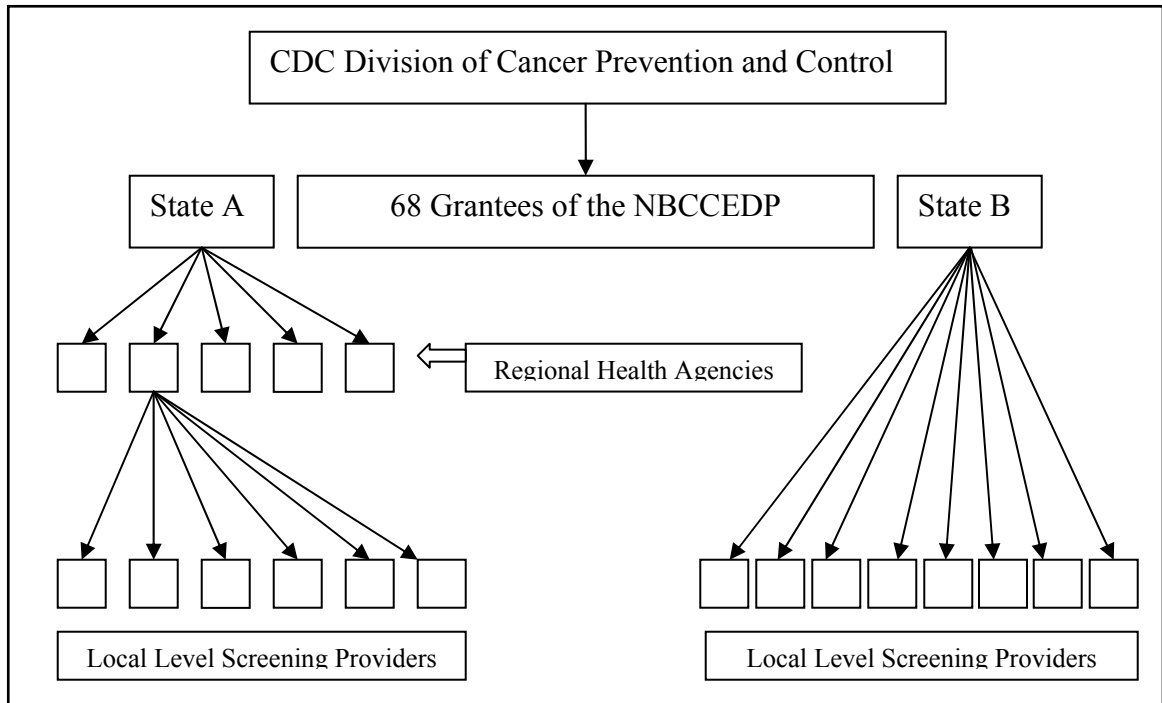
**Figure 4. Vertical Network Structure for the National Breast and Cervical Cancer Early Detection Program [Cancer Detection Program]**

For all cases, the vertical relationships typically extend from Health and Human Services to CDC to the grantee to local level partners. These vertical relationships within the implementation networks are almost always formalized through funding relationships. CDC funds its grantees through cooperative agreements and grantees use varied funding mechanisms (e.g., grants, contracts, memorandum of understanding) to support local-level partners. State, city, territorial, and tribal infrastructure differs and their individual policies or laws may specify required relationships between, for instance, state and local government. The vertical network should *not* be assumed to be entirely intergovernmental – three of the four cases (i.e., STD Program, Cancer Detection Program, Tobacco Control Program) fund local-level, non-governmental agencies such as

community-based organizations, family planning clinics, private health care providers, and community health centers.

### 8.2.2.2 Network Structure: Horizontal Relationships

All four cases involve lateral partners at the federal, grantee, and local levels. For instance, the Office on Smoking and Health collaborates with the Substance Abuse and Mental Health Services Administration, National Cancer Institute, American Cancer Society, and others at the federal level; Tobacco Control Program grantees typically work with other divisions within their department of health, advocacy organizations (e.g., Campaign for Tobacco Free Kids), and non-profits (e.g., American Cancer Society); and local-level agencies interact with partners for coalition building and the implementation of program activities. Similarly, the Division of STD Prevention works with Health Resources and Services Administration, other Divisions within CDC, and the National Coalition of STD Directors. STD Program grantees collaborate with other health-related departments (e.g., HIV/AIDS) and state peers. And agencies funded at the local-level as part of the STD Program work closely with jails, juvenile detention facilities, family planning clinics, and private health care providers. While the vertical relationships within the implementation structure are, for the most part, formalized through funding mechanisms, the horizontal relationships across the four cases are most often informal and un-funded. Consequently, while grantees have some level of authority over their vertical partners, formal authority over horizontal partners is absent.

Horizontal relationships serve different purposes for the four cases. For the Preparedness Program and the Tobacco Control Program, these relationships are essential to achieving program goals and are an integral part of overall program implementation.

322

For instance, an effective response to pandemic flu will require not only the leadership of the official health agency in a state, city, or territory, but also necessitates the involvement of other departments and sectors to ensure a coordinated, integrated, and effective response. Schools may need to be closed, requiring the involvement of the Department of Education and local school boards; hospitals will have a critical role in caring for the sick; personal mobility may be restricted demanding the participation of the Department of Transportation and even the Police Department; and commerce may be enlisted to assist with mass dissemination of medication. Similarly, the four program goals established for effective tobacco control cannot be achieved by CDC or Tobacco Control Program grantee institutions alone. A compliment of coordinated strategies involving community interventions and mobilization; counter-marketing; policy development and implementation; and surveillance are all needed to affect relevant outcomes.

In contrast, horizontal relationships within the STD Program and the Cancer Detection Program primarily support improved access to priority populations for sexually transmitted disease testing and treatment and for breast and cervical cancer screening, respectively, as well as service integration. For example, STD Program grantees encourage their local-level partners to establish relationships with jails and juvenile detention facilities in order to provide screening for syphilis, gonorrhea, and chlamydia among inmates. As the Division of STD Prevention continues to promote a broader "community perspective" for the STD Program and introduces population-based goals, horizontal partners will play an increasingly important role in achieving those goals. At

present, however, the STD Program is best characterized as leveraging its horizontal relationships to extend access to priority populations.

In the Cancer Detection Program, horizontal partners provide important public education messages and help identify and refer program-eligible women for screening. In addition, these partners assure greater integration of service delivery. For instance, Cancer Detection Program grantees may work closely with sister programs such as WISEWOMAN,™[54] comprehensive cancer control programs, and newer state or federally-funded colorectal cancer screening programs to facilitate integration of service delivery. A program-eligible woman coming to a health clinic for a Cancer Detection Program-funded mammogram ideally would also be offered cardiovascular health screening funded by the WISEWOMAN™ program and be assessed for colorectal cancer screening.

8.2.2.3 Network Function: Control and Authority within the Network

In all four cases, CDC's control and authority over program implementation is compromised given the decentralized, networked structure. As indicated by the descriptions above, program activities are typically conducted two to three steps removed from CDC and one to two steps removed from the grantees. The Preparedness Program, Cancer Detection Program, and Tobacco Control Program all use cooperative agreements as the funding tool with their awardees; the STD Program moved from a grant to cooperative agreement in January 2009. While this particular mechanism supports CDC involvement with its grantees in terms of program-related decision making, grantees also

---

[54] Well Integrated Screening and Evaluation for Women Across the Nation

have a fair degree of discretion in program implementation. In addition, the level of authority *exercised* by CDC through the cooperative agreement and grant varies across cases. Program requirements (as articulated in the funding announcement and funding award) differ for each program as does the political will to impose sanctions on grantees perceived to be out of compliance. With the absence of a direct funding relationship, CDC control weakens considerably at the local level. CDC is dependent on its grantees to exercise authority over local level partners through their funding tools or other management practices.

Of the four cases, the Division of Cancer Prevention and Control was observed to exert the greatest control over its Cancer Detection Program grantees within the vertical chain based on the funding awards' requirements and other management systems in place. For instance, the Division of Cancer Prevention and Control has established an extensive data monitoring system that allows staff to identify potential implementation problems and intervene as needed. In addition, Division of Cancer Prevention and Control management imposes sanctions on grantees viewed as out of compliance with program requirements, particularly if there are concerns about the care of women served through the program. The authority relationships between grantees and their local-level partners are less understood given that few representatives from grantee agencies were included in the study. However, control and authority between grantees and their funded partners is also likely influenced, in part, by the funding tool used. Several interviewees with the Cancer Detection Program described grantees' use of performance-base contracting and other reimbursement policies that afford them greater control over local-level providers and, consequently, service delivery.

In all four cases, the CDC, grantees, and local-level funded agencies typically lack formal authority over informal, horizontal partners involved in program implementation. For the Preparedness Program and the Tobacco Control Program that rely on horizontal partners to achieve program goals, significant dependencies on partner cooperation are created. For instance, the 2006 legislation titled the Pandemic and All-Hazards Preparedness Act of 2007 requires that Preparedness Program grantees develop pandemic flu plans, and a 2007 initiative issued by the White House (Homeland Security Presidential Directive, #21) requires that grantees ensure local capacity to distribute countermeasures (i.e., prophylaxis treatment) to all citizens within 48 hours of an incident (e.g., anthrax attack). Both initiatives demand extensive cooperation of horizontal partners at the grantee-level and local-level, but neither grantees nor local-level vertical partners have formal authority to demand the participation of their horizontal network partners.

8.2.2.4 Network Function: Shared Goals and Priorities within the Network

The four cases differ in the degree to which they share goals and priorities within their individual networks. As noted earlier, the Cancer Detection Program and Tobacco Control Program enjoy strong consensus among both vertical and horizontal network partners in regard to their program goals and objectives. For the Office on Smoking and Health, the broader community of organizations involved in comprehensive tobacco control has embraced the Tobacco Control Program's four goals related to adult cessation, youth prevention, second-hand smoke, and reduced disparities. These goals drive comprehensive tobacco control efforts nationally, not just for the Tobacco Control Program grantees funded by CDC. Similarly, the Cancer Detection Program's goal to

screen low-income, under-insured women and detect cancer early when it is more effectively treated is embraced widely by network partners.

Conversely, the Preparedness Program and the STD Program experience priority and goal differences that may affect program implementation. For instance, the Preparedness Program grantees have struggled to engage some horizontal partners with differing priorities in efforts to develop pandemic influenza plans. Data suggest these partners, often representing other sectors (e.g., emergency response, transportation, education), may not yet understand the public health role in preparedness or may, themselves, face competing demands that preclude participation. Likewise, STD Program grantees have been challenged to involve jails and juvenile detention centers that have starkly different missions (i.e., public safety) than sexually transmitted disease prevention. Within the network context of these four cases, shared goals and priorities seems to facilitate the cooperation of both vertical and horizontal partners in program implementation while conflicts in mission, goals, and priorities serve as a barrier to coordinated implementation efforts.

### 8.2.2.5 Network Function: Context, Capacity, and Resources

Common across all four cases is the extensive *variability* among grantees in terms of their context, capacity, and resources. First, the extent and nature of the public health problem varies across individual grantees – the incidence and prevalence of specific sexually transmitted diseases in Florida differ dramatically from those in New Hampshire or South Dakota. The risk for a bioterrorist attack is likely greater in New York than in Iowa. And the demographic characteristics and cultural norms of women eligible for the Cancer Detection Program in West Virginia are quite different from those in Texas. Other

contextual factors reflect variability across grantees in the four programs as well –
differing geography (e.g., rural, urban, size in square miles), state laws and regulations,
and the strength and influence of particular industries (e.g., tobacco industry).

Capacity, which is influenced by resource levels, also varies from grantee to
grantee. Capacity is often reflected in the number of staff assigned to the program, the
expertise available to the program, and infrastructure (e.g., data management systems).
The structure and capacity of the larger network within a state or territory also varies;
local-level capacity and overall network infrastructure have important implications for
program implementation.

And, as reflected in chapters 4-7, the range of the funding award size is significant
across all cases resulting in variable resource levels for grantees. For instance, the award
size for STD Program grantees ranges from $43,609 $6,711,083. In addition, state
contributions to individual programs will vary as well. In the past, the state legislature of
California provided over $100 million to augment the Cancer Detection Program funding
from CDC while other states did not contribute any additional resources.

Together, all three of these factors – context, capacity, and resources – influence
each grantee's program priorities, activities, and ability to achieve goals. Study
participants for all four cases emphasize that grantee performance should be interpreted
based on the individual program context. Comparing grantees on measures of
performance was viewed as futile given their differences. Across cases, each grantee
program is considered unique.

### 8.2.3 The Performance Measurement Systems

The four performance measurement systems studied represent ones in different stages of development. All were initiated after the passage of GPRA suggesting the importance of that particular policy in advancing performance measurement as a management tool for government programs. This section summarizes the process of developing performance measurement systems for the four cases, their design features, and each case's current use of performance data.

8.2.3.1 Development Process

Although all four programs introduced their first set of performance measures between 2003 and 2004, the cases represent systems at different levels of development. The Cancer Detection Program performance measurement system is the most mature; minimum data elements providing the basis for the measures have been collected since the program's inception in 1991 and the data quality indicator guide, that includes most of the current 11 performance measures, was introduced nearly fifteen years ago in 1993. The 11 Cancer Detection Program performance measures were formally implemented by the Division of Cancer Prevention and Control in 2004 as "core program performance indicators." Overall, the Cancer Detection Program has the most sophisticated performance management system of the four cases and utilizes its performance data for a range of purposes including accountability, program improvement, and budgeting. The minimum data elements and performance measures are considered vital to program management and are widely accepted and viewed favorably by Division management and grantees.

The other three cases are more similar in the time line of their development process, but still vary in regard to their developmental progress. The Division of State and Local Readiness launched its first set of over 120 performance measures for the Preparedness Program in 2003, but the program has experienced a significant expansion in scope and extensive staff turnover since that time. A revised set of 47 measures was put forth in 2004, followed by 35 measures in 2005, and the current set of 6 measures in 2007. Currently, staff in the Division's monitoring and evaluation branch are working with programmatic and scientific experts, internal and external to CDC, to develop a conceptual framework built around five areas of preparedness-related capability (e.g., incident management, biosurveillance). The framework is intended to better define the Preparedness Program and provide a foundation from which to identify future performance measures. As in the Division of STD Prevention, the Division of State and Local Readiness faces issues of organizational support, particularly among program consultants, and resistance from grantees. Data validity has been a perennial problem since the Preparedness Program first launched indicators.

The Division of STD Prevention initiated efforts to develop performance measures in 1999, piloting over 60 measures in seven states over a two-year period beginning in 2001. Based on the pilot, twelve measures were included in the 2004 STD Program five-year grant agreement as a reporting requirement for grantees. Revisions to the measures were made in 2007. Development efforts have continued, and the five-year STD Program cooperative agreement funded in January 2009 includes a set of 17 measures, some from previous years and others newly introduced. Although the STD Program performance measures have evolved over time, organizational support for the

system remains inconsistent, data validity is problematic, and use of the performance data is limited.

Finally, the Office on Smoking and Health started their development process in 2001. In 2005, the *Key Outcome Indicator Guide*, inclusive of 120 measures for the Tobacco Control Program was published. Grantees currently report on measures most relevant to their programs, although reporting on the key outcome indicators is not a requirement of their funding award. In 2006, the Office on Smoking and Health began working on a set of core indicators. A list of 30 approved indicators is being finalized in 2009, and staff hope that these will provide a "national snapshot" of the program for stakeholders "up" the vertical chain (e.g., Health and Human Services, Office on Management and Budget). The Tobacco Control Program is considered the least developed among the four cases because the Office on Smoking and Health has not completed and implemented the core set of performance measures.

The process to develop performance measures for these four cases can be characterized as negotiated and evolving. Adoption of the systems and data use seems to occur slowly and incrementally over time. First, given the network structure of the four cases, the development of measures becomes a negotiated process between the CDC program and those who are accountable for performance and responsible for the relevant data collection, reporting, and management – primarily the grantees. Consequently, stakeholder involvement is recognized by all four cases as essential to the development process. In developing their performance measurement systems, staff in all four cases have included representatives from associations representing grantee programs (e.g., Directors of Public Health Preparedness, National Coalition of STD Directors, National

Breast and Cervical Cancer Early Detection Program Council of Program Directors) and grantee staff. These stakeholders and others have been included in performance measurement workgroups (i.e., STD Program, Cancer Detection Program), expert panels (i.e., Preparedness Program, Tobacco Control Program), pilot efforts (i.e., STD Program), and vetting processes (i.e., STD Program, Cancer Detection Program, Tobacco Control Program). Participants in all four cases emphasized the importance of including stakeholders to ensure that program performance measures are feasible, meaningful for practice, and valid – a unique challenge with 58 or more grantees. In addition to the expertise stakeholders contribute to the development process, study participants stressed the importance of stakeholder involvement to building the necessary buy-in needed to assure serious participation and data quality. As noted above (section 8.2.1.4), stakeholders also possess a level of political power and influence – consequently, the development of performance measures becomes a negotiated process between stakeholders and CDC.

In addition, for these four cases, the development of performance measures has been evolving. In general, performance measurement systems are expected to be dynamic. Periodic revisions to measures based on changes in program priorities, improved measurement, increased availability of data, and, for public health, advancements in scientific understanding, are expected. The incremental and evolving nature of these four performance measurement systems, however, may best be attributed to the fact that these systems are in their early development and because the measures are meant to apply to broadly decentralized, networked programs. The evolution has often been dramatic: the Preparedness Program moved from 120 measures in 2003 to 6

measures in 2007; the STD Program from 60 measures in 2001 to 12 in 2004 and 17 in 2009; the Cancer Detection Program evolved from a large monitoring dataset with over 100 variables to the data quality indicator guide in 1993 and 11 core measures in 2004; and the Office on Smoking and Health is transitioning from 120 key outcome indicators for the Tobacco Control Program in 2005 to 30 core indicators in 2009.

For the Preparedness Program, the dramatic change in measure composition has largely been a factor of its political environment and turnover among evaluation staff. For the STD Program, the evolution reflects the development process itself. The program started by piloting a large set of potential indicators with a subset of grantees. The Division of STD Prevention and Control has gradually refined these over time based on field testing and grantee feedback. The Cancer Detection Program implemented its large monitoring data set at the time the program was initiated in 1991; that monitoring system is still maintained. Core performance measures for the program were introduced later as a means to communicate program priorities to the Cancer Detection Program grantees and begin performance-based budgeting. In the Office on Smoking and Health the 120 key outcome indicators remain intact for the Tobacco Control Program, but given program variability, not all 120 measures are relevant or available for all grantees. Consequently, the smaller set of 30 core indicators currently under development is intended to provide a common set of data that can be assessed across all grantees and used to provide a national summary of the program.

Finally, the adoption of performance data seems slow and somewhat incremental. Performance measurement becomes more challenging in the networked environment where 58 to 68 grantees must embrace the effort and communicate measures to hundreds

or thousands of local-level partners involved in implementation. For these programs, the implementation and adoption of performance measurement systems seems to advance in a slow, step-wise fashion that must encompass three to four levels in the vertical chain – measures are introduced, data management systems developed for 50 plus grantees, data collected and reported, and data quality addressed and improved over time. Data use seems to follow as CDC staff begin to use and promote data, grantees take up use of the data, and local-level agencies become aware of the importance of the data.

8.2.3.2 Design Characteristics

A summary of characteristics for the four performance measurement systems is provided in table 30; a more detailed synthesis is included in appendix K.

**Table 30. Design Characteristics of Performance Measurement Systems by Case**

| | Preparedness Program | STD Program | Cancer Detection Program | Tobacco Control Program |
|---|---|---|---|---|
| **Purpose** | Accountability | Accountability and program improvement | Accountability, program improvement, budgeting | Accountability |
| **Level of Measurement** | Grantee-level | Local and grantee-level | Local-level (patient-level clinical data) | Primarily grantee but some local-level |
| **Types of Measures** | Process measures that reflect program capabilities | Process and immediate outcome measures | Process, short-term and intermediate outcome measures | Short, intermediate, and long-term outcome measures |
| **Use of Targets** | Time-based targets for some measures set by CDC | Baseline and 3-year targets set by grantees | Common standards set by CDC | Baseline and targets set by grantees |
| **Quality Assurance** | Yes | Yes | Yes | Yes |

8.2.3.2.1 Purpose

All four cases identify accountability as an important purpose for their

performance measurement systems. Given the current fiscal crisis, participants across the

cases described the importance of collecting performance data to help preserve funding

levels and defend against potential budgetary cuts. The priority for the Preparedness

Program and Tobacco Control Program is to address accountability concerns from

"above" – that is, for the Division of State and Local Readiness and the Office on

Smoking and Health to provide performance measurement data to stakeholders at Health and Human Services, the Office of Management and Budget, the Assistant Secretary for Preparedness and Response, and other federal-level stakeholders in order to demonstrate accountability for the national program and its management. The Preparedness Program and Tobacco Control Program grantees may also use the performance data to demonstrate accountability with their own state-based (city-based, territorial-based) stakeholders to defend state or local contributions to the programs if they receive such funds.

While the Division of STD Prevention and the Division of Cancer Prevention and Control also recognize the importance of accountability to federal-level stakeholders, CDC staff are more focused on using the performance measurement systems to hold their STD Program and Cancer Detection Program grantees accountable for their performance. These two programs emphasize the purpose of program improvement, along with accountability, as an important driver of their performance measurement systems.

In regard to purpose, the most interesting dichotomy across the four cases is that between the Preparedness Program and the Cancer Detection Program. The Division of State and Local Readiness is working to develop unique sets of performance measures to serve individual purposes – the six existing performance measures for accountability "up," a new set of measures to meet the performance-based budgeting requirements imposed by the Pandemic and All-Hazards Preparedness Act of 2007, and a larger data set of capacity measures to support program monitoring and improvement. The Division of State and Local Readiness' need to develop unique set of measures is influenced, in part, by the nascent stage of its overall program development, the nature of the program and complexity of the problem, and the political context around the Preparedness

Program. In contrast, the Division of Cancer Prevention and Control uses its one set of eleven Cancer Detection Program measures for all three purposes – accountability, program improvement, and budgeting. As noted above, the Division of Cancer Prevention and Control also has its large set of monitoring data (minimum data elements) for the Cancer Detection Program.

8.2.3.2.2 Level of Measurement

For all cases, much of program implementation occurs at the local-level – preparedness exercises, STD case finding, breast cancer screening, and school-based tobacco control programs. Performance measures for the STD Program and Cancer Detection Program – data are collected at the local-level and the performance measures typically represent individuals served through the program (e.g., among clients of sexually transmitted disease clinics, the proportion of women with positive gonorrhea tests that are treated within 14 and 30 days of the date of specimen collection; percentage of women screened through the Cancer Detection Program diagnosed with invasive carcinoma with time from date of diagnosis to treatment started less than 60 days). Data are aggregated at the grantee-level for submission to CDC.

In contrast, state- or territorial-level survey data will primarily be used to support the core Tobacco Control Program performance measures, although a few measures will rely on locally-collected data (e.g., extent and type of retail tobacco advertising and promotions). For the Preparedness Program, challenges related to data availability, local-level capacity, and measurement have all contributed to the reliance on grantee-level measurement for now. A common sentiment across three of the four cases (Preparedness Program, STD Program, and Tobacco Control Program) has been to select "low hanging

fruit" – measures for which data are available and the capacity to collect and report data exists.

8.2.3.2.3 Types of Performance Measures

Aside from the Tobacco Control Program performance measures, process and short-term outcomes reflect the dominant measure type used for the three other cases. For example, the Cancer Detection Program includes process, short-term, and intermediate-level measures reflecting service delivery (i.e., screening and diagnostic services) in its set of 11 indicators. All are consistent with the program's focus on the women served through the program (rather than population-level measures). For all three programs Preparedness Program, STD Program, Cancer Detection Program CDC has been sensitive to introduce measures and targets that are viewed by grantees as feasible and achievable. Participants working with the Preparedness Program and STD Program described an intention to introduce more challenging, intermediate-level outcome measures over time as grantees become more comfortable with the measurement process. Again, this demonstrates the evolving and incremental development process of these systems discussed above in section 8.2.3.1.

In addition, grantees have successfully argued for measures that they can "fairly" be held accountable for, that is, measures that are largely under their control. This inclusion criterion, whether explicit (Preparedness Program) or not (STD Program, Cancer Detection Program), has effectively dismissed longer-term outcomes that are often influenced by multiple factors and even some short-term outcomes that are dependent on partners over which grantees have little authority. In contrast, the Tobacco Control Program key outcome indicators and proposed core measures *only* include

outcome measures (short-term, intermediate, and long-term). Two factors influenced this

choice. First, from the outset, the development team aimed to identify outcome-level

measures given the emphasis by federal-level partners in addressing accountability.

Second, with 58 grantees implementing "best practices" in different ways, the Office on

Smoking and Health found it impossible to identify a common set of process measures.

In focusing on outcome-level measures, the Office on Smoking and Health recognizes

that the measures represent the results of comprehensive tobacco control efforts, not

simply the activities supported with CDC funds. In other words, Tobacco Control

Program grantees are not expected to influence performance on these measures alone.

The Office on Smoking and Health recognizes that the outcomes will only be achieved by

the joint efforts of network partners in the tobacco control community. Of note, the

acceptance of the outcome-based core measures by Tobacco Control Program grantees is

untested given that the measures have not been introduced. In addition, the planned use

for the performance data seems limited to federal-level accountability purposes for now.

Distal measures, typically representing population-level changes in morbidity and

mortality, have largely been rejected by all cases for inclusion in their performance

measurement systems. For public health programs, this level of change typically takes

years to achieve and requires resources beyond those allocated for any of these four

cases. GPRA/PART measures for the STD Program, Cancer Detection Program, and

Tobacco Control Program include population-level, distal measures[55] which participants

---

[55] For example, one GPRA/PART measure for Division of STD Prevention is as follows,
"Reduce pelvic inflammatory disease (PID) in the United States." Interviewees explained
that there is not even consensus on how to define PID epidemiologically. One of the
GPRA/PART measures for DCPC is the "age-adjusted annual rate of breast cancer
mortality per 100,000 female population." As noted in Chapter 6, resources for the

in this study generally viewed as unrealistic to achieve based on program funding levels. Given that preparedness is not disease-specific and, therefore, lacks incidence and prevalence data, the Division of State and Local Readiness has successfully negotiated three GPRA/PART measures with the Office of Management and Budget that are directly tied to a subset of their six performance measures.

8.2.3.2.4 Use of Targets

Both the Preparedness Program and the Cancer Detection Program have established standards for their performance measures (e.g., time for primary staff with public health agency Incident Command System functional responsibilities that the public health agency's Emergency Operations Center is being activated – target 60 minutes or less; percentage of abnormal screening results with time from screening test result to final diagnosis greater than 60 days – target 25% or less). Neither the STD Program nor the Tobacco Control Program have set targets for their performance measures due to variability in disease burden across grantees and the absence of adequate trend data to help substantiate a standard. The Division of STD Prevention encourages grantees to document a baseline for their performance measures and set individual targets based on their epidemiology, resources, and capacity. Similarly, the Office on Smoking and Health currently requires Tobacco Control Program grantees to set a baseline and three-year target for any key outcome indicator that grantees report through CDC's web-based reporting system.

---

Cancer Detection Program support screening for approximately 1% of women in the United States. Consequently, interviewees for this study argued that the Division of STD Prevention cannot fairly be held accountable for achieving the related benchmark for this measure.

8.2.3.2.4 Quality Assurance Efforts

All programs with performance measurement systems in place (Preparedness Program, STD Program, Cancer Detection Program) require their grantees to submit data twice annually to CDC. Web-based systems are used by all three programs to report data or submit a data file. The Tobacco Control Program uses a web-based system for voluntary reporting of the key outcome indicators already in place. Data quality (e.g., complete data, valid data) emerges as a predominant concern for all four cases given their extensive decentralization. For programs that rely on local-level data collection and reporting like the STD Program and Cancer Detection Program, data quality is particularly challenging. And even for the Preparedness Program and Tobacco Control Program that involve grantee-level data collection, CDC has experienced data quality problems. Recognizing that the data will be relatively meaningless if quality is poor, staff in all four programs are making efforts to address quality assurance.

Data quality problems may largely reflect the developmental stage of the performance measurement systems. For instance, the Cancer Detection Program seems to have the strongest data quality systems instituted of the four cases – cooperative agreement funds support a data manager for each grantee, a standard data management software is available to all grantees, on-going technical assistance for data management is provided by a data contractor, regular edit checks are conducted at the grantee level and by the Cancer Detection Program's data contractor, edit reports are produced, and data quality is closely examined as part of semi-annual data reviews for each grantee. In addition, a national data validation study of the minimum data elements was conducted from 2003-2007 and results reflected high quality data for the national program.

In contrast, the Preparedness Program and STD Program have both faced significant data quality problems which the Division of State and Local Readiness and the Division of STD Prevention are working to address. Participants in the Division of STD Prevention suggested that data quality is improving with time and increased technical assistance efforts. Both the Preparedness Program and STD Program provide grantees performance measurement guides with detailed information about each measure's intent, operational definition, data sources, numerator and denominator specification, and references supporting the measures. However, staff working with the Preparedness Program and STD Program expressed frustration for grantees' lack of compliance with the guidance.

8.2.3.3 Use of the Performance Measurement System and Data

Use of the performance measurement data varies across cases. The Cancer Detection Program is the most sophisticated user of their performance data employing it for the multiple purposes of accountability, program improvement, and budgeting. Data use for the Cancer Detection Program is supported by two key factors. First, the Cancer Detection Program has established a strong data-driven program culture – as noted, the minimum data elements have been an integral part of the program since its inception in 1991. As described in chapter 6, the Division of Cancer Prevention and Control has long-supported a comprehensive and well-funded performance management system. For instance, grantees are provided resources through their cooperative agreement award to support data collection and reporting (e.g., staff salary for a data manager, data system support). In addition, the Division of Cancer Prevention and Control funds a data contractor to manage the national data set for the Cancer Detection Program. Technical

consultants working for the data contractor are assigned to work with grantees to provide on-going data management support. And Cancer Detection Program staff at CDC include a senior data manager and programmer to work with the data contractor and help manage the national dataset. Second, the performance measures are overwhelmingly viewed as meaningful and valid measures that represent program priorities and support program management. In other words, the perceived value of the performance measures support and reinforce their use.

Consistent with the Division of State and Local Readiness' stated purpose for its performance measures, staff are using Preparedness Program performance measurement data to respond to accountability demands from the Office of Management and Budget and other federal-level partners, although data quality problems have limited its utility. More importantly, however, the Division of State and Local Readiness staff and grantees interviewed for the study recognize that the current set of six measures is not perceived as particularly meaningful or useful to grantees. Grantees would prefer performance measures that better reflect program priorities and that will support program improvement. Staff with the Preparedness Program recognized that wider adoption and use of performance measures for the program will be influenced by their ability to identify more relevant measures as part of the current development process. In addition, program consultants in the Division have had little to no involvement in the development efforts; as the primary CDC contact with grantees, their participation and investment in the effort will be needed to support data quality and data use by grantees.

The use of the STD Program performance measures has been influenced by data quality concerns, the lack of organizational support within the Division of STD

Prevention from management and program consultants, and resistance from grantees. Staff working on the STD Program performance measures have been frustrated by the lack of buy-in for and support of the effort – while some grantees are using the measures and even developing their own measures, adoption across the broader network is perceived to be limited.

Use of the Tobacco Control Program core performance measures is, as yet, untested given that the indicators are still in development. Grantees voluntarily report key outcome indicators of their choice as part of regular progress reporting. Of interest, the Office on Smoking and Health did not mandate reporting of the new core performance indicators (or key outcome indicators) in the new five-year funding announcement issued in 2009.

Overall, several factors seem to influence the use of performance measurement data for the cases included in this study. These include the perceived relevance of the measures to grantees, how well the measures reflect program goals and priorities, data quality, the existence of a performance management system supporting data use, and stakeholder buy-in– especially by CDC program consultants, management, and grantees. In addition, results suggest that adoption of performance measures in these large networked contexts takes time and an investment of resources to facilitate their acceptance and use.

### 8.3 Cross-Case Results

Findings from the cross-case analysis are presented in this section. Table 31 provides a summary of the cross-case results organized by Jennings and Haist's

hypotheses and the research questions for this study. As a reminder, a listing of results

from the *with-in* case analysis is provided in appendix N.

**Table 31. Summary of Cross-Case Findings**

| | |
|---|---|
| **Jennings and Haist, hypothesis #13:** The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed (2004, p.185). | |

| **Research Question #1.** How does networked public management affect CDC's observability of program outputs and outcomes? | **Finding 8.3.1.1** Observability (measurement) of program outputs and outcomes is primarily related to the complexity of the public health problem, but network differences in program priorities, activities, epidemiology, and resource levels do challenge the ability to observe program outputs and outcomes across all grantees. |
| | **Finding 8.3.1.2** While networked public management is necessary to achieve program outcomes, it compromises observability (attribution) and fragments accountability. |
| **Research Question #2.** How does networked public management influence CDC's use of performance measurement and the types of performance measures used? | **Finding 8.3.2.1** The use of performance measurement is an expectation of the political and administrative context and is not influenced by the network structure. |
| | **Finding 8.3.2.2** The types and choice of performance measures are affected by network variability, the political influence of network stakeholders, and the network structure which compromises control over outcomes. |

| | |
|---|---|
| **Jennings and Haist, hypothesis #14**: Measurement will be more common and will have greater impact when agencies have greater control over outcomes. (2004, p.185) | |

| **Research Question #3.** Does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? | **Finding 8.3.3.1** Dependencies on both vertical and horizontal network partners diminish CDC and grantee control over outcomes, restricting the design and impact of performance measurement. |

**8.3.1 How does networked public management affect the observability of CDC's program outputs and outcomes?**

As noted in the introduction to this chapter, Jennings and Haist (2004) define observability as it relates to performance measurement in two ways – the ability to *measure* outputs and outcomes and the ability to *attribute* outcomes to the program of interest. Two related findings are presented below.

8.3.1.1 Observability (measurement) of program outputs and outcomes is primarily related to the complexity of the public health problem, but network differences in program priorities, activities, epidemiology, and resource levels do challenge the ability to observe program outputs and outcomes across all grantees.

The ability to measure outputs or outcomes for these four cases is primarily influenced by the complexity of the problem each addresses rather than the network implementation structure. The nature of the program (e.g., preparedness, infectious disease control and prevention, health care service delivery, prevention of chronic diseases), the available evidence base, the complexity of the problem – these are the most important factors influencing whether the four programs' outputs and outcomes can be measured.

As a service delivery program, the Division of Cancer Prevention and Control can easily measure both outputs and short-term and intermediate outcomes for the Cancer Detection Program. A CDC staff person commented,

> This program certainly has advantages in that we're able to quantify things in a way that other people can't. But that's purely because we are a direct service delivery program, where you can count the number of women screened, you can count the number of tests provided, you can count how many [tests] were abnormal, you can count how long it took to get from an abnormal screening result to a follow-up test, and from diagnostic determination of cancer to [cancer] treatment. We have the luxury of doing that.

346

The Preparedness Program, however, represents a much more complicated

problem – preparedness for the public health consequences of disasters as diverse as

flooding and bioterrorist attacks. During a presentation at the national preparedness

conference (2008), a lead evaluator from the Division of State and Local Readiness

suggested that public health emergency preparedness is not an "end state" reflected in

specific outcomes, but rather, an on-going process to build and maintain infrastructure,

partnerships, and plans for demonstrating capability to respond in real emergencies.

Unlike many other public health areas, preparedness lacks disease-specific incidence and

prevalence data. One participant said,

> We're not so sure what it means to be prepared; is emergency
> preparedness all about getting everybody trained in incident command and
> setting up an emergency op [operations] center and getting everybody
> there in 15 minutes from the time that you know there's an emergency, is
> that what emergency preparedness is all about? Well, that might be a little
> part of it, but it's also about how quickly you stand up your points of
> distribution following an anthrax event…I would say the biggest challenge
> [for performance measurement] is defining what it is we really need to
> measure that will have a meaningful interpretation in the world of public
> health emergency preparedness.

Even for infectious diseases like sexually transmitted diseases that benefit from

robust surveillance data, some important outcomes for the STD Program are difficult to

measure. For instance, although one of the three GPRA/PART measures for the Division

of STD Prevention is to reduce pelvic inflammatory disease, researchers lack consensus

on how to measure it. Likewise, while the Division of STD Prevention (STD Program),

Division of Cancer Prevention and Control (Cancer Detection Program), and the Office

on Smoking and Health (Tobacco Control Program) all intend to reduce racial and ethnic

disparities in disease impact, an evidence-base supporting *how* to reduce disparities (and,

therefore, what to measure) is lacking. From a CDC staff person working with the STD

Program,

> Another major Division commitment is to reduce "disparities." Okay. What does that mean from a process measure, not from just showing this is the rate in whites, blacks, and then it happens to come down, but [measuring] what are you doing [to reduce disparities]? Let's say you're tackling African-Americans – what is it that you want to do that you think is on the pathway to improving [STD] rates among African-Americans? We're way, way far from having any kind of agreement. Is it having a care home, is it having access to care? Short of being able to find that 'what', there's no measure.

Likewise, sexually transmitted disease infections *prevented* or tobacco

consumption *prevented* is nearly impossible to measure. And in all four cases, even when

outcomes *are* measurable, population-based outcomes such as morbidity and mortality

may take years to observe making related measures impractical to include in a

performance measurement system.

Although networks do not affect the *ability* to measure outputs or outcomes,

networks can challenge the capability to *observe* outputs and outcomes consistently

across the grantees for any given case. That is, outputs and outcomes for the four cases

are not equally *observed* across the grantee networks due to differences in epidemiology,

program priorities and activities, network capacity, and resources. This variability within

the network has implications for the performance measurement system and is discussed

in 8.3.2 below.

8.3.1.2 While networked public management is necessary to achieve program outcomes, it compromises observability (attribution) and fragments accountability.

For all four cases included in this study, program outcomes are, to some extent,

jointly produced by both vertical and horizontal partners within the implementation

348

network. The complexity of the public health problems addressed by each case demands networked approaches not only to achieve longer-term goals, but also, for some cases, to achieve outputs, short-term outcomes, and intermediate-level outcomes.

Dependencies on vertical *and* horizontal partners are greatest for the Preparedness Program and Tobacco Control Program given that established goals will only be achieved through collaboration. For example, moderating the effects from a bioterrorist event requires an integrated response across levels of government, numerous departments, and diverse sectors. In such an emergency, a local health department is dependent on an array of other agencies and organizations at the local level as well as state and federal government. A representative from a grantee agency funded as part of Preparedness Program described the networked response in preparedness.

> I think in this arena [preparedness] there's an enormous component that is
> relationships, that is coordination. A great deal of what you do is
> absolutely not under your control but is critical to any type of success.
> You really are mobilizing and enabling others, coordinating across
> networks, managing loosely defined networks.

Similarly, eliminating nonsmokers' exposure to secondhand smoke cannot be achieved by government alone; other agencies and organizations must bring resources to bear in order to meet this goal. Educational, clinical, regulatory, economic, and social strategies are all needed in order to have an impact on secondhand smoke.

As the Division of STD prevention expands its goals for the STD Program to assume the broader 'community perspective' described earlier (rather than a narrow focus on public STD clinics), the role of partners becomes increasingly important. As noted, trends of health care access and delivery have shifted; patients are less likely to seek treatment at publicly-funded sexually transmitted disease clinics than in the past, instead

turning to private, primary care providers. A participant working with the STD Program said,

> Syphilis used to be ninety percent or more treated in public STD clinics.
> Now more than half are treated outside of STD clinics. Particularly men
> who have sex with men, a lot of them are HIV infected. They have care
> providers and they get treated by those care providers. And so they never
> enter the public system. So in order to reach people with syphilis, you've
> got to reach beyond just the STD clinics.

Consequently, engaging agencies and partners that interface with populations at higher risk of sexually transmitted diseases such as private health care providers, schools, departments of corrections, and drug treatment facilities is necessary to effectively address the problem. Even the Cancer Detection Program, a program that predominantly relies on partners within the vertical implementation chain (i.e., CDC, grantee, local health providers) to achieve outcomes, depends on horizontal-level community partners to conduct public education and support client recruitment efforts. To an extent, effectively screening women is a product of collaboration across partnering agencies – some that make the public aware of the importance of screening through educational efforts, others that help identify program-eligible women and refer them to participating providers, and the physicians, radiologists, and surgeons who screen and diagnose them. The statement below, from the most recent funding announcement for the Cancer Detection Program, reflects the program's emphasis on partnerships for all of its program components.

> Program components of the NBCCEDP [Cancer Detection Program]
> include program management; screening and diagnostic services to
> include case management for follow-up of abnormal results; data
> management; quality assurance/quality improvement; evaluation;
> partnerships; and professional development and recruitment. These
> program components are carried out at the local, State and national levels

through collaborative partnerships with State health agencies, community-based organizations, tribes and tribal organizations, universities, a variety of medical care providers and related agencies and institutions, and the business and voluntary sectors.

For the four cases in this study, collaboration with network partners, whether intentional, planned, or otherwise, is meant to support the realization of outcomes. All four programs leverage network partners to achieve synergies, increase resources, and promote the diverse strategies needed to address these complicated public health problems that are influenced by multiple factors. At the same time, however, assigning attribution for specific results to a specific program becomes much more challenging in these networked environments where outcomes (and even outputs in some cases) are jointly produced. For instance, even with a strong evidence base supporting causal pathways, disentangling which outcomes should be attributed to what agency becomes impossible without employing sophisticated evaluation approaches. Typically, performance measurement does not entail the methodological rigor to unravel attribution for outcomes.

In addition, when multiple network partners are all contributing to improved outcomes, accountability becomes fragmented. For example, which agency is ultimately accountable for achieving the Tobacco Control Program's short-term outcome measure #2.3.7, the "level of support for creating tobacco-free policies in public places and workplaces?" As discussed in chapter 7 (section 7.3), the Office on Smoking and Health has made a conscious decision to only include outcome-level measures in its set of Tobacco Control Program core indicators. Recognizing that the outcomes are jointly produced by many network partners involved in comprehensive tobacco control efforts,

assigning accountability for these outcomes to a single agency is impossible. In many ways, this example best represents one of the conundrums of performance measurement as applied in networked contexts – that is, while these four cases all state accountability as a primary purpose for their performance measurement systems, it is often difficult to assign accountability for jointly produced outputs and outcomes. This issue is discussed in more depth in 8.3.3 below.

Although not the focus of this study, per se, accountability for GPRA/PART measures was often discussed by study participants. In theory, CDC programs are held accountable for these measures and program results are used to inform federal-level budget decisions. For the cases of the STD Program, the Cancer Detection Program, and the Tobacco Control Program that all have population-level, distal outcome indicators as part of their GPRA/PART measures, participants were adamant that CDC alone should not be held accountable for achieving them given several factors including limited resources, the network structure that diminishes control over implementation, and the fact that longer-term program outcomes are typically influenced by multiple factors. A participant from Tobacco Control Program said,

> At CDC we're the translation agency so we're not, at least in chronic, we're not doing the bench science that's really easy to measure your outcome. We're about setting up programs. We give money [to grantees], we give program guidance, and we kind of hope that things happen, and we think we know what the outcomes should be if the guidance is implemented. But we don't have any direct control over it…You don't have direct control like you do over your checkbook, and so you say, well I can do this and if good things happen I think I was a part of it but I can't claim credit.

And from the STD Program,

> Accountability is awfully hard to measure. These are very broad – this is *population-based* national stuff, and activities that impact on whether

people are being screened, whether they're being treated, whether their partner's being screened and treated – that's all happening three steps removed from us [CDC] with money that comes to us, to send down there [to grantees] that is completely inadequate. And getting more inadequate by the day when we have this pathology of flat funding when flat funding means less every year.

**8.3.2 How does networked public management influence CDC's use of performance measurement and the types of performance measures used?**

8.3.2.1 The use of performance measurement is an expectation of the political and administrative context and is not influenced by the network structure.

The networked public management context did not influence the decision for staff

in any of the four cases to *use* performance measurement. Instead, data suggest that

program staff and stakeholders assume that performance measurement is the best means

to address demands for accountability. The study provides no evidence that

representatives from these cases assess the appropriateness of performance measurement

for their particular program or context. In part, this finding is based more on what was *not*

said by participants than what they discussed. Results suggest that measurement has

become an expectation for programs, one widely accepted and unquestioned. The

assumption that performance measurement should be employed is probably influenced by

GPRA, PART, and other performance-based initiatives established at CDC. One person

said, "HHS wants it, the top at CDC wants performance measures." Reflecting a common

sentiment among participants, one person from the Preparedness Program remarked,

I believe in performance measures. You know, 'what gets measured gets done.' I've always believed that. And I've always believed in looking at things that measure whether or not your program is meeting what you want your program to do. I'm not sure we're there yet. And we still have a long way to go [in developing measures].

A participant from the Division of STD Prevention discussed how GPRA and

PART, while not particularly useful to their program, are used to justify the importance

of accountability and the use of performance measurement with grantees.

> It's very useful politically for us to say, 'There's an era of greater
> accountability, and we need to be documenting and demonstrating,' which
> is not untrue. GPRA is not coming and going, they were smart – it passed
> Congress and so it's a law. The actual [GPRA] measures for us are a little
> less useful and we don't really manage around it, so it has not really
> directed what happens out there. But as a concept, as something that isn't
> going to be going away, as part and parcel of this greater era of
> accountability – that is quite useful for us.

Overwhelmingly, study participants in all cases suggested the budget crisis in

government has further intensified the need for performance measurement as a means to

demonstrate measurable progress so that programs can defend resource allocations. A

representative from a Tobacco Control Program grantee program said,

> I was able to convey to the staff the idea that we always have to be able to
> say what we do. And we always have to be able to measure stuff, because
> the program has been cut so badly. And part of that [having their budget
> cut] was because we were not able to articulate what we did and what the
> numbers were and how the program and the state and the citizenry would
> be harmed by cuts to the program.

And from someone with the Preparedness Program,

> We've got to have information to defend this program, that's the bottom
> line. And, and maintain it. We're faced with funding cuts on a yearly basis
> and although it is highly funded as a program, its intent is to build the
> public health infrastructure so that we're in a much better position to
> respond to emergencies. And I think if we can demonstrate that [being
> prepared] through our measurement system, this program is going to be
> much stronger going forward.

<u>8.3.2.2 The types and choice of performance measures are affected by network variability, the political influence of network stakeholders, and the network structure which compromises control over outcomes.</u>

In these four cases, the networked implementation influences the types and choice of performance measures in three different ways. First, as evident in the description provided in section 8.2.2.5, the decentralized program implementation of the four cases results in extensive variation across the program grantees and their local level partners. Second, the network structure inherently results in a greater number of stakeholders with whom the performance measurement system must be planned and negotiated. And third, the network structure lessens the control that CDC and grantees have over outcomes. All three, network variability, the political influence of network members, and compromised control within the network, have implications for the types and choice of measures included in the performance measurement system. In this section, the first two factors are addressed; the last is discussed in section 8.3.3 below.

Network variability in all four cases is represented in terms of the extent and nature of the public health problem (e.g., infectious versus chronic disease, epidemiology, demographics of those affected); context (e.g., state laws, presence of advocates and/or adversaries); organizational priorities, resources, and capacity (e.g., grantees' activities, budget size, staffing, expertise); and network capacity (e.g., local-level capacity, infrastructure). Together, these kinds of variability affected the types and choice of measures included in the four performance measurement systems.

Variability in disease incidence and prevalence challenged CDC's ability to identify a *common* set of performance measures across grantees. As example, the

epidemiology of sexually transmitted disease varies across states – syphilis is much more

prevalent in Texas than Vermont. A representative of STD Program said,

> Absolutely there are different disease burdens. I think our challenge has
> been to identify a set of measures that is *useful* for the *majority* of project
> areas, both small and large and that's been an internal debate in the
> performance measure work group. There are a number of participants on
> the work group who will come out and say that small states don't matter,
> because the real outcome measure is reducing morbidity and reducing
> syphilis morbidity. In New England where all together they have 90 cases
> of P&S [primary and secondary] syphilis a year compared to Texas that
> has 90 in a month, there are advocates for not worrying about the small
> states…Some of these are specifically for syphilis high morbidity areas.
> There are a couple of measures that we've created so that the non-syphilis
> morbidity areas have something else to do. If you don't have syphilis, then
> you should be concentrating on gonorrhea and here's a measure to take a
> look at that.

Likewise, in regard to preparedness, New York is more vulnerable to an anthrax

attack than New Mexico. A participant with the Preparedness Program said,

> There's a lot of concern from grantees about, you know, that West
> Virginia doesn't need to be as prepared for a hurricane as Florida does. Or,
> Idaho doesn't need to be as prepared for a biochemical attack as New
> York City does. So what should our measures truly try to address? And is
> it as critical that certain states be able to do stuff as quickly or whatever as
> others?

For the three cases that confront variability in disease incidence and prevalence

(STD Program, Cancer Detection Program, and Tobacco Control Program), programs

have been forced to incorporate a degree of flexibility into their performance

measurement systems. For instance, in the STD Program, a grantee does not have to

provide data for performance measures that may not be relevant for their state, city, or

territory as long as they provide a rational justification. For the Cancer Detection

Program, if a grantee has a sample size less than ten for a specific performance measure

(e.g., percentage of women diagnosed with breast cancer with treatment started, target –

90% or greater), the indicator is not calculated given that the small numbers skew the

measurement. And in the cases of both the Tobacco Control Program and STD Program,

grantees will establish their own targets based on baseline data specific to their own state

or territory. From a participant with the Tobacco Control Program,

> See that's [setting targets] the challenge, especially when you fund 50
> states [laughing]. I don't know how those would be set. The target would
> have to be different for each program because every state has different
> resources, and different goals, and different activities, so I'm not quite
> sure [about setting a national target].

And from a staff person with the STD Program,

> We've opted against things like benchmarking or setting goals. We're
> going to have to revisit that because I think there's value in that. But
> because there's such variability between programs we didn't set a national
> goal [grantees set their own]. And to get people used to the idea of setting
> their own goals, realizing, OK, you know better than we do what your
> resources are, what your limitations are. So don't focus on how you
> compare to Texas or how you compare to Washington State. Focus on
> where you are now compared to where you could be and then set your
> goals accordingly. What we've [CDC] learned is that we need to do some
> training on how to set goals. They're all over the place.

Variability in context as well as organizational priorities, resources, and capacity

also had implications for the selection of performance measures. The unique needs of a

given grantee's population along with its resources and capacity will inform the program

goals it establishes and program activities it implements. As example, CDC has

established four national goals for the Tobacco Control Program, but grantees prioritize

them differently based on need, resource levels (e.g., CDC award size, state

contributions), or capacity (e.g., the network relationships needed to address a particular

goal area). Consequently, program activities implemented within a given state or territory

will vary as one grantee focuses on secondhand smoke and another on preventing the initiation of tobacco use among young people. The Office on Smoking and Health has found that even for the Tobacco Control Program that benefits from a strong science-base and identified best practices, implementation of program activities differs tremendously, even within a state (i.e., different communities within a state implement different activities). In part, this led to the initial development of 120 key outcome indicators which CDC is struggling to narrow down to a core set of national outcome indicators. Also, given the extensive variability in implementation efforts, the Office on Smoking and Health found it impossible to identify a limited and common set of process and output measures. One person representing the Tobacco Control Program said,

> They [Office on Smoking and Health] really wanted to identify process indicators that mapped onto short-term outcomes. But when we started going out and interviewing states and talking to them, we realized that every one of them was doing it *so differently* that coming up with one or two or three process indicators that all the states were using was going to be just a huge task – almost impossible.

This kind of variability in regard to processes and outputs was observed in three of the four cases, the Preparedness Program, STD Program, and the Tobacco Control Program. In contrast, however, the Cancer Detection Program is a service delivery program which is an anomaly among programs supported by CDC. For the Cancer Detection Program, grantees are responsible for the delivery of the same activities screening and diagnostic tests for breast and cervical cancer. Because the performance measures are almost entirely tied to the delivery of these services, the network's variability has had less affect on the types and choice of performance measures. All 68 of the Cancer Detection Program grantees work with local health care providers to recruit

and screen women, conduct follow-up to assure timely diagnostic follow-up, and ensure that women diagnosed with cancer initiate needed treatment. These activities are easily measured, regardless of the type of measure (i.e., process, short-term outcomes, and intermediate-level outcomes).

Next, variability in network capacity also affects the performance measurement systems for these four cases. Variability in network capacity is represented by differences in local-level expertise, resources, staffing, and infrastructure that affect the availability of data, the ability to collect and report valid data, and the adequacy of data management systems within a given state. Consider the range of funding, alone. For the Preparedness Program, grantee awards range from $330,000 to over $50 million; for the Cancer Detection Program, awards range from $75,000 to nearly $9 million. Resource level is a major factor affecting all aspects of an individual grantee's program, including staffing and data management systems, and the amount of resources available for local-level implementers.

The lack of local-level infrastructure and capacity in the Preparedness Program led to the Division of State and Local Readiness' decision to include only grantee-level measures in its current set of performance measures. For the Preparedness Program, STD Program, and Tobacco Control Program, variability in network capacity led all three programs to initially select what several participants referred to as "low hanging fruit" or, in the case of the STD Program, "get right at 'em" measures. For these programs, the idea of "low hanging fruit" or "get right at 'em" measures not only reflects the need to identify measures that are acceptable to most grantees in order to build buy-in, but also the need to select measures for which data are available and that are feasible for grantees to collect

and report given their diversity and capacity. That is, CDC must identify performance

measures for which data are available across the grantee network *and* measures for which

local-level partners have the capacity to collect and report the needed data. From the

Preparedness Program, a participant talked about their choice of performance measures

that are meant to be manageable, and therefore, acceptable to grantees.

> I think it's a good move for now from the measurement perspective – I
> think we considered these six 'low hanging fruit' and we thought that they
> [grantees] would too, so there were fewer complaints (laughing). But I
> think that it was partially to get their buy-in and so that they would agree
> that, 'Yes, these are reasonable measures, and yes, we do these things
> anyway, and it's not out of order for CDC to ask us for data on these [six
> measures].'

A representative from the STD Program also commented on the challenges of selecting

measures for which all grantees have the capacity to collect and report.

> I guess probably it's [variability across grantees] *the biggest challenge*
> because even being able to come up with the 12 or 14 measures that
> everybody [all grantees] can actually collect data on and do it in a
> reasonably [valid] sound way was a major accomplishment.

Finally, for all cases, variability limits the ability to compare performance across

grantees. The unique contextual characteristics of each grantee necessitate careful

interpretation of performance data; CDC staff consistently expressed the importance of

assessing grantees' performance based on their individual situation. From the STD

Program,

> I can't compare North Carolina and South Dakota. They're two entirely
> different environments, as far as what they're operating in, what they're
> trying to do. So I don't look at what one project area is doing versus
> another. I can't see the benefit to that. But I will look at the individual
> grantee, knowing what a project area does, and then look at those
> performance measures.

The network also influences the type and choice of performance measures through the political influence of its members, primarily grantees, but also other stakeholders in the network. In contrast to a contract, the cooperative agreement funding tool used by these four cases is premised on a collaborative relationship between CDC and grantees. In all cases, data suggest that grantees hold and exercise political influence in regard to performance measurement systems as well as in other areas of program-related policy and decision making. For instance, an observation of a performance measurement workgroup meeting revealed that the Division of STD Prevention management was extremely sensitive to grantee concerns about a proposed set of new performance measures. As each individual performance measure was discussed, representatives from the management team inquired about any concerns that grantees had expressed about that particular measure, giving those concerns significant weight in the considering whether or not to include the measure. In an interview, one person described how the Division management sometimes relents to pressure from grantees on issues related to the performance measures. "Well, they [grantees] didn't like some measure we had, and so they put pressure on him, that they don't want this as a measure. And he does give in to the field [grantees] a good bit."

Similarly, the Cancer Detection Program grantees will use political influence when necessary to affect decision making related to the performance measures. A representative from a grantee who was interviewed suggested grantees could have Congressmen calling CDC if some performance measures were introduced that grantees opposed.

> Some of the things that CDC's talking about now [ideas for future
> performance measures] that aren't performance measures of clinical things

– you may have Congressmen calling! I mean, it's different. These [current measures] are tangible – you can count days, you can count [screening] results, you can count the number of women who have had cancer diagnosed and, whether or not they got into treatment.

A staff person with the Cancer Detection Program also remarked about the potential backlash from grantees if benchmarks for the performance measures were made more difficult.

> [Laughing] I think we'd all get shot if we raised the benchmark because we've been standing behind them so long that they're adequate the way they are. Because in the same way we refused to lower them, I think we'd get in much more trouble with grantees if we tried to raise them for the same reason.

For the Preparedness Program, federal-level stakeholders have been influential in shaping the performance measures through the imposition of specific policies that require related performance measures and even specify the type of measure or measure itself. As detailed in chapter 4, policies such as the Pandemic and All-Hazards Preparedness Act of 2007 and Homeland Security Presidential Directive #21 have significant implications for the Preparedness Program's performance measurement system. Homeland Security Presidential Directive #21 states,

> (a) within 270 days after the date of this directive, (i) publish an initial template or templates meeting the requirements above, including basic testing of component distribution mechanisms and modeling of template systems to predict performance in large-scale implementation, (ii) establish standards and performance measures for State and local government countermeasure distribution systems, including demonstration of specific capabilities in tactical exercises in accordance with the National Exercise Program, and (iii) establish a process to gather performance data from State and local participants on a regular basis to assess readiness; and

> (b) within 180 days after the completion of the tasks set forth in (a), and with appropriate notice, commence collecting and using performance data

362

and metrics as conditions for future public health preparedness grant funding.

A participant remarked,

> I think it was in the fall or maybe December [2007] that HSPD-21 was approved. But it gave us six months to develop measures for countermeasure distribution. The data, the first report has to be within a year. I mean, it's just totally unreasonable.

The Preparedness Program grantees also exert political influence around the selection and development of performance measures. As observed at both a conference and a performance measurement workgroup meeting, grantees were forceful in sharing their views about the performance measurement design and the choice of measures that would be acceptable to the grantees.

**8.3.3 How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement?**

8.3.3.1 Dependencies on both vertical and horizontal network partners diminish CDC and grantee control over outcomes, restricting the design and impact of performance measurement.

As reflected in section 8.2.2.3, CDC's control over outcomes weakens as the network descends both vertically and horizontally. Diminishing control over implementation and subsequent outcomes due to network structure was observed in all four cases, albeit to different degrees. First, in regard to vertical network relationships, CDC and grantee control is exerted based primarily on institutional and funding relationships. However, a central challenge of the vertical structure is CDC's diminishing level of influence on program implementation (and outcomes) with each step down the chain. While these four CDC programs have some formal authority over grantees based on the cooperative agreement or grant, programs have no direct relationship with vertical

partners at the local-level. Instead, CDC is dependent on grantees' control over local-level implementation which will be influenced, in part, by grantees' choice of funding tool and model of decentralization. From the Cancer Detection Program,

> I think it [control over the local level] depends on how they're [grantee programs] structured. I know, for instance, that a state that works with county health departments – maybe where the county health departments are their own unique entity – they [state grantee] don't have as much control over service delivery as a grantee that has individual contracts with providers where the state could cancel their contract, or decrease their funding, or something like that.

A representative from the STD Program also discussed how the infrastructure within a state affects control over the local level.

> Their [grantees] administration is different. In some areas the disease intervention specialists, the DIS, who would be doing the interviewing, they're state employees in a central location. And in other areas they're [DIS] maybe county employees. And so those kinds of things make it a little more difficult [to control implementation]…there is so much variation in staffing and the types of staffing and it just makes it more difficult.

Of the four cases, the Cancer Detection Program is most adept at using management tools to enhance control and authority within the vertical network, even down to the local-level. CDC closely monitors clinical service delivery through its large monitoring dataset and is able to intervene with grantees as necessary. The Division of Cancer Prevention and Control has not shied away from imposing sanctions on grantees (e.g., withholding funds, requiring a corrective action plan) when quality of care issues are involved. A representative of the Cancer Detection Program noted,

> In our case, we have the MDEs [minimum data elements] that we're always analyzing. We're always looking at it [the data]. We're always providing feedback to the states and working with them. And we do impose sanctions, if I could call it that, when a program that we fund gets too far out of compliance.

Use of the Cancer Detection Program performance measures as part of the budgeting formula also serves as an incentive for grantees to enhance *their* mechanisms of control over local-level performance. Several participants mentioned grantees' use of performance-based contracts and reimbursement policies as means to exert influence over local-level partners.

In contrast, the Division of STD Prevention has afforded greater latitude to STD Program grantees over the years leading to a culture of grantee autonomy that has been difficult for CDC to shift. A participant described the tension between the Division of STD Prevention and its grantees, "You know, [grantees think] it's *my* money, and I get to do this with it, and I'm running *my* program. But, aren't there expectations?" This institutionalized culture has compromised CDC's influence over grantees. The Office on Smoking and Health also provides extensive leeway to its Tobacco Control Program grantees as reflected in the statement below.

> We [CDC] have traditionally been non-prescriptive. We want states to use our money for what they need our money for. If that [a particular goal area] is not a priority to the state, then we may not require them to do anything on it. We want the programs to be as effective as they possibly can and as specific as they possibly can to the needs of their state.

Dependencies on horizontal partners impose the greatest challenge for control and authority in these four networks given that the relationships are typically unfunded and informal – that is, neither CDC nor grantees have any formal authority over their efforts. Horizontal partnerships serve different purposes in the networks of these four cases. For the Preparedness Program and Tobacco Control Program, outcomes are jointly produced; therefore, significant dependencies emerge between grantees and the horizontal partners

needed to achieve results. As noted earlier, an effective response to an emergency,

whether a hurricane or bioterrorist attack, cannot be addressed by CDC grantees alone –

they are dependent on partners in varied sectors and at multiple levels of government to

contribute to the effort. But without formal authority, Preparedness Program grantees

must rely on their ability to *influence* horizontal partners, an endeavor that itself can be

made more difficult when priority and goal conflicts exist. The Director of a state grantee

program for the Preparedness Program said,

> You're dependent on all these programs to have the agency be prepared.
> You need to have them come to the table for planning, and for
> participating in training and exercises. But you're just another *"ask"* on
> top of the work they already have to do.

Like the Preparedness Program, the Tobacco Control Program operates from a

collaborative, networked model. Tobacco control will only be achieved through the

implementation of multiple, complimentary interventions and community mobilization

that address diverse areas such as policy, public education, and clinical care. Because

CDC funds are primarily retained by the grantees to support infrastructure, data revealed

less about their relationships with network partners, including grantees' ability to exert

control over either vertical or horizontal partners. However, one can assume that similar

issues of control may exist.

The STD Program and Cancer Detection Program grantees along with their local-

level vertical partners rely on horizontal network members to facilitate access to target

populations and support the integration of service delivery. As in the cases of the

Preparedness Program and Tobacco Control Program, formal control over horizontal

partners is absent. For STD Program grantees, the ability to influence them is further

challenged by mission and goal conflicts. One interviewee noted,

> There's challenges in working with entities whose mission is not public
> health. You know, even those of us who have set up jail screening
> programs, or worked with provider visitation, or worked with laboratories,
> it's still a challenge that comes up. Convincing people to do our work for
> us without any money is what we're doing and [helping them] understand
> how they fit into the public health equation is difficult. It's a challenge.

In these cases, the loss of control over implementation and outcomes imposed by

the network structure has implications for their performance measurement systems. In

particular, the Preparedness Program, STD Program, and Cancer Detection Program all

experience grantee resistance to performance measures viewed as "outside their control,"

specifically, those indicators for which grantees (or their local-level implementers) are

dependent on horizontal partners over whom they have no authority. In the statement

below, a representative with the STD Program argues that grantees should not be held

accountable through performance measurement for processes that are not "our work."

> If we are trying to measure our work, the STD Program, it makes more
> sense to me to measure from date of assignment rather than specimen
> collection. [Grantees can't control how long it takes laboratories to
> process and report a case to the health department]. It [influencing the
> laboratories] is potentially a piece that we can help to better, but it isn't
> actually our work.

A staff person with the Cancer Detection Program described grantees' concerns over a

performance measure they viewed as "outside their control."

> If a referral is made [by a network partner] to a woman with an abnormal
> PAP to go to the B&C program [an Cancer Detection Program-funded
> provider] down the street, they [grantee] don't have control over how long
> it takes that woman to show up, and that delays their follow-up time [a
> performance measure]. And we counted the clock [calculated the measure]
> from the time that woman was diagnosed with an abnormal result, not

from the date they entered our program. So programs felt disadvantaged because they can only control the timeframe from the time the woman walks in the door, as opposed to when they were originally screened and diagnosed to have an abnormal test.

Staff in the Preparedness Program and STD Program also expressed concern about the reliance on horizontal partners for data collection and reporting. For instance, STD Program grantees are dependent on jails and juvenile detention facilities to screen new admittees. Given that these partners are typically unfunded by STD Program grantees, concerns about both data access and data quality arise. One participant discussed why grantees are opposed to measures that depend on jail personnel.

I think it's pretty clear why grantees don't like that – you're calling it performance, which is a reflection on them. They may not be performing well because someone in their community isn't supporting them or there are things that are beyond their control. And other people just have the luck of being in a place where they have a cooperative prison system or something.

In response to grantee resistance, all three programs the Preparedness Program, STD Program, and Cancer Detection Program – have altered one or more of their measures to strengthen grantee control over performance. As discussed in more depth in the individual case chapters, the Division of State and Local Readiness has added a criterion of "under public health's control" for the selection of future performance measures for the Preparedness Program; STD Program grantees rejected several proposed measures for 2009 that were viewed as "outside their control;" and the Division of Cancer Prevention and Control recently revised the calculation of a measure to reflect performance that is strictly under the control of agencies in the vertical network. In all cases, the consequence has been a tendency to select measures that are more closely tied

to the work of the grantee or those in the vertical chain. Typically, these include process, output, and short-term outcome indicators. A representative from the Preparedness Program described how CDC has tried to revise measures in ways that maximize control for the grantee.

> There's a whole lot of things that they [grantees] don't have control over because of working with such diverse partners. For example, if they're doing an exercise with five other partners at the state, but they're not leading it [the exercise], that could cause delays in certain things. It wouldn't be their fault. So we had to really play with the language to find out how we can phrase this measure so that they provide what they're responsible for verses depending on somebody else to facilitate something. I think that's a good move.

Overall, the network implementation structures represented in these cases lead to diminished CDC and grantee control over outcomes. Grantees, whose buy-in is needed to assure data quality and data use, are resistant to the inclusion of performance measures that rely on unfunded, informal horizontal partners. For the Preparedness Program and Cancer Detection Program that will use performance data to inform budgeting decisions[56], grantees seem to have a "zero-tolerance" for measures viewed as "outside their control" given the implications of the performance data. As a consequence, the choice and types of measures included in the performance measurement systems for the Preparedness Program, STD Program, and Cancer Detection Program have all been constrained by the need to assure that grantees have "control" over related performance.

---

[56] The Division of Cancer Prevention and Control currently uses seven of its 11 performance measures as part of a broader, funding formula. The Division of State and Local Readiness will begin using performance measures to inform budgeting in 2010 based on the requirements of some federal policies (e.g., Pandemic and All-Hazards Preparedness Act of 2007).

A representative from the Preparedness Program articulated the important compromise this limitation presents in a networked context,

> In the long term, I don't know if we can move that way [including measures over which grantees have less control]. I don't know if we can get the buy-in from them [grantees] if their situation at the home front doesn't change to where they feel like, 'Yeah, we can be responsible and accountable for measures that really test that synergy of working among different partners. I think that's important. That's critical. Because that's what makes or breaks a good response in public health or any other area.

## 8.4 Summary

Cross-case findings suggest that networked public management has important implications for the design of performance measurement systems in the four cases studied here. While the involvement of vertical and horizontal partners is essential to effectively address these complex public health problems, networks impose limitations and force compromises in the design of performance measurement systems. For these four programs, networks make the attribution of outcomes to specific agencies or programs difficult, if not impossible. For the Preparedness Program and the Tobacco Control Program, where program outcomes are dependent on collaborative network efforts, accountability becomes fragmented. In addition, networks create dependencies that challenge grantees' ability to control performance for important outputs and outcomes, especially when they are reliant on unfunded, horizontal partners over whom they have no formal authority.

Variability across the grantee network in regard to the extent or nature of the public health problem, context, organizational features, and network capacity influenced the types and choice of measures selected, challenged CDC to define common measures,

set national targets, and forced programs to incorporate flexibility into their performance measurement systems. Finally, the political power of network stakeholders, especially the influence of grantees and federal policy initiatives, has shaped the performance measurement systems for these four cases. Even for the Cancer Detection Program where the nature of the program (i.e., service delivery) has facilitated the selection of common performance measures and primary reliance on vertical partners has limited dependencies, variability across the network and stakeholders' influence have affected the design, choice of measures, and target setting.

# CHAPTER 9

# DISCUSSION

This chapter begins with a brief review of the study purpose, theoretical framework, research questions, and design. Next a summary of the findings are presented situated within the broader literature related to networked public management and performance measurement. Third, implications for theory, practice, and policy are provided reflecting on the theory building of Jennings and Haist and considering the relevance of this study for federal-level performance measurement in public health. Fourth, recommendations for future research are provided. The chapter closes with a summary of the study's conclusions. As in chapter 8, shortened program names will be used for the four cases (see table 27, chapter 8).

## 9.1 Summary of Study Purpose and Design

The purpose of the study was to investigate the implications of networked public management on the design, implementation, and utilization of federal performance measurement systems. My intention was to contribute both to the practice of performance measurement and to expand early theory building initiated by Jennings and Haist (2004). While much has been written about the importance of performance measurement for managing in the public sector, little research has addressed its practice in settings where implementation is carried out through collaborative networks. In particular, there is a lack of empirical research exploring how networked contexts affect the design, implementation, and utilization of performance measurement. Given the limited understanding of this topic, a qualitative, descriptive study design was utilized.

Qualitative methods are most suitable when a subject is insufficiently understood and researchers are exploring "how" and "what" (Creswell 2007).

Although exploratory in nature and not intended to test a particular hypothesis, the research was informed by a set of 25 hypotheses proposed by Jennings and Haist (2004) related to the impact of performance measurement (appendix A). Jennings and Haist offer a set of hypotheses about the potential impact of performance measurement given varied conditions and factors, with the intent to set a theoretical foundation for performance measurement and spur empirical study. While none of Jennings and Haist's 25 hypotheses explicitly addresses decentralized or networked implementation structures, two related to agency type were selected to guide this study given their emphasis on organizational context and function:

- Hypothesis #13: "The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed." (p.185)

- Hypothesis #14: "Measurement will be more common and will have greater impact when agencies have greater control over outcomes." (p.185)

These two hypotheses provided the theoretical focus for the study; from them, three research questions were developed to further guide the investigation, each centered on the role of networked public management.

- How does networked public management affect the observability of CDC program outputs and outcomes?

- How does networked public management influence CDC's use of performance measurement and the types of performance measures used?

- How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement?

Given the exploratory nature of the research, a multiple, instrumental case study design was used. Four public health programs funded by CDC and implemented nationally through vertical and horizontal network structures were included as unique cases. For the four cases, data collection included 50 in-depth interviews with 52 individuals, formal review of 57 documents and websites, and 12 formal observations. Standard qualitative analysis techniques, including those derived from grounded theory (Charmaz 2006) were applied. A detailed summary of the research methodology is provided in chapter 3. Results of individual case analysis are presented in chapters 4 through 7; findings are summarized in appendix N. Cross-case findings, presented in chapter 8 and organized around each research question, provided greater insight by studying the similarities and differences across the four cases in how the networks affect design and implementation of federal-level performance measurement systems.

### 9.2 Summary of Findings

Cross-case findings suggest that the networked implementation structures for these four federal-level public health programs have important implications for the design of the performance measurement systems. Specifically, the performance measurement systems were affected by four consequences of the implementation networks:

- the political influence of collaborative stakeholders,
- network variability,
- dependencies on voluntary, horizontal network partners to achieve outputs and outcomes, and

- jointly produced outcomes that compromise assigning agency-specific attribution and accountability.

While these four factors did not deter the *use* of performance measurement by any of the programs, all had important consequences for the development and subsequent design of the performance measurement systems, including limiting the choice and types of measures, level of measurement, potential uses of the measures (e.g., accountability, comparing performance across grantees), and resources needed to implement and support the system. In addition, results suggest that the adoption of these systems across vastly decentralized implementation structures takes considerable time. After briefly describing the networks involved in the four cases studied, I will address each of the four factors listed above in turn. The section closes with a discussion about CDC's decision to use performance measurement in these four cases.

**9.2.1 Network Description**

The networks of these four public health programs involve formal and informal network partners, some intergovernmental, but others representing different agency types (e.g., community-based organization) and sectors. Although all four programs involve networks with both vertical and horizontal dimensions, the Preparedness Program and the Tobacco Control Program are more reliant on horizontal partners to achieve outcomes than the STD Program or the Cancer Detection Program. In general, these networks maintain traditional hierarchical, vertical relations based on funding relationships, but they also extend horizontally in order to reach high risk populations, integrate service delivery, and facilitate coordinated prevention strategies more effectively. This finding is consistent with earlier descriptions on network structure (Heinrich, Hill and Lynn 2004;

Milward and Provan 2004), although this study identified ways in which programs

benefited from these horizontal relationships that had not been discussed previously (e.g.,

extending program reach to high risk populations, facilitating coordinated prevention

strategies). During the past several years, researchers have proposed various typologies of

network types (Agranoff 2003; Milward and Provan 2006); these cases may best reflect

service implementation networks characterized by Milward and Provan (2006).

 While public health has traditionally relied on vertical relationships with state and

local partners to implement programming appropriate to the needs of individual

communities, the involvement of horizontal partners in these four cases may be explained

by the growing recognition about the complexity of the problems faced (e.g.,

preparedness, health care system issues) and the need to bring multiple sectors,

nontraditional partners, diverse strategies, and greater resources to bear. This supports, in

part, O'Toole's (1997) proposition that awareness of more "wicked" social problems is

contributing to increased networks in public management.

### 9.2.2 The Political Influence of Collaborative Stakeholders

 As detailed in chapters 1 and 2, network governance represents a shift from

traditional, hierarchical government to an approach that recognizes the public sector's

interdependent relationships with vertical and horizontal partners at multiple levels (Kettl

2002; Stoker 1998). For the four cases in this study, a collaborative relationship between

CDC and its grantees is reinforced through the choice of policy tool for funding, a

cooperative agreement. Although CDC maintains some authority, grantees enjoy a level

of autonomy that allows them to structure program implementation in ways appropriate

to community needs and priorities. In regard to local-level partners, CDC concedes direct

control entirely. And the extent of grantees' control and authority over local-level implementers may be dependent, in part, on the funding tool used, the existing state, tribal, or territorial infrastructure for implementation, and relevant laws or policy[57]. CDC and grantees both lack formal authority over voluntary, horizontal partners that may exist at multiple levels, although dependencies exist as their participation is needed to achieve program outputs and outcomes. Some suggest that the interdependencies, whether between vertical partners or vertical and horizontal partners, created in a networked environment lead to power differentials, rendering networked governance an inherently political concept (Lynn, Heinrich, and Hill 2000; Peters and Pierre 1998; Stoker 1998). This is the case as it applies to the development of the performance measurement systems studied here. Findings suggest that grantees and their partners hold substantial political influence in their relationship with CDC that forces a degree of bargaining and compromise in the design of the performance measurement systems. For the majority of cases studied, the development of the performance measurement systems is a negotiated and collaborative endeavor between CDC, the grantees, and, in some cases, other stakeholders.

The importance of involving stakeholders has long been recognized by experts of performance measurement to ensure that systems are feasible and that buy-in, needed to support data quality and data use, is achieved (Hatry 2007; Poister 2003). But results from this study also suggest that for these cases, network stakeholders at multiple levels

---

[57] The relationship between grantees and their local-level partners was not the focus of this study. A limited number of stakeholders representing the grantees were included in the study given limits on the number of non-federal employees that could be included in the research. Consequently, less is understood about the authority relationships between grantees and their local-level, funded partners (sub-grantees).

exercise political influence that shapes aspects of the performance measurement systems

including the choice and types of measures and system design. Grantees are willing to

flex their political muscle to influence CDC decision-making related to the performance

measurement systems, especially if performance data will be used to inform budgeting.

This finding lends support to another of Jennings and Haist's (2004) hypothesis #12:

"Significant mobilization of interests adverse to the measures will reduce the impact of

performance measurement" (p.182). Findings suggest that in these networked contexts, a

participative process of system design was required. Like Radin (2006) who dismisses

the one-size-fits-all approach for performance measurement, the network context

demands a more nuanced, collaborative, and negotiated strategy to develop systems

appropriate to the needs not only of CDC, but to a variety of stakeholders. This idea of a

collaborative approach to the development of performance measurement systems has

been forwarded by Goddard and Mannion (2004) as more appropriate in a networked,

horizontal environment (see chapter 2, table 6).

### 9.2.3 Network Variability

In this study, network variability was identified as an important consequence of

network structure and a key factor influencing the performance measurement systems.

Variability within the networks of each program was observed in the extent and nature of

the public health problem; the state, tribe, or territorial context; grantee priorities,

resources, and capacity; and network capacity. This variability challenged programs to

identify common measures, led to the selection of "low hanging fruit" for initial sets of

measures, necessitated that programs incorporate a degree of flexibility in their

measurement systems to accommodate the diversity across grantees, affected the level of

measurement for one program, required programs to incorporate extensive data quality systems, and compromised the utility of comparing performance across grantees within a particular program (e.g., Preparedness Program, STD Program). Variability continues to challenge programs once performance measures are developed and implemented; in particular, the sheer number of grantees along with variability in network capacity contributes to on-going challenges in regard to data quality and validity. Studies of GPRA by the Government Accountability Office (GAO) and others have identified similar issues related to variability for programs implemented through decentralized intergovernmental structures (GAO 1997a; GAO 1998a; GAO 2004; Mihm 2004).

The extent of variability within each of these four programs should not be underestimated. Not only does the Cancer Detection Program work with 68 different and unique grantee programs, but an additional 22,000 local providers are involved in service delivery and, consequently, with the collection and reporting of data that support the performance measures. This magnitude of decentralization, the related variability, and the implications for performance measurement cannot be easily dismissed. As Fossett, Gais, and Thompson (2001) contend, "Proclaiming the virtues of mission-driven federalism is, of course, one thing and success in implementing it quite another. Intergovernmental arrangements complicate virtually all aspects of performance management – agreement on key goals, the development of indicators, the timely collection of pertinent and valid performance data, the interpretation of these data, the implementation of an incentive system (e.g., rewards for strong performers), and more" (p. 208).

**9.2.4 Dependencies on Voluntary, Horizontal Network Partners to Achieve Outputs and Outcomes**

Study results provide empirical support for some of the critical challenges related to managing in networks that have been described in the literature. In nearly all cases, grantees struggled to engage voluntary, unfunded network partners needed to achieve outcomes. Grantees in Preparedness Program found it difficult to bring other partners to the table for preparedness planning; local-level public health agencies with STD Program wrestled to make in-roads with jails to allow for STD screening; and grantees with the Cancer Detection Program were frustrated in attempting to ensure timely referrals from unfunded, partner agencies for women needing diagnostic services. O'Toole and Meier (2004) and Agranoff (2003) have dubbed this the joint production problem -- the challenge of achieving program success in an environment that involves voluntary collaboration with actors over whom public managers have little formal authority. Among the four cases studied here, there was also evidence of mission and priority conflicts that challenged collaboration further.

In this study, the consequence of the joint production problem was grantees' effective dismissal of performance measures reflecting jointly produced outputs or outcomes. Such measures were viewed as "outside of the control" of grantees given inherent dependencies on the compliance of others over whom they had no formal authority. Such power dependencies created in networked environments have been described by others (Salamon 2002; Stoker 1998). In several cases, grantees successfully argued that it was "unfair" to include such measures – the Division of State and Local Readiness had gone so far as to include an *inclusion* criteria for measure selection that

assured proposed indicators were "within the control of public health." By effectively

excluding important outputs and outcomes given this lack of control, potential limitations

on the impact of performance measurement are created. In particular, performance

measures will more likely reflect process and short-term outcome measures more closely

tied to the efforts of those funded in the vertical chain.

### 9.2.5 Jointly Produced Outcomes that Compromise Assigning Agency-Specific Attribution and Accountability

As discussed in chapter 2, accountability is another major challenge in networked

environments. The dilemma of accountability in a governance framework is well

established in the literature (Kettle 2002; O'Toole 1997; Peters and Pierre 1998; Stoker

1998). In the four cases studied here, the layers of decentralization and joint production

of outcomes both contribute to a fragmentation of accountability structures. While all

four cases prioritized accountability as a primary purpose for their performance

measurement systems, programs either had to re-conceptualize the notion to one of

"shared accountability" as seems to be the case with the Tobacco Control Program, or

select measures more closely tied to the work of the grantees for which agency-based

accountability can be assigned. The latter represents another potential limitation to the

impact of performance measurement; similar to the joint production problem, process and

short-term outcome measures may predominate.

Among the four cases, performance measurement seemed most effective for the

Cancer Detection Program, a service delivery program. For this program, issues of

variability are diminished because grantees provide, for the most part, the same set of

services. In addition, dependencies are limited given primary reliance on vertical, funded

network partners. And *accountability* problems are minimized since providers in the vertical chain, over whom there is some authority, are mainly responsible for producing program outcomes of interest. Although the case benefits from these attributes, the extent of decentralization (i.e., over 22,000 local providers) is significant and presents obstacles for performance measurement. But to its credit, the Division of Cancer Prevention and Control has been relatively successful in managing these challenges. The Division effectively uses policy and management tools (e.g., cooperative agreement, a large monitoring data base) to preserve authority and support accountability within the vertical chain. Although not the focus of this study, there is some evidence that some grantees of the Cancer Detection Program use performance-based contracts or reimbursement strategies that strengthen *their* authority over local-level providers, thus preserving vertical authority throughout the implementation chain. In addition, the Division of Cancer Prevention and Control has made an enormous investment in its performance management system (e.g., data contractor, available software, extensive monitoring system, data validation study) that contributes to institutionalized data use and to strong data quality and validity, perennial challenges in such decentralized structures.

But, as a service delivery program, the Cancer Detection Program is an anomaly in public health at CDC; most programs are more similar to the other three cases -- grantee variability is extensive, outcomes may be more difficult to measure and dependent on voluntary, network partners to produce, and the ability to hold individual agencies accountable is difficult. These cases represent more complex situations for assessing performance than within a hierarchical context where accountability structures are clear (Frederickson 2003; Goaddard and Mannion 2004).

### 9.2.6 Use of Performance Measurement

Interestingly, cross-case analysis suggests the compromises to the potential impact of performance measurement imposed by the network structure were not a factor in determining whether or not to *use* performance measurement. Instead, findings suggest an *expectation of use* implicit in the administrative and political climate that drives the uptake of performance measurement by staff with these programs. This finding supports others' view that "performance" has been embraced by public managers, often without critical deliberation (Behn 2003; Blalock and Barnow 2001; Coplin, Merget, and Bourdeaux 2002; Grizzle 2002; Radin 2006). Although Jennings and Haist (2004) hypothesize that the use of performance measurement will be tied to the observability of outputs and outcomes, the authors also acknowledge that political and social imperatives can take precedence in demanding performance-based accountability. This seems to be the situation here. Participants across all four cases reiterated the adage, "what gets measured, gets done" as endorsement for their commitment to develop and implement federal-level performance measurement systems to support accountability. At CDC, this expectation of use may reflect the institutionalization of performance measurement for assessing public health programs. Similar to the Office of Management and Budget using GPRA measures to hold federal agencies accountable, CDC is developing program-specific performance measurement systems to hold grantees accountable and improve programs, although without critical assessment of its potential impact as a monitoring tool given the complexity of the program or its networked context.

## 9.3 Implications for Theory, Practice, and Policy

Although these research findings are not generalizable beyond the four cases included in the study, results nonetheless offer insights for public health, public management, and public policy. This exploratory study provides empirical evidence for potential limitations in the use of performance measurement outside of traditional, hierarchical contexts where its practice has primarily been described. As the adoption of networked governance expands and federal-level public health leaders continue calls for greater integration of program strategies, a better understanding of how performance measurement works in these settings and its potential limitations as a management tool is needed. In this section, implications of this study for the theory of performance measurement, its practice, and related policy are presented. Particular attention is given to federal-level performance measurement in public health.

### 9.3.1 Implications for Theory and Research

As summarized in 9.1, two hypotheses from Jennings and Haist (2004) guided this study.

- Hypothesis #13: "The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed." (p.185)

- Hypothesis #14: "Measurement will be more common and will have greater impact when agencies have greater control over outcomes." (p.185)

With these two hypotheses, Jennings and Haist theorize that performance measurement will vary in its impact depending on the type of agency involved. Using

384

Wilson's (1989) typology of agency-type, the authors suggest that performance measurement used in production agencies, those organizations with observable outputs and outcomes, will have greater impact than when applied in craft, procedural, or coping agencies, all of which have either outputs or outcomes (or both) that are not observable. Jennings and Haist define observability in two ways – the ability to measure outputs and outcomes and the ability to attribute outcomes to the efforts of a particular agency. As noted in chapter 1, none of the hypotheses developed by Jennings and Haist, including the two used to guide this research, explicitly addresses the potential influence of decentralized program implementation structures characteristic of the four cases included here. Consequently, this study has provided the opportunity to better understand how network implementation structures influence the design and implementation of performance measurement.

For the four cases studied here, findings suggest that network structures invite extensive political influence of stakeholders, introduce significant variability within the network, weaken control over outcomes given dependencies on horizontal network partners, and compromise observability (when defined as attribution) based on jointly produced outputs and outcomes  These aspects of the programs' networked implementation structures influenced the design of their performance measurement systems by limiting the choice and types of measures, level of measurement, potential uses of the measures (e.g., accountability, comparing performance across grantees), and resources needed to implement and support the system. Based on results, one may speculate that the networked structure compromises the impact of performance measurement.

Consequently, I propose adding three[58] new hypotheses, expanding upon those offered by Jennings and Haist (2004), in order to elaborate the implications of networked environments for performance measurement:

- Network variability (e.g., extent and nature of the problem, program context, network capacity) imposes limitations on the design and utilization of performance measurement.

- Dependencies on voluntary, horizontal network partners for program implementation weaken principal and agent control over outputs and outcomes and, consequently, reduce the impact of performance measurement.

- Networked implementation structures compromise the degree to which outputs and outcomes can be observed, limiting the inclusion of outcome measures and, therefore, lessening the impact of performance measurement.

**9.3.2 Implications for Practice and Policy**

Results of this research may have implications for the practice of federal-level performance measurement when applied to decentralized, public health programs implemented through collaborative networks. Given that this was a small, qualitative study, individual readers must assess the relevance of these recommendations for their particular program and setting. Several recommendations for practice are proposed including:

- Evaluate the decision to develop a performance measurement system and set reasonable expectations for what the system can accomplish.

---

[58] A new hypothesis related to the political influence of stakeholders is not proposed given that Jennings and Haist's hypothesis #12 discussed in section 9.2.2 above adequately encompasses this issue.

- Involve network stakeholders in the development process.

- Utilize available management tools to strengthen authority within vertical network structures.

- Allocate adequate resources to support the success of the system.

- Recognize that time is needed for the adoption of performance measurement systems in vastly decentralized and networked settings.

- Consider more rigorous evaluation methods if agency-based attribution is important to determine.

First, my findings suggest that programs may benefit from critically assessing aspects of the program itself and the implementation network that may influence the potential impact of performance measurement before endeavoring on the development process. Not unlike evaluability assessment, described by Wholey (2004), results here suggest that program staff should consider several factors in weighing the decision to develop a measurement system and setting expectations for its contribution to program management. Factors such as the complexity of the program (e.g., the ability to measure outputs and outcomes); the types of stakeholders with an investment in the program; the extent of vertical and horizontal decentralization that encompasses program implementation; the level and kinds of variability across implementation sites; dependencies on horizontal, voluntary partners needed to achieve outputs and outcomes; and the available management tools to support authority relationships within the vertical network dimension should all be considered. In better understanding the extent of program complexity and networked implementation context, program staff will gain insights for structuring the development process and set more reasonable expectations

about the types of measures that can be included in a system, the possible uses for the performance measurement system (e.g., accountability, program improvement, budgeting), resources needed to support a system, and the potential impact of the performance measurement system.

Second, results suggest that characteristics of performance measurement may differ from what has been traditionally described (i.e., use in two-party, hierarchical contexts where accountability structures are relatively straightforward) when applied in networked settings. In particular, conducting a participative and collaborative process to develop performance measurement is recommended in order to design systems responsive not only to federal-level interests, but also to the needs of state and local stakeholders. In addition, flexibility in system design may be needed to accommodate the diversity of grantees and their contexts. If performance measures are included that reflect jointly produced outputs or outcomes, developers must be sensitive to *how* data are used. For instance, grantees will likely oppose the use of performance data to inform budgeting decisions if they have minimal control over the performance on the measures. Further, agency-specific accountability may need to be recast as "shared accountability" for jointly produced outputs and outcomes.

Next, while participative approaches may be more appropriate for the development of performance measurement systems in networks, programs should use available mechanisms to support control and authority. As evident from this study, these opportunities will primarily rest within the vertical network relationships and involve funding tools, management practices, and partner relationships. Performance

measurement in networked contexts may necessitate achieving a balance between where authority can be exercised and where collaboration is needed.

Fourth, in vastly decentralized and networked programs like the cases studied here, performance measurement systems require a significant investment of resources to ensure data validity and data use. The Cancer Detection Program has spent over fifteen years and millions of dollars to develop its monitoring system that provides the data needed to calculate their performance measures – this commitment has contributed to a performance measurement system that supports accountability, program improvement, and budgeting. As described in chapter 6, the Division of Cancer Prevention and Control maintains a data contractor who supports data management software available to all grantees and provides technical consultants who are assigned to individual grantees to provide on-going data management support. In addition, a systematic, semi-annual data review process is institutionalized to support program improvement. Finally, training is routinely provided given staff turnover at both the state and local levels and quality assurance efforts are on-going. Consequently, leaders of national programs must carefully consider the resources needed to support an effective performance measurement system before committing to the effort. However, once in place, the performance measurement systems can be leveraged to help *build* capacity in the areas of monitoring and evaluation among state and local level partners.

Results also hint that, given the vast decentralization reflected in these four cases, adoption of performance measurement by grantees and local-level partners likely requires significant time. Although training and technical support may facilitate adoption, the development and implementation of data management systems, development of data

collection protocols and related technical assistance materials, provision of training, and

pilot testing will take time. Even when a data management software package is made

available, grantees may choose to develop their own in order to integrate with existing

state-based data systems. More importantly, building grantees' understanding and

appreciation of the system's value will take time and depend, in part, on the perceived

success of the system. Particularly challenging is making local-level partners aware of the

system. But as reflected in the following comment from a CDC staff person with the

Cancer Detection Program, this seems an important goal,

> I think communicating [the measures] to every level is essential…I'm
> talking about the fact that it is decentralized. You have a federal central
> level, and then you have a state, and then the state is divided up into a
> variety of counties and communities. And each one of those may have
> either a centralized provider group or individual providers. So in order to
> make this work [performance measurement], the lowest level of function
> has to be working on this in order to make it work at all.

Finally, if programs need to assess agency-based accountability – that is, to

confidently attribute specific outcomes to a particular program – other means of

evaluation should be considered. The methodological limits regarding performance

measurement and attribution are well established (Hatry 2001; Poister 2003).

Unfortunately, there remains an inaccurate assumption that performance measurement

assumes causality (Blalock and Barnow 2001; Frederickson and Frederickson 2006;

Radin 2006). This limitation must be more effectively communicated to both policy

makers and federal agency leaders.

## 9.4 Recommendations for Research

Although this was a relatively small, qualitative study, it offers insight for areas of future investigation that will help researchers better understand and define the practice of performance measurement in networked contexts. First, further qualitative studies that encompass all levels of the network may help illuminate factors that influence both the adoption of performance measurement throughout a given network and its impact. A focused study of state and local level partners involved in a federally-funded grantee program like the ones studied here, would be useful in expanding understanding about the design, implementation, and adoption of federal-level performance measurement.

Second, this study involved a particular type of network (i.e., collaborative). Other network types have been described (Agranoff 2003; Milward and Provan 2006)[59]. Similar descriptive studies of performance measurement as applied in other network types may also begin to shed light on these more challenging contexts.

Although not the focus of this study, these results point to two areas of particular interest adoption time and cost. What factors facilitate adoption of performance measurement in these networked contexts? What are the long-term costs of developing, implementing, and maintaining performance measurement systems for federally-funded grantee programs that are vastly decentralized? Quantitative studies, including cost analysis, may be most appropriate to address these questions.

---

[59] See chapter 2, section 2.6.3. Agranoff (2003) describes information networks, developmental networks, outreach networks, and action networks. Milward and Provan (2006) describe service implementation networks, information diffusion networks, problems solving networks, and community capacity building networks.

Finally, this study identified potential factors related to network structures that compromise the impact of performance measurement. These factors include the extent of vertical and horizontal decentralization, stakeholder involvement, network variability, the extent of dependencies on voluntary, horizontal partners to achieve outputs and outcomes, the extent of jointly produced outputs and outcomes, and funding mechanisms used within the vertical chain. These factors may contribute to modeling the potential impact of performance measurement in a variety of network contexts and testing my proposed hypotheses.

## 9.5 Conclusions

My interest to conduct this study stems from my experiences at CDC in managing and evaluating federally-funded public health programs implemented through our state and local partners. While strongly committed to community-based approaches, I also know, first-hand, the challenges network structures present for program management. At CDC, program implementation typically takes place two to three steps removed from the agency; thousands of local-level public and non-profit agencies may be involved in carrying out program activities or delivering services of a single national program. In addition, horizontal, voluntary partners are engaged at every level in order to provide more integrated service delivery, facilitate coordinated public health program efforts, extend our reach to vulnerable populations, and contribute resources.

Given the emphasis in government today on performance measurement as a tool of public management, my study aimed to better understand how networked public management affects the design, implementation, and utilization of federal-level performance measurement systems. With pragmatic intentions, but also an interest to

392

advance nascent theory building, this exploratory study contributes to a small, but emerging literature addressing performance measurement in the context of networked public management. Results suggest that for the national public health programs included in the study and others like them, networked structures impose important limitations on the practice of performance measurement and likely compromise its impact.

While further research inclusive of the full network is needed to better comprehend how these implementation contexts affect performance measurement, results provide insight for both practice and theory. In particular, in practicing performance measurement, a balance must be struck between leveraging existing authority relationships within the vertical chain while also recognizing the limits to performance measurement introduced by the variability within the network, the political influence wielded by network partners, the inherent dependencies introduced by network relationships, and compromises to agency-based attribution and accountability brought by jointly produced outputs and outcomes. In these contexts, a more collaborative and negotiated approach to performance measurement is called for -- one that reflects compromises made among all stakeholders in order to best meet their varied needs and accommodate this difficult implementation context. In addition, in developing theory that attempts to explain the impact of performance measurement, this study suggests that factors related to networked contexts must be considered.

# APPENDIX A

## JENNINGS AND HAIST'S TWENTY-FIVE HYPOTHESES FOR PERFORMANCE MEASUREMENT

## Twenty-five Proposed Hypotheses for Performance Measurement
## (Jennings and Haist 2004)

| | Code | Description |
|---|---|---|
| **Incentives and Accountability** | H1 | The more widespread the sharing of performance information, the greater the likelihood of positive effects on performance (p.176). |
| | H2 | The greater the belief of service providers and administrators that officials and citizens care about and will use performance information, the greater the likelihood of positive effects on performance (p.176). |
| | H3 | The more extensive the mechanisms of control available to principals, the more likely it is that agents will react positively to performance measurement (p.176). |
| | H4 | The more an organization's resources are dependent on measured performance, the greater the impact of performance measurement (p.177). |
| | H5 | The greater the competition faced by the organization, the greater the impact of performance measurement (p.177). |
| | H6 | Incomplete systems of performance measures are likely to distort agency behavior and reduce attainment of policy goals (p.178). |
| | H7 | Performance measures have greater impact when principals give them more attention, have more resources to shape agent behavior, and use the resources to pursue performance goals (p.178). |
| | H8 | Performance measures have greater impact when service providers and managers believe they are accurate reflections of performance (p.179). |
| | H9 | Performance measures have greater impact on outcomes when managers believe that the organization's performance on the measures affects their own economic well-being, career opportunities, power, or professional prestige and the organization's access to resources of authority and finance (p.179). |
| **Political Context** | H10 | The greater the agreement among principals with respect to goals and measures, the greater the effect of measurement (p.181). |
| | H11 | The greater the agreement between principals and agents with respect to goals and measures, the greater the impact of performance measurement (p.181). |
| | H12 | Significant mobilization of interests adverse to the measures will reduce the impact of performance measurement (p.182). |
| **Agency Type** | H13 | The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed (p.185). |

| | | |
|---|---|---|
| | H14 | Measurement will be more common and will have greater impact when agencies have greater control over outcomes (p.185). |
| | H15 | The impact of performance measurement depends on the design of a set of measures appropriate to agency task and goals (p.186). |
| **Organizational Compatibility** | H16 | The closer to the level of service delivery a performance measure is implemented, the greater the impact on output or outcomes (p.187). |
| | H17 | Performance measures that are compatible with the existing use and distribution of skills, tasks, and resources will result in greater and more immediate impact. Conversely, performance measures that require significant disruption of skill utilization or task organization may experience delayed or dysfunctional response in adoption (p.187). |
| | H18 | Performance measures will have greater impact in more homogeneous organizational settings (p.187). |
| | H19 | The more consistent performance measures are with an agency's culture, the greater the likelihood of a positive impact (p.188). |
| | H20 | Performance measures that enhance the mission of the organization will have a positive effect on outcomes (p.188). |
| | H21 | Agencies with low agreement among members regarding the organization's mission (agencies experiencing low policy consensus) are more likely to experience dysfunctional responses to performance measurement (p.188). |
| | H22 | Agencies staffed by professional whose norms are aligned with the goals of performance measures are more likely to experience positive response in adoption of performance indicators (p.188). |
| | H23 | Agencies staffed by professionals with strong norms of autonomy are more likely to experience delayed or negative response to adoption of performance measures (p.189). |
| **Agency Leadership** | H24 | Leadership that demonstrates high commitment to mission, effective goal setting ability, and ability to cope with (external) political and (internal) administrative challenges specific to performance measures will result in positive impact of performance measures (p.190). |
| | H25 | For organizations staffed largely by professionals, recognition of professional competence of the leader is critical to the impact of performance measures (p.191). |

# APPENDIX B

# SUMMARY OF EMPIRICAL RESEARCH ON PERFORMANCE MEASUREMENT

**Summary of Empirical Research on Performance Measurement (PM)**

| | Findings | Research Methods | Level of Study | Citation |
|---|---|---|---|---|
| **Features of PM System** | Vertical and horizontal features of PM systems are present | Systematic review | National-level, United Kingdom | Goddard and Mannion 2004 |
| **Adoption of PM** | Widespread adoption of performance budgeting among states based on legislative and administrative mandates | Survey | States | Melkers and Willoughby 1998, 2004 |
| | Widespread adoption of management for results systems among all 50 states | Survey | States | Mohnihan and Ingraham 2003 |
| | Adoption of PM influenced by technocratic and rational factors such as goal orientation, resources, etc.; Implementation of PM influenced by political and cultural factors | Survey | State and local | Julnes and Holzer 2001 |
| **Level of Use of PM** | 68% of respondents said either all departments or at least half of their departments use PM | Survey | State and local | Melkers and Willoughby 2005 |
| | 40% or less of cities used PM in "meaningful ways" to support management and decision processes | Survey | City | Poister and Streib 1999 |
| | 22% of cities used PM to track progress in meeting goals and objectives of strategic plans | Survey | City | Poister and Streib 2005 |
| | 33% of counties used PM of some type; only 20% demonstrated "high use" | Survey | Counties | Berman and Wang 2000 |
| | Although the use of administrative data is imperfect for the measurement of results, PM information is still useful | Survey | Federal | Heinrich 1999 2002 |
| **Factors Affecting Use of PM** | Technical capacity and | Survey | Counties | Berman and |

| | Findings | Research Methods | Level of Study | Citation |
|---|---|---|---|---|
| | of PM; widespread use increases satisfaction with the impact of PM | | | |
| | Managerial commitment to the use of PM, decision making authority, and training in PM techniques have a significant influence on its development and use | Survey | Federal | Cavalluzzo and Ittner 2003 |
| | Lack of authority to make changes based on performance information limits use | Interviews | Federal | Hatry, Morley, Rossman, and Wholey 2004 |
| **Types of Use** | PM found useful for communications and management activities such as strategic planning | Survey | City/County | Melkers and Willoughby 2005; Willoughby and Melkers 2000 |
| | PM found useful for managerial purposes | Interviews | Federal | Frederickson 2003 |
| | PM not used for budgeting and fiscal decision making | Survey`Interviews | City/County/ Federal | Melkers and Willoughby 2005; Willoughby and Melkers 2000; Frederickson 2003 |
| | Mixed results regarding the use of PM for program improvement | Survey | Federal | GAO 2004 |
| | Desire to use PM for improved decision making and accountability, not simply reporting | Survey | Cities | Poister and Streib 1999 |
| | PM used to trigger action, identify and encourage best practices, motivate, and plan | Interviews | Federal | Hatry, Morley, Rossman, and Wholey 2004 |

| | Findings | Research Methods | Level of Study | Citation |
|---|---|---|---|---|
| **Outcomes of PM** | Federal managers report having more outcome measures as required by GPRA, identified positive effects of GPRA on requirements for planning and reporting, and had more results-based goals in place | Interviews | Federal | GAO 2005 |
| **Challenges in Implementing PM** | Goal related Challenges:<br>• Goal conflicts<br>• Translating long term goals into annual performance goals<br>• Coordinating across federal programs<br>• Mission fragmentation | Interviews and Case Studies | Federal | GAO 2004; Frederickson 2003; GAO 2000a; GAO 1999; GAO 1997a; GAO 1997b |
| | Measurement Challenges:<br>• Developing outcome measures<br>• Identifying useful outcome measures<br>• Obtaining valid and reliable data<br>• Obtaining timely data<br>• Old data not useful<br>• Outcomes can take years to achieve<br>• Data not disaggregated | Interviews and Case Studies | Federal | GAO 2004; GAO 2000a; Hatry, Morley, Rossman, and Wholey 2004 |
| | Data Collection Challenges<br>• Collecting outcome data | Interviews and Case Studies | Federal | GAO 2004; GAO 2000a |
| | Causal Attribution Challenges:<br>• Inability to attribute changes in outcomes to particular programs | Interviews and Case Studies | Federal | GAO 2004; GAO 2000a; GAO 1997b; GAO 1996b |
| | Challenges of Decentralized Implementation Structures: | Interviews and Case | Federal | GAO 1998a; GAO 1998b; |

| | Findings | Research Methods | Level of Study | Citation |
|---|---|---|---|---|
| | • Compromises data quality and introduces variability in data collection across sites<br>• Difficult to control data collection collected by 3<sup>rd</sup> parties<br>• Challenges to aggregate data nationally | | | |
| | Capacity and Training Challenges:<br>• Limited understanding about how to use performance information<br>• Limited analytic and methodological expertise<br>• Lack of training in PM techniques | Interviews and Case Studies | Federal | Hatry, Morley, Rossman, and Wholey 2004; GAO 2004; GAO 2000a |

**APPENDIX C**

**INDIVIDUAL INTERVIEW GUIDE**

## Individual Interview Guide
## Performance Measurement Study


**INTRODUCTION**

Thank you for agreeing to participate in this interview. The interview will take about 60 minutes and I will be asking you about different aspects of the performance measurement system for [Program Name]. Your participation in this interview is voluntary and you can end the interview at any time. Data collected will be kept confidential. Results will be presented in aggregate for your program and quotes will not be attributed to any specific individual. The risks associated with your participation are minimal, however, you will have the opportunity to review this interview transcript before analysis is conducted and remove any statement(s) for which you have concern.

I am conducting this interview as part of my dissertation research. The research and interview have been approved by both the GSU and CDC human subjects review boards [pending]. Would you please read and sign a copy of the consent form if you are comfortable with it [provide at interview]. [If the interview is conducted by telephone, the consent form will be sent electronically in advance and the participant will be asked to fax a signed copy to the researcher in advance of the interview.] Please keep one copy for yourself.

Do you have any questions before we get started? May I have your permission to audio record the session?  ☐ Yes     ☐ No

1. What are the more immediate program outputs for the program?
   a. What factors affect those outputs?
   b. How does your program contribute to achieving those outputs?
   c. From your perspective, how much influence does your program have on affecting those outputs?

2. What are the long term goals or outcomes for the program?
   a. What factors affect those outcomes?
   b. How does your program contribute to achieving those outcomes?
   c. From your perspective, how much influence does your program have on affecting those outcomes?
   d. Has a program logic model been developed? [If so, request a copy]

3. Describe the implementation structure for your program.
   a. What are the critical activities that the grantee staff must conduct to achieve program goals?
   b. Typically, what types of agencies does the grantee collaborate with at the state or more local levels, both funded and non-funded, to achieve its goals and outcomes?
   c. How important is collaboration (at the grantee level) in achieving program goals and outcomes? If it's important, why is it important?

4. Describe your performance measurement system.
    a. What is the purpose or purposes of the system? [Probe on accountability purpose]
    b. How did you develop and select the measures (e.g., from a logic model, strategic plan? with stakeholder involvement?)
    c. Describe the measures. What types of measures are they (e.g., process, intermediate outcome)?
    d. What factors influenced the selection of measures and/or measure types?
    e. How did the implementation structure influence the design of the system? How did it influence measurement selection?

5. What were the most significant challenges you faced in developing your performance measurement system?
    a. How did you address the challenges?
    b. What compromises do you feel you had to make, if any?

6. How does the performance measurement system relate to GPRA or PART requirements for your program?
    a. How does your Division leadership perceive the performance measurement system as it relates to policy? What purpose do you think the leadership views the performance measurement system serving?
    b. How do you perceive the performance measurement system? What purpose do you think the performance measurement system serves?

7. How are you (CDC) using the performance data?
    a. What factors influence CDC's use (or non-use) of the data?
    b. How are grantees using the data?
    c. What factors influence grantees' use (or non-use) of the data?

# APPENDIX D

# INFORMED CONSENT FORM

**Title:** Federal-Level Performance Measurement in Networked Public Management Environments

**Principal Investigator:** Theodore Poister, Ph.D.
**Student Investigator:** Amy DeGroff, MPH

**Purpose**
Georgia State University and the Centers for Disease Control and Prevention (CDC) invite you to participate in a research study. Your participation is entirely voluntary. Please ask questions if there is anything you do not understand. The purpose of this research is to better understand the performance measurement systems developed for some of CDC's public health programs that are implemented nationally through decentralized program structures. The information will be used to better understand the issues related to developing and implementing performance measurement in these contexts.

**Procedures**
This research is being conducted by Amy DeGroff, a CDC employee and doctoral candidate at Georgia State University, as part of her dissertation research. The research involves approximately 32 individual interviews with staff and key stakeholders who have been involved in the development, implementation, and / or management of the performance measurement system. One focus group comprised of a subset of 8 interviewees will also be conducted. Based on your involvement with the [program name] performance measurement system, you have been selected to participate in a 60 minute interview. The interviews are one-time episodes, although the researcher may contact participants again for clarification of comments made in the interviews. Again, your participation is entirely voluntary. Interviews will be conducted at your convenience in a private office at CDC. If it is not possible to conduct the interview in person, the interview will be conducted by telephone; the researcher will conduct the interview from a private office.

**Confidentiality**
Any information you provide will not be associated with your name and the information you provide will be kept private to the extent allowed by law. The researcher will not reveal the names of participants to anyone. The interview will be audio recorded with your permission; individual names will not be included in any transcriptions, the dissertation, reports, or manuscripts. However, the larger program (e.g., the National Diabetes Prevention and Control Program) will be named and identified in transcriptions, the dissertation, reports, and/or manuscripts. The broad role of a person associated with a quote, such as CDC staff or CDC stakeholder, may be noted in the dissertation, reports, and/or manuscripts. The researcher will permanently delete all audio files from

interviews as written transcriptions of the interviews are completed. Transcriptions will be maintained in a locked filing cabinet.

**Risks**

Risks associated with participation are minimal but may involve inferred identity by colleagues who may have a close understanding of the program and its performance measurement system. You will have an opportunity to read the interview transcript and delete any statements for which you have concern. The researcher may contact participants again for clarification of comments made in the interview.

**Benefits**

Although you will not benefit personally by participating in this study, your participation will contribute toward a better understanding of the development of performance measurement systems at CDC and, optimally, to the improved practice of performance measurement at CDC.

**Voluntary Participation and Withdrawal**

Your participation in this research is entirely voluntary. You have the right to refuse to answer any question. You also have the right to end your participation at any time. Whatever you decide, you will not lose any benefits to which you are otherwise entitled.

**Contact Persons**

If you have any questions about this study, please contact Theodore Poister at 404-413-0129 or Amy DeGroff at adegroff@cdc.gov or 770-488-2415. If you have any questions about your rights as a study participant, you may contact Susan Vogtner in the Office of Research Integrity at Georgia State University at 404-413-3513. You may also contact CDC's Acting Deputy Associate Director for Science at 1-800-584-8814 and leave a message with your name, phone number, and refer to CDC protocol # 2550, and someone will call you back.

If you are willing to volunteer for this research and be audio recorded, please sign and date below.

You have been told about the study. You have been allowed to ask questions. You had all of your questions answered. You would like to be in the study.


_____          _____
Signature of Participant                                          Date


_____          _____
Signature of Principal Investigator                          Date

**APPENDIX E**

**DOCUMENT REVIEW FORM**

**Document Review Form**
**Performance Measurement Study**

| Date Reviewed | |
|---|---|
| Document Name | |
| Document Date | |
| Document Author | |

**General Description of Document:**

**Information about networked structure:**

**Information about observability of program outputs and outcomes:**

**Information about CDC's influence/control over program outcomes:**

**Information about performance measurement design:**

**Information on types of performance measures:**

**Information on use of performance measurement system and resultant data:**

**APPENDIX F**

**SUMMARY OF DOCUMENTS REVIEWED BY CASE**

**Summary Of Documents Reviewed By Case**

| | |
|---|---|
| **Comprehensive STD Prevention Systems** | Federal Register Announcement, vol. 69, no. 139, Wednesday, July 21, 2004 / Notices. Funding announcement  pp. 43595-43595 |
| | Performance Measures – Quick Reference Guide (2007) |
| | STD Prevention System Logic Model (2000) |
| | Division of STD Prevention webpage http://cdc.gov/nchstp/dstd/aboutdiv.htm  (accessed Nov 2008) |
| | 2007 Performance Measures Companion Guidance; CSPS, IPP, SE Program Announcement (2006) |
| | STD Performance Measures: 2005-2006 Data Report (2007) |
| | 2009 Performance Measures (2008) |
| | Charter of the Performance Measurement Workgroup, January December 2007 (2007) |
| | PM Evaluation Criteria Definitions (2001) |
| | ExpectMore, Program Assessment, National Center for HIV/AIDS, Viral Hepatitis, Sexually Transmitted Diseases, and Tuberculosis Prevention (accessed 2008) |
| | Performance Measures Learning Tour Guide (2006) |
| | 2009 Performance Measures: Feedback from Grantees (2008) |
| | HHS, CDC, Request for Applications 9S09-902 2008. Five year funding announcement for CSPS (2008) |
| | Hubley T. Lessons from a project to create performance measures for public health. *Evaluation and Program Planning*, Vol. 31 (2008): 410-415 |
| | Materials from the Maine state program about their 2007 Regional Program Performance Measures and Benchmarks (2007) |

| | |
|---|---|
| **Public Health Emergency Preparedness** | Public Health Preparedness Cooperative Agreement (AA154) Budget Period 08, Continuation Guidance (2007) |
| | Public Health Preparedness Cooperative Agreement (AA154), Budget Period 08, Performance Measures, Definitions and Guidance (2007) |
| | CDC Connects (CDC Intranet): "States, Localities Better Prepared" (February 2008) |
| | Key Findings from Public Health Preparedness: Mobilizing State by State; A CDC Report on the Public Health Emergency Preparedness Cooperative Agreement (February 2008) |
| | PHEP Evaluation Workgroup Meeting: January 24-25, 2008, Arlington, VA – Draft Summary Notes |
| | PHEP Cooperative Agreement: Performance Measures Data Quality Assurance and Improvement Process – Letter Report, November 27, 2007 |
| | Division of Strategic National Stockpile: Local Technical Assistance Review – October 2007 |
| | Division of Strategic National Stockpile: State Technical Assistance Review – October 2007 |
| | Historical Overview of the Evolution of the Evaluation Focus for the Public Health Emergency Preparedness Program –presentation (2007) |
| | PART Measures for PHEP (2008) |
| | Pandemic and All-Hazards Preparedness Act of 2006, December 19, 2006 |
| | Homeland Security Presidential Directive 21, October 2007 |
| | US Office of Management and Budget website: ExpectMore.gov (accessed May 2008) |
| | FY 2009 Congressional Justification – Department of Health and Human Services, Centers for Disease Control and Prevention (2008) |

| | |
|---|---|
| **National Breast and Cervical Cancer Early Detection Program** | Public Health Preparedness Cooperative Agreement (07-703) Budget Period 08, Continuation Guidance (2007) |
| | Public Health Preparedness Cooperative Agreement (07-703), Budget Period 08, Performance Measures, Definitions and Guidance (2007) |
| | NBCCEDP Data Management Web Conference Series: De-mystifying MDE Feedback Measurements (presentation hand-out) (2006) |
| | Core Program Performance Indicators (2006) |
| | MDE Data Quality Indicator Guide Template (2006) |
| | DP07-703 NBCCEDP Performance Measures assessment tool (2008) |
| | Program Data Review Worksheet Summary: NBCCEDP October 2007 MDE Submission (2008) |
| | ExpectMore.gov Detailed Information on the Chronic Disease Prevention Assessment (accessed April 2008) |
| | Update to the FY 2008 NBCCEDP Funding Process (presentation handout) (2008) |
| | NBCCEDP Policies and Procedures Manual (not dated) |
| | Funding Announcement for the National Comprehensive Cancer Control Program, the National Program for Cancer Registries, and the National Breast and Cervical Cancer Early Detection Program; United States HHS, CDC NCPCP; Billing Code: 4163-18-P (2007) |
| | Summary of MDE Validation Project – Internal CDC Report (2007) |
| | The National Breast and Cervical Cancer Early Detection Program: 1991-2002 National Report (http://www.cdc.gov/cancer/nbccedp/bccpdfs/national_report.pdf) (not dated) |
| | |
| **National Tobacco Control Program** | *Key Outcome Indicators for Evaluating Comprehensive Tobacco Control Programs* (2005) |
| | *Introduction to Program Evaluation for Comprehensive Tobacco Control Programs* (2001) |
| | Funding Announcement Program Announcement 03022 – (2003) |
| | OSH Consensus Core Indicators (2008) |
| | *Best Practices for Comprehensive Tobacco Control Programs* (2007) |
| | *Introduction to Process Evaluation in Tobacco Use Prevention and Control* (2008) |

| | |
|---|---|
| | Measurement development options for non-standard OSH core indicators (internal correspondence report, 2007) |
| | Use of Outcome Indicators for Planning and Evaluating NTCP presentation given by Paul Hunting, MPH, at the National conference on Tobacco or Health, Minneapolis, MN (2007) |
| | ExpectMore.gov Detailed Information on the Chronic Disease Prevention Assessment (accessed April 2008) |
| | CDC OSH Website, http://www.cdc.gov/tobacco  (accessed December 2008) |
| | CDC NTCP Website, http://www.cdc.gov/tobacco/tobacco_control_programs/stateandcommunity/index.htm  (accessed December 2008) |
| | Core Indicator Workgroup meeting notes (July 2007) |
| | Collaborative Chronic Disease, Health Promotion, and Surveillance Program Announcement: Healthy Communities, Tobacco Control, Diabetes Prevention and Control, and Behavioral Risk Factor Surveillance System CDC RFA- DP 09-901 FOA (November 2008) |

**APPENDIX G**

**OBSERVATION GUIDE**

## Observation Guide
## Performance Measurement Study

| | |
|---|---|
| **Date of Observation** | |
| **Role of Observer** | |
| **Description of the Observation Event** | |

**Description of Setting/Context:**

**Description of Participants:**

**General Observations:**

**Observed Nonverbal Communications:**

**Observed Informal Interactions:**

**Detailed Field Notes (to be completed following observation):**

**APPENDIX H**

**ANALYTIC CODEBOOK**

**Dissertation**
**Codebook**

---

Code Number:        001

Code Name:          Context: Fed level

Brief Description:   Contextual factors at the federal level

Long Description:    Contextual factors at the federal level which are relevant to CDC
                     programs and the development or use of performance measurement

When to Use:         Apply code to text referencing federal level contextual factors such
                     as diminishing federal resources, federal level policies and
                     legislation, health care reform, HHS-related (HP 2010), federal
                     emphasis on accountability or results.

When NOT to Use:     Do not use this code for GPRA/PART issues.

Coding Rules:        If GPRA/PART related, code one of the two "GPRA/PART"
                     codes. If related to political factors directly affecting the
                     development of a particular PM system, double code as "Process:
                     Political Factors."

Example:             "…but they also understand that in light of these funding
                     restrictions and/or reductions, they need to start demonstrating
                     accountability and that this is a very difficult area to do, to do so."

---

Code Number:        002

Code Name:          Context: CDC

Brief Description:   Contextual factors at the CDC level

Long Description:    Contextual factors at the overall CDC level which are relevant to
CDC programs and the development or use of performance measurement

When to Use:  Apply code to text referencing overall CDC level factors such as agency
mission, goals; issues of alignment with agency strategic goal planning; CDC Office of
the Director views on value of performance measurement; issues of overall agency
resources.

When NOT to Use:    Not for CDC program-specific level (e.g., COTPER, NBCCEDP) unless the program specific issues relate to larger CDC context (e.g., mission, goals). Do NOT use for political factors at CDC level (e.g., CDC OD/OD – Gerberding)

Coding Rules: Double code text that relates to CDC context and program-specific PM system (e.g., alignment). If policy/political factors at CDC level, code as "Process: Policy/Political Factors"

Example:                 "they're [the PM] tied to the nine [CDC] preparedness        goals."

"If the political system lets us do it.  And if CDC lets us do it."

Code Number:          003

Code Name:            Culture: Shift to PM

Brief Description:     PM as a shift in organizational mgmt/culture

Long Description:      Descriptions of the adoption of performance measurement as requiring a shift in organizational management or culture. The shift may be evident in change in how work is done, shift in how PM is perceived, shift to new view of PM as institutionalized, PM as new, established culture norm, shift to new view of valuing PM, diminishing resistance to PM.

When to Use:  Apply code to text that describes the adoption of PM as a mgmt or cultural shift, the need for a culture shift in order to adopt PM, or the lack of acceptance at a broad organizational level for recognizing value or importance of PM.

When NOT to Use:    Do not use when culture shift refers to move toward networked structures and approaches.

Coding Rules: If relates to culture/organizational shift to networks, code as "Culture: Shift to Networks"

Example:        "I personally think it's a really critical piece.  I don't think we do enough of it in public health.  I don't think we have the skill set, I mean, well we have lots of data analysis skill sets, I don't think we have a staff level understanding of it as much as we should.  I think the danger is we tend to develop performance measures without developing a performance management system."

Code Number:          004

Code Name:            Culture: Shift to Networks

Brief Description:     Cultural shift to use of network approaches

Long Description:      View of move to network approach as a cultural shift in perspective. In public health, programs have more traditionally viewed their responsibility as related to the program activities funded by CDC and the carrying out of public health work through traditional, vertical, intra-governmental relationships.

When to Use:  Apply code when text describes recognition that working through networks is important, that such a shift may require a change in role for the public health

worker, the consequences of moving to a network, the need for moving to network approaches.

When NOT to Use:    Do not use this code for reference to characteristics of the networks themselves (vertical dimensions, variability inherent in networks)

Coding Rules: Code any references to network characteristics as any of the Network Characteristics code family. If relates to shift in culture/organizational mgmt to adopt PM, code as "Culture: Shift to PM"

Example:       "You realize that public health does not work in a vacuum"

"It was obvious right from the beginning you know that you needed a broader perspective."

Code Number:          005
Code Name:            Design: Control over Performance on Measures

Brief Description:     Level of control over the performance on PM

Long Description:      Factors may affect the grantees' level of control over their performance on a particular measure

When to Use:  Apply code when text reflects issues that compromise control over the performance on performance measures – system-level challenges (e.g., provider capacity to perform colonoscopies), dependencies on network partners (e.g., labs, jails), level of outcome (e.g., mortality), other factors affecting outcome (e.g., homelessness, drug use, poverty). Also apply code when text addresses the intentional development of PM for which grantees will have extensive control

When NOT to Use:

Coding Rules: If appropriate, double code to the relevant code in "Network Characteristics" family of codes. May be instances when it's appropriate to double code to "Design: Control within Network" (e.g., text about lack of control over jails for whose performance you're dependent).

Example:       "So, at this point, I feel like our, the grantees have a fair amount of control over this set of six measures."

"It's, it's difficult to understand who owns the performance at the state level or even at the local level.  And what, what we can say about that networking piece is that states want a lot of credit for that"

Code Number:          006
Code Name:            Design: Control within Network

Brief Description:     Level of control over network partners

Long Description:      Given network structure, CDC and grantees often do not have direct relationships or command and control authority over local level implementers.

When to Use:  Apply code when text references limited control over network partners [e.g., local implementers, lack of direct relationship with local levels, network partners over which grantee or CDC has little to no control (e.g., jails), structural challenges that impede control over partners (e.g., governance structures)].

When NOT to Use:    Do not use when text relates to bargaining, consensus building, or building networks. Do not use when text relates to control over the performance on a particular measure.

Coding Rules: Code references to bargaining, etc. within a network as "Network Characteristics: Bargaining, Consensus Bldg, Bldg Networks". Code references to control over performance on measures as "Design: Control over Performance on Measures"

Example:        "I mean we don't have a direct relationship, with the exception of 4 or 5 cities, you know, any direct relationship with the counties.  We can't ask them to do anything so, or get really any information from them it all has to go through the state."

"I think people still think, you know, they're concerned about the things that they control in their program and if you go on a statewide or a jurisdiction wide basis, they feel like they're being held accountable for things that they don't really have power over."

---

Code Number:        007

Code Name:        Design: Conveying Priorities

Brief Description:        Performance measures reflect program priorities

Long Description:        Performance measures typically reflect key priorities of a given program and are intended to influence program behavior by providing a focus in key performance areas. In decentralized contexts, it may be difficult to identify measures that are priority and meaningful for both the federal and state levels. Some programs may not have a strong sense of what their priorities should be.

When to Use:  Apply code when text refers to whether or not the program's performance measures reflect program priorities (or goals) for the national program, act as incentives for programs to focus on certain areas ("what gets measured gets done"), or if programs face challenges identifying key priorities.

When NOT to Use:    Do NOT apply "PM: Value"

Coding Rules:        If text relates to other purposes/uses/value of PM, code as "PM: Value"

Example:        "I'm not so confident in our measures to say that if somebody modeled a program after them in a sense that they were really trying to focus in on something that we focused in on, that it would do good and not harm, you know"

---

Code Number:          008

Code Name:            Design: Fairness

Brief Description:     PM perceived as "fair" (or not); or applied "fairly" across network; or used in a "fair" way

Long Description:      PM may or may not be perceived as "fair" – whether they are seen as fair may influence their acceptability for purposes such as accountability and budgeting. In networks, the variability inherent among actors challenges the notion that a "fair" set of indicator can be developed, or that "fair" targets can be set for all. Likewise, the use PM data may be perceived as unfair (e.g., ranking).

When to Use:  Apply code when issues of fairness are implied or explicit in the text.

When NOT to Use:

Coding Rules: If text also relates to fairness of targets, double code to "Measurement: Targets" or "Measurement: Target Challenges."

Example:       "To me as long as the measures are clear, the performance expectation is clear, the technical assistance to help programs get to the level of performance we expect is clear, that we apply it fairly across all funded programs is transparent and clear, it should be fine."

Code Number:          009

Code Name:            Design: Flexibility

Brief Description:     Flexibility needed in PM system

Long Description:      In network context, flexibility may be needed in the design or implementation of the PM system (e.g., choice of measures, choice of targets, data system, data sources).

When to Use:  Apply code when text addresses issues of flexibility (e.g., incorporating flexibility, not allowing flexibility) in the design and/or implementation of the PM system.

When NOT to Use:    Do NOT use code when text addresses issues focused on variability of activities, data collection, etc.

Coding Rules: Code issues of variability under appropriate "Network Characteristics" family of codes. If text relates to flexibility and program priorities, double code to "Design: Conveying Priorities"

Example:       "So, we've tried to steer clear of optional measures or allowing them to, you know, on the one, on the Chlamydia screening in juvenile detention, we do allow some flexibility in which sites they pick but we're, you know, for the big cities with more than 500 [people in the facility], you've got to report on all of them"

Code Number:          010

Code Name:            Design: Frameworks

Brief Description:     Frameworks to structure development of PM system

Long Description:      Frameworks such as logic models, process mapping, evaluation frameworks, and other unique constructs may be used to guide the PM development process.

When to Use:  Apply code to text that refers to such a framework or rationale that guides development of the PM system.

When NOT to Use:

Coding Rules:

Example: "The second set of information is on capabilities which are what can they do with those capacities that they just bought. How can they demonstrate that they can do something with that, that infrastructure?"

Code Number:          011

Code Name:            Process: Time

Brief Description:     Time is needed to develop PM

Long Description:      Recognition that developing PM is a (learning) process that takes time, usually a significant amount of time. Process may be incremental in nature, an iterative process of trial and error, piloting, etc. Recognition that time is needed to build a worthwhile PM system.

When to Use:  Apply code when text addresses the inevitable passing of time involved in the process of developing PM. Text may mention frustration with amount of time needed, importance of the process itself, the learning process inherent as time passes, and idea that PM must be gradually (i.e., incrementally) implemented.

When NOT to Use:

Coding Rules: Double code text that relates to the idea that it takes time to gain stakeholder buy-in to the appropriate "Stakeholders" family of codes.

Example:        "We want to improve program, you know, and it's, it's been a long haul, quite frankly."

"That it's going to take us time to get to an established set of measures that everybody is going to feel comfortable in gathering the data in the same way and find that those measures are useful for program planning and program improvement."

Code Number:          012

Code Name:            Process: Measures Evolve

Brief Description:     PM / PM system evolves as system is developed

Long Description:      Programs often started with large set of "pilot" measures or worked to fine tune or improve measures as the PM system was developed and improved upon. In other cases, programs are trying to develop new, additional measures to capture more complicated constructs. This approach may relate to the incremental nature of developing a PM system in a decentralized system – need to start more simply before moving toward the more complex.

When to Use:  Apply code when text addresses change, adaptation, or evolution in the measures or when text describes efforts to move toward adding targets, adding more local measures, or removing measures that were unsuccessful or didn't work.

When NOT to Use:

Coding Rules:            Double code text to "Stakeholders Buy-in: Strategies" when appropriate.

Example:        "The, the first set was very, for the most part, very specific. We had a set of syphilis measures that looked at traditional STD program activities such as interviewing index patients and interviewing cluster patients and suspects and associates, looking how quickly programs were able to do that."

Code Number:          013

Code Name:            Process: Stakeholder Involvement

Brief Description:     Stakeholder involvement in the development of the PM

Long Description:      Given the network, many different stakeholders may be involved in the development of the PM system. There may be national groups, state grantees, or other partners who play some role in the development of the measures (e.g., reviewing draft measures, participating on PM work groups, piloting measures). This may reflect a collaborative approach to development of the measures and some give/take between CDC in the selection of measures.

When to Use:  Apply code when text reflects stakeholder involvement in the development of the PM system (e.g., suggest measures, pilot measures, provide feedback) or the range of stakeholders that should be engaged in the process.

When NOT to Use:    Do NOT use when text relates to stakeholders or stakeholder involvement outside of their participation in the development of the PM system.

Coding Rules: Code other stakeholder-related text using the "Stakeholder" family of codes.

Example:       "Well, a lot of them [the measures] came from them [grantees]."

"I think we, we'd get in much more trouble with programs if we tried to raise them [targets] for the same reason."

Code Number:          014

Code Name:            Process: Resources

Brief Description:     Resources needed to support PM

Long Description:      Resources are needed to support the development, implementation, and maintenance of PM. Resources may include consultants hired to help develop the system, staff expertise, funding for data system development and maintenance, funding for contracts that support data management, funding support for on-going maintenance, technical assistance, training, etc.

When to Use:  Apply code for text referring to or describing resources used or needed to support PM (i.e., development, implementation, or maintenance).

When NOT to Use:

Coding Rules: If appropriate, double code to "Use – Technical Assistance/Tools" if resources relate to the training and TA.

Example:        "Well, part of it has been not having a budget internally [to pay for TA to the grantees]. I mean, for us to, to function with it, to never know how much money and, you know, it's been good that we've been able to have a consultation on an annual basis."

Code Number:          015

Code Name:            Process: Policy/Political Factors

Brief Description:     Political factors affecting the development/implementation of PM

Long Description:      Policy and political factors seem to affect the development of the PM system. These factors could include requirements passed down from other Federal agencies (e.g., Dept of Defense' Target Capabilities List), policies (e.g., PAHPA, HSPD #21), investigations (e.g., GAO, IG), Congressional inquiries or hearings, CDC OD OD (Gerberding), or political scrutiny and pressures.

When to Use:  Apply code to text reflecting policy or political factors that affect the process of developing or implementing PM.

When NOT to Use:    Do NOT use for broader political pressures and policies at the Federal level that do not directly affect the development of the PM. Do NOT use for GPRA/PART related text.

Coding Rules: Code broader federal level political pressures and policies as "Context: Federal". Code GPRA/PART issues as one of the two "GPRA/PART" codes.

Example:        "The challenge is around the political agendas and the fact that Washington, ASPR, HHS, are completely unreasonable in what they expect from a measurement perspective. If you look at HSPD 21 that was just released in December I think…"

Code Number:          016
Code Name:            GPRA/PART: Perceived Value, Accountability, Attribution
Brief Description:     Perceived value of GPRA/PART – related issues of accountability, attribution

Long Description:      Both GPRA and PART are federal policies that require the use of performance measurement for federal programs. The implementation of PART for CDC programs has changed over time – not all of the four cases were directly reviewed for PART by OMB. Perceptions of the GPRA/PART policies vary from useful to "absurd." Both policies aim to increase government accountability for performance. Issue of attribution are also relevant here if GPRA/PART are used for accountability and for Federal (Congressional) level funding appropriations.

When to Use:  Apply code when text addresses the perceived value of GPRA/PART (either good or bad); views of the PART/GPRA measures themselves; text about the experience of being reviewed, text addressing GPRA/PART and issues of accountability,

attribution, and federal level funding appropriations from Congress; or text re: alignment of program level PM with GPRA/PART measures

When NOT to Use:
Coding Rules:
Example:      "But it (GPRA) is as a concept, as something that ain't going to be going away, as something that's going to one, stay with us, and is part and parcel of this greater era of accountability, that is quite useful for us starting the programs, and this is part of it you know. PART is a huge pain in the rear end."

"The first thing I do is look at it and I say, you know, 'this is ridiculous' [GPRA].   You know, these are really bad ways to evaluate this program."

"A whole other sort of thing about you know, GPRA did a good thing GPRA asked 'what's your business'? But whether your business was a lot of what you were measuring, it wasn't it went to sort of, not clear what your business was about."

Code Number:        017

Code Name:          Management: Policy Tools in Networks

Brief Description:   Mechanisms used to assert control in networks

Long Description:    Policy tools used to assert control in network contexts such as cooperative agreements, CDC field staff, formal agreements/MOUs, funding restrictions, PM itself, performance-based contracts, etc. Can include limits of those tools, as well

When to Use:  Apply code for text referring to any policy tools or other means used to assert control within network structures.

When NOT to Use:    Do NOT use when text refers to whether or not a particular agent has control or not over network partners.

Coding Rules: Code text related to whether a particular agent has control or not over network partners (e.g., local implementers) as "Design: Control within Network"

Example:        "You know, telling, telling providers, yes, we'll pay for something when you should be telling them, no, we don't pay for that." [provider reimbursement as a form of control over network partners]

 "And, you know, they're putting out, like, mutual aid agreement templates and stuff for grantees to be able to use because it's so clear that there is so much reliance on different entities."

Code Number:        018

Code Name:          Management: Unintended consequences

Brief Description:   Unintended consequences of PM

Long Description:    Both positive and negative consequences / effects that may occur based on the development and implementation of PM.

When to Use:  Apply code when text reflects any type of unexpected or unintended consequence of the PM system (e.g., improved relationships, more collaboration, unintentionally encouraging unfavorable practice, unintentionally creating frustration for grantees because PM don't reflect everything they do)

When NOT to Use:	Do not use when text describes efforts of "gaming"

Coding Rules:	Code text describing "gaming" of the PM as "Measurement: Gaming"

Example:	"It, it has, our challenge with that has more been implementation in terms of we already kind of put out exercise schedules when that first came in so we didn't, like, redo all of our schedules and things and there was also some preliminary that needed to be done to be able to put the right groups of people in there to do that."

Code Number:          019

Code Name:            Measurement: Common Measures

Brief Description:     Identification of a common set of federal-level PM

Long Description:      Federal-level PM typically involves identification of a common set of PM that can be used to describe the program nationally/in aggregate. Network structures make this difficult given variability in several key areas, including program activities.

When to Use:  Apply code when text describes issues or challenges related to the identification of a common set of federal-level PM.

When NOT to Use:

Coding Rules: When appropriate, double code text to "Network Characteristics" family of codes (in particular, ones related to variability of activities, data collection, etc.).

Example:      "Listeria.  And, but there are a whole host of other.  We just chose those two [infections] because they had enough numbers that everybody [all grantees] would be reporting on."

 "I think it's very appropriate for all programs, if you screen a woman and they have an abnormal Pap, that they get into diagnostic [care], and if they have a diagnosis of cancer they get into treatment.  So these to me are very, I'm not sure the right word is generic, but these are measures that are relevant to everybody [all grantees]."

Code Number:          020

Code Name:            Measurement: Challenges

Brief Description:     Challenges to measurement

Long Description:      A variety of challenges to measurement confront the development of a PM system for "wicked problems" in a decentralized context. These may include a lack of science or scientific standards, undefined programs where outcomes aren't clear – don't know what to measure, sample sizes that are small, analytic challenges, could be measuring the wrong thing, or finding that a measure doesn't work well.

When to Use:  Apply code to text that reflects challenges to measurement

When NOT to Use:     Do NOT use when measurement challenge relates to what programs find hard to measure (e.g., level of catastrophe, prevention, PID)

Coding Rules: If text relates to things that are hard to measure, code as "Measurement: Hard to Measure"

Example:       "So, can we have, like, the same measure [local-level measure] and then just roll it up and have it be an aggregate that represents what's happening at the state level or are they not really roll-upable?  You know, like, how, how is it that you actually measure what's happening at the local level and somehow aggregate it to get a sense of how prepared a state is."

Code Number:          021

Code Name:              Measurement: Hard/Impossible to Measure

Brief Description:      PM that are difficult or impossible to measure because they are complex or data is unavailable

Long Description:      Indicators or constructs that are analytically challenging or impossible to measure such as indicators that require a proxy, vague constructs like prevention or collaboration, staff capacity, or indicators for which data are unavailable.

When to Use:  Apply code when text implies indicator is difficult or impossible to measure

When NOT to Use:    Do NOT use for more general measurement challenges

Coding Rules: For measurement challenges that do not meet this definition, code as "Measurement: Challenges"

Example:        "You know, we have this continued idea or we have a need to figure out some way to measure syphilis prevention during pregnancy.  We just haven't figured out how to do that yet because pregnancy is not reportable.  We've tried various ways to come at it and nobody's been happy with the outcome."

Code Number:          022

Code Name:              Measurement: Criteria

Brief Description:      Criteria used for measurement selection

Long Description:      Criteria are frequently used to select performance measures from a larger set of candidate measures or to simply rate a set of proposed measures. Criteria may include face validity, extent of a science-base, consistent with goals/objectives, and relevancy (whether the measure is "meaningful" from the perspective of various stakeholders), among others.

When to Use:  Apply code when text describes or refers to criteria used to judge potential performance measures.

When NOT to Use:

Coding Rules:

Example:    "It's looking to those people who actually do it to say, does this make sense? We want these measures to be relevant, we want them to be feasible, we want them to be reliable and we want them to be valid."

Code Number:          023

Code Name:            Measurement: Interpretation/Definitions

Brief Description:     Interpretation of the PM and how it is calculated

Long Description:      Given the network structures, network actors (e.g., grantees, local level implementers) may interpret or define the performance measures differently – this refers to both an understanding of the intent of the measure and the calculation of the measure.

When to Use:  Apply code when text refers to grantees interpreting the measure differently, interpreting the intent of the measure differently, defining data sources differently, other definitional issues, lack of consistency in how things are measured

When NOT to Use:    When text refers to variability factors that are un-related to the consistent interpretation of the measure itself

Coding Rules: If relevant to the interpretation of the measure, may also double code to "Network Characteristics: Variability – Data Sources, Collection, Reporting, Systems"

Example:        "And that was one issue.  The other issue is, even though we sent out the guidance, they didn't read it so they defined things their own way."

Code Number:          024

Code Name:            Measurement: Gaming

Brief Description:     Grantees' / or CDC's manipulation of the PM system or calculation of a particular measure

Long Description:      "Gaming" can occur when there are incentives that drive programs to manipulate data or circumstances to improve the level of performance on a particular measure. Gaming is used to make the data seem better than they actually may be. For CDC, gaming may involve developing measures to "game" political requirements like PAHPA.

When to Use:  Apply code when text describes examples of gaming.

When NOT to Use:

Coding Rules:

Example:        "I think that, especially those 75% of women screened [for breast cancer], I've seen programs structure their funding to meet that, they have algorithms that they've developed to make sure that, we've seen programs be rated 75.1%.  It's hard to tell whether programs might hold data that is not complete, like get it complete so that it doesn't effect it [the PM], I would not be surprised if that's happening"

Code Number:        025

Code Name:          Measurement: Process

Brief Description:   Process-level PM (typically program activities/outputs)

Long Description:    Process measures typically reflect the activities and outputs that lead to outcomes

When to Use:  Apply code when text reflects information about process measures as part of a PM system

When NOT to Use:    Do NOT use for reference to proximal or longer-term outcome measures, unless text relates the two (relates process and outcome measures in some way)

Coding Rules: Reference to outcome measures should be coded as "Measurement: Outcomes." Double-code text that refers to both process and outcome measures, to "Measurement: Outcomes."

Example:        "…and in some ways, the more typical of CDC, than some other government kind of programs that for a lot of government programs you have a shit load of process measures and not a whole lot of outcome measures, you know, and in fact it's not even clear what your outcome is or there is less agreement about outcome whereas, it seems we got turned topsy-turvy, you know we can agree we want to reduce rates of disease and have measures of rates of disease but we just don't know what to do here [black box in between activities and outcomes]."

Code Number:        026

Code Name:          Measurement: Outcomes

Brief Description:   Outcome-level PM (typically reflect changes resulting from program activities)

Long Description:    Apply code when text reflects information about outcome measures, regardless as to whether they are immediate, intermediate (proximal), or long-term. In public health outcome measures may be tied to changes in behavior, knowledge, attitudes, policies, morbidity, mortality, etc.

When to Use:  Apply code when text reflects information about any type of outcome measure, whether immediate, intermediate (proximal), or long-term

When NOT to Use:    Do NOT use for reference to process measures, unless text relates the two (see example below)

Coding Rules: References to process measures should be coded "Measurement: Process." Double-code text that refers to both process and outcome measures.

Example:      "Yet the reality is we're still very much in process monitoring rather than outcome monitoring.  We're working towards developing those measures but it's taken some time because it does take time if you want to do it right."

Code Number:          027

Code Name:            Measurement: Targets

Brief Description:     Established targets or benchmarks for performance measures

Long Description:      Targets may or may not be used as part of a PM system.
Approaches to setting targets vary (e.g., based on science, policy, past performance,
individual grantee).

When to Use:  Apply this code when text describes views on targets and approaches to
establishing targets

When NOT to Use:    Do NOT use code when text describes challenges related to setting
or using targets with PM.

Coding Rules: References to target-related challenges should be coded "Measurement:
Target Challenges"

Example:       "Right, so that was part of the process.  And, you know, and that's
sometimes the argument I get from some programs is they say, we can't possibly have all
the women meet that [the target] and I say, well, you only need to have 75% to meet it.
You know?"

---

Code Number:          028

Code Name:            Measurement: Target Challenges

Brief Description:     Challenges to using/setting targets or benchmarks for PM

Long Description:      Given "wicked problems" and a networked context, challenges
exist in setting or using targets or benchmarks for PM. For instance, variability in
programs may inhibit the ability to set a standard; a lack of science or trend data may
challenge the ability to set a defensible target; variability in context and programs may
compromise the utility of comparing sites based on targets.

When to Use:  Apply code when text refers to challenges in using/setting targets and
when text addresses the utility of comparing sites based on their performance.

When NOT to Use:    Do NOT use this code when text refers to approaches to setting
targets and views on the use of targets for decentralized public health programs.

Coding Rules: References to approaches to setting targets or views on the use of targets should be coded as "Measurement: Targets"

Example:     "I mean, I can't compare North Carolina and South Dakota.  I mean, it's two entirely different environments, as far as what they're operating in, what they're trying to do."

 "There's no factor in our measure that take into consideration scalability and that goes back to, you know, 48 hours is, is the same for Boise as it is for New York City."

Code Number:          029

Code Name:            Measurement: Data Quality / Validity Challenges

Brief Description:     Perceptions of data quality and concerns about data validity

Long Description:      Ensuring the validity of data is important for ultimate interpretation and confidence in the performance data. In networked contexts, data validity is challenged by a number of issues including the multitude of different people/agencies involved in data collection and reporting, the variability in data sources and data systems, and differentials in staff capacity and resources.

When to Use:  Apply this code when text refers to perceptions of data quality and issues of data validity / reliability

When NOT to Use:     Do NOT code text that refers to methods of ensuring data quality – use "QA: Methods".

Coding Rules: If text refers to methods to ensure data quality, code as "QA – Methods." Double code text to others that may be relevant such as those in the Network Characteristics family (e.g., variability: data collection) or the Measurement Challenges/Measurement: Hard to Measure.

Example:       "I think part of it is because they have, that, that it's too difficult to get the right data and they will sort of estimate. They will give us estimates rather than accurate. You know, than actual data, they'll estimate because it's too difficult to get the correct data."  [Double code to Measurement: Hard to Measure"]

Code Number:          030

Code Name:            Network Characteristics: Horizontal

Brief Description:     Reference to horizontal or lateral network partners/relationships

Long Description:      Networks are typically comprised of both vertical and horizontal dimensions. Horizontal relationships may exist at multiple levels (e.g., federal level agencies – CDC, HHS, NCI; within CDC – DCPC, OSH; and at state, regional, and local levels).

When to Use:  Apply this code when text references or describes horizontal relationships (or the need for these relationships) as they relate to performance measurement. This can

be when text names those partners (e.g., fire dept.) or use code for text that references indicators for which performance relies, to some extent, on horizontal partners.

When NOT to Use:     Do NOT use when text refers to vertical dimensions.

Coding Rules: If text refers to both horizontal and vertical dimensions, double code for both; If text refers to vertical dimensions only, code as "Network Characteristics: Vertical"

Example:       "Yeah, and I think a lot of project areas have done this community thing, and they have established a relationship with the CBOs and the agencies within their community already.  That's a given, that's established, that's not going anywhere because they understand that relationship and that partnership is beneficial to both.  But, I think we haven't done as much with the private docs, you know, in even the laboratories as well."

Code Number:         031

Code Name:           Network Characteristics: Vertical

Brief Description:    Reference to vertical network partners/relationships

Long Description:     Networks are typically comprised of both vertical and horizontal dimensions. Vertical relationships are often inter- or intra-governmental (e.g., CDC – State – local), but could reflect vertical relationships between other agencies as well (e.g. National organization, State health department, local advocacy group). The vertical relationships may also reflect more traditional hierarchical arrangements in government.

When to Use:  Apply this code when text references or describes vertical relationships (or the need for these relationships) or structures as they relate to performance measurement.

When NOT to Use:    Do NOT use when text refers to horizontal dimensions.

Coding Rules: If text refers to both horizontal and vertical dimensions, double code for both; If text refers to horizontal dimensions only, code as "Network Characteristics: Horizontal"

Example:        "Not because anybody, again, is doing anything on purpose, but because when you have layers and then you're moving vertical and then horizontal and then doing that, coming back…"  [double code with "Network Characteristics: Horizontal"]

"I'm not sure how deeply involved with STD history you've been but the STD program had been run top down for a lot of years, very strongly managed, well forcefully managed."

Code Number: 032

Code Name: Network Characteristics: Value and Goal Conflicts

Brief Description: Conflicts/issues between network partners around values, goals, mission, priorities, etc.

Long Description: Given the network structure, conflicts or issues may arise in developing performance measurement related to differing agency values, goals, mission, priorities, areas of responsibility / turf, culture, etc. These conflicts may impede the development of a common set of measures or the collaboration needed to collect/report performance related data.

When to Use: Apply this code when text refers to or describes tensions between network partners that is related to fundamental differences in values, goals, etc.

When NOT to Use:

Coding Rules:

Example: "Human nature, again, I could be part of this myself if I was out there, I have been out there in the past, but you've got your own little kingdoms and queendoms and fiefdoms and all that and the HIV STD programs are a great example of that. They're still not really working together out there and it's now 2008."

"You know, I think each State determines who their screening population is going to be, and in Washington it's not just Department of Health at the site, we have stakeholders, and we have a really strong Komen presence in our State" [referring to partners who want women aged 40-50 screened vs. CDC policy of screening women 50-64]

Code Number:          033

Code Name:    Network Characteristics: Bargaining, Consensus Building, Building
Networks

Brief Description:       Efforts of bargaining and consensus building in networks

Long Description:       Bargaining and consensus building are recognized as important
strategies in a network context – taking the place of traditional command and control
structures/approaches in hierarchical arrangements. Building relationships with network
partners is essential – through consensus, compromise, collaboration, respect, etc.

When to Use:  Apply code when text reflects efforts of bargaining or consensus building,
efforts of reciprocity or give/take, negotiation, collaboration, etc. – Or other strategies to
build network relationships (at all levels, including CDC).

When NOT to Use:

Coding Rules:

Example:       "I think so, but I think it's because we have developed really good
relationships with our prime contractors and, you know, they're there for us when we,
you know, need something or we need something done differently, or we've identified an
issue, and we're there for them, you know, so we're reciprocating."

Code Number:          034

Code Name:              Network Characteristics: Variability Activities

Brief Description:       Variability across grantees in the program activities implemented
at each site.

Long Description:       Public health endorses the use of approaches and strategies suited
to the unique context and needs of particular communities. For a national program
operating through a network structure, this results in extreme variability in program
activities which make identification of common codes difficult.

When to Use:  Apply this code when text refers to the variability across grantees in their
activities/ in how they do things and/or when text refers to the effects of this variation on
the development of PM

When NOT to Use:

Coding Rules: Double code text to "Measurement: Common Measures" if appropriate

Example: "Because they're so different, every place is so God damn different and it does not help you very much, sometimes, but this is a big problem I've had conceptually across the states and even in terms of the groups they think they get a lot more having people come in telling them what to do, looking over your state, looking over the fence. They would be far more advised to look closely at their own system and understand the difference."

Code Number:          035

Code Name:              Network Characteristics: Variability – Capacity/Resources

Brief Description:      Differences across grantees or other network partners in staff capacity, agency resources, agency infrastructure

Long Description:      Given the network structure, there is variability in the capacity, infrastructure, and resources of individual grantees and other network partners. This variability has implications for performance measurement in terms of collecting and reporting data, availability of adequate data systems, staff to manage and interpret data, etc.

When to Use:  Apply code when text refers to differences in capacity or resources across the grantees or network partners – may be both a lack of or a richness of resources, capacity, infrastructure

When NOT to Use:    Do NOT code text that refers to issues of resources needed to support the development/implementation of the overall PM system (national level).

Coding Rules: If text relates to broader resource issues for the overall PM system, code as, "Process – Resources"

Example:        "The diversity in the programs is, is a big issue because large programs have larger staff and, and more opportunity to develop the infrastructure necessary to do some things sometimes.  They also have much more disease and a far more complex program.  Smaller programs don't have any resources or staff."

Code Number:          036

Code Name:   Network Characteristics: Variability – Data Sources, Collection, Reporting, Systems

Brief Description:      Variability in grantees' data systems and data sources, collection, and reporting processes

Long Description:      Given the network structure, there is variability in grantees' data systems and in the data source for the measures, as well as the way the data is collected and reported. This poses several challenges, including ones of data validity, for the PM system.

When to Use:  Apply code to text that refers to variability across grantees in data sources, data systems, data collection, or data reporting.

When NOT to Use:

Coding Rules: Double code text to others that may be relevant such as "Measurement: Data Validity"

Example:       "They collect data differently and they, you know, store it differently, and they can report on it differently.  So, that's something that's always been, you know, hit and miss with us, as far as data we're going to get, and can this, is it going to be, you know, valid and useable."

Code Number:        037

Code Name:          Network Characteristics: Variability – Context (comparing too)

Brief Description:      Contextual variance across network members (e.g., grantees)

Long Description:      In a networked structure, programs may vary extensively in regard to their overall context (e.g., size, capacity, disease burden, cultural norms, demographic factors). Given this, the interpretation of PM data may need to account for those contextual factors – similarly, these contextual differences may compromise the value of comparing the PM data for state A to state B.

When to Use:  Apply code when text reflects issues of variance in grantee context; or when text references the role of context in the interpretation of data.

When NOT to Use:    Do NOT use when text refers to variability in activities (what grantees do)

Coding Rules: If text relates to variability in grantee activities, code "Network Characteristics: Variability – Activities"

Example:        "I think it's [grantee context] very important, I think Bill eluded to it before, because some of the indicators you really need to know how the program is structured and whatever other things are going on, just like with the funding source, so a program that has a large non-federal grant, or money coming in, has the flexibility to pinpoint which mammograms they're paying for with federal funds…"

Code Number:        038

Code Name:          Network Characteristics: Value

Brief Description:      Perceived value of networks and networked approach to public health

Long Description:      The literature suggests that networks provide a number of advantages in addressing complex problems – together, network partners can tackle the varied factors contributing to social problems, networks leverage scarce resources, and networks extend programs' reach (among other things).

When to Use:  Apply code when text reflects perceptions of the value of networks and networked approaches to public health – also the value of particular network partners

When NOT to Use:

Coding Rules:

Example:	"I mean, my understanding is that is, it's [collaboration] extremely important because a lot of the, I think a lot of the, the work that STD needs to have done or things that STD needs to accomplish is, are in areas that are essentially outside their direct control.  Like, for instance, with the jails and the family planning clinics and things like that.  And so to be able to actually work with those people in other, other programs to, to get them to buy into your goals and, and to participate in your process is, seems to me to be very critical."

Code Number:          039

Code Name:            Performance Measurement: Value

Brief Description:     Perceived value of performance measurement

Long Description:      Performance measurement may be perceived as an invaluable management tool, a meaningless burden, a defense against funding cuts, etc. Value may be expressed generally, but also reflected in intended uses or purposes.

When to Use:  Apply this code when text reflects the value (or lack of value) of performance measurement for their program, include text about potential use/users and purposes.

When NOT to Use:

Coding Rules:

Example:      "[The biggest benefit of PM has been in] defining the program."

"So, they get that, but they also understand that in light of these funding restrictions and/or reductions, they need to start demonstrating accountability and that this is a very difficult area to do, to do so.  So they really want, there's a, there's a fairly large constituency of folks who really want to get this right."

Code Number:          040

Code Name:            Program Characteristics: Developmental Stage of Program

Brief Description:     Program's maturity or developmental stage

Long Description:      Programs vary in their developmental phase or level of program maturity. The STD program is one of the oldest at CDC; PHEP is one of the newest. A program's developmental stage may be reflected in its articulated goals, program experience, level of institutionalization, etc..

When to Use:  Apply code when text refers to the maturity of a program, their stage of development, experience – and when text explicitly recognizes a relationship between developmental stage and the PM system.

When NOT to Use:    Do NOT use code when text refers to the developmental phase of the performance measurement system – unless a direct correlation is made.

Coding Rules: If text refers to the performance measures or measurement system, code as "Process: Measures Evolve." If text makes a direct correlation between developmental phase of the program and developmental phase of the PM, double code with "Process: Measures Evolve."

Example:    "You know, so our STD programs are well entrenched and some of them have trouble moving with the environment.  So, yeah, I think the first set of measures was our attempt to get them used to the idea that, you know, you really do need to collect data that gives you information about what you're doing and how you're performing.  There were obvious deficiencies with the current set as far as aggregating the data for a national measure."

Code Number:            041

Code Name:              Program Characteristics: Disease Burden

Brief Description:      Variability or changes in disease burden or emergency incidents across grantees / network

Long Description:       Given the variability in disease burden (e.g., syphilis), emergency incidents (e.g., anthrax attack in NY vs Nebraska; scale of incidents), and populations to be served (breast cancer screening in Alaska vs. California), performance measures may be more or less meaningful to a particular grantee/jurisdiction. Similarly, disease burden may shift in regard to certain high risk populations, with the potential to affect the utility of a particular PM.

When to Use:  Apply code when text references disease burden, variability in types of emergency hazards and scale of incidents, etc.

When NOT to Use:    Do NOT use code for descriptive text about program context in general

Coding Rules: May double-code text that is related to both disease burden and measurement (e.g., Measurement: Hard to measure); May double code text that is also related to identifying common measures (e.g., Measurement – Common Measure); If relates to program context more generally, code "Program Characteristics: Goals & Scope of Program"

Example:        "We're, you know, we're finding gonorrhea and Chlamydia in, in these settings and, and it's a worthwhile activity and it's important.  You know, that's something that could change if, you know, like syphilis, you know, historically it was a heterosexual disease and now it's becoming much more MSM disease."

" A couple of big states are a bit concerned because they have overwhelming numbers and they don't normally follow up on Chlamydia and gonorrhea because they've got so much syphilis.  But we think this is going to be a good measure over time."

Code Number:          042

Code Name:            Program Characteristics: Goals, Scope, Context, Culture of CDC Program

Brief Description:     Information about the goals and scope of the national program

Long Description:      General information about the nature of the program, its goals, and its scope (e.g., the STD program addresses several diseases, preparedness is really broad, etc.)

When to Use:  Apply this code to text that reflects aspects of the national program's scope, its goals, and/or its inherent challenges also characterizations of the program or the program context / program culture

When NOT to Use:

Coding Rules:

Example:        "I think we're just in an evolutionary phase of our, of, of this as a public health discipline and I think of it a little bit loose --, because while it is a new section or sector within public health, certainly new responsibilities and new scope of activities, you know, where everyday public health ends and public health preparedness begins is a very fuzzy line."

████████████████████████████

Code Number:          043

Code Name:            Purpose: Accountability Fiscal

Brief Description:     Accountability as stewards of federal funds

Long Description:      As a public agency, CDC has responsibility to act as proper stewards of federal funds (tax dollars).

When to Use:  Apply this code when text refers to CDC's accountability for federal dollars.

When NOT to Use:   Do NOT use when accountability is related solely to demonstrating performance.

Coding Rules: If text relates to accountability for performance, code as "Purpose: Accountability – Performance" Double code with "Accountability – Performance" if text addresses both (see example below).

Example: "Yeah, I think it, I think it's tremendously valuable because I think for one thing just simply to, to, to justify, you know, what the money is being used for and how the, how the program is actually performing in these core areas and to be able to, you know, essentially to, to put a number on it or at least to have a formal definition and be able to monitor people's progress because we've never done that, you know, consistently in the past.

Code Number:          044

Code Name:            Purpose: Accountability Performance

Brief Description:     Accountability for performance

Long Description:      Performance measurement is intended to promote accountability for achieving outputs and outcomes, to demonstrate success.

When to Use:  Apply to text referring to issues of accountability or responsibility for outputs, outcomes, or for performance in general. Also code for text that addresses challenges around accountability as it relates to performance, issues of responsibility, or policies (other than GPRA/PART) that dictate accountability (e.g., PAHPA) – and for instances where text describes grantees avoiding accountability.

When NOT to Use:     Do NOT use when accountability is related solely to fiscal accountability. Do not use when addressing GPRA/PART.

Coding Rules: May double code with "Performance Measurement: Value" if appropriate (e.g., text refers to PM as an incentive for holding people accountable). May double code with "Purpose: Accountability – Fiscal" if text refers to both. If relates to GPRA/PART, code as "GPRA/PART: Perceived Value, Accountability, Attribution"

Example:        "You're responsibility for the whole state in terms of your data, but there seemed to be no movement within the programs to take responsibility for the activities that would affect that whole state's rate."

"Because all of them, you know, just, they'd just say give me the money and, you know, let me run my, let me run my program the way I want."

Code Number:          045

Code Name:            Purpose: Accountability Beliefs

Brief Description:     Individuals' beliefs about accountability, attribution

Long Description:      People have differing views about what accountability means, the importance of accountability

When to Use:  Apply code when text refers to the interviewee's beliefs about accountability, also when text refers to issues of attribution

When NOT to Use:     Do NOT use if text relates to fiscal accountability, accountability for performance, or GPRA/PART.

Coding Rules: Code fiscal issues as "Accountability – Fiscal"; code performance issues as "Accountability – Performance"; code GPRA/PART issues as "GPRA/PART"

Example:        "And then I guess we took credit for what else was going on [laughing]. But we never really, you know, we never really reported it that way, we just knew that, and we didn't have to, to Congress or anybody, but we knew that if our dollars weren't going to support the infrastructure in Massachusetts or New York, they couldn't have, they couldn't have implemented their program"

"You think about the STD program, Division of STD, has never really collected anything other than case-based surveillance data.  Now all of a sudden they want hard data about program.  Well, it's a little intimidating for people in the field that never had to report on what they're doing.  It's absolutely appropriate and necessary."

"Well I think program improvement is joined at the hip with accountability."

Code Number:          046

Code Name:            Purpose: Budgeting

Brief Description:     Performance measurement and its use/relationship to budgeting

Long Description:     Performance measurement may be used to inform budgeting decisions. Some suggest using PM for budgeting encourages improved performance while others have concerns it will lead to gaming, misuse, teaching to the test, etc. Fears that PM may be used (punitively) for budgeting also infuses skepticism into the process (trust) and may compromise the level of buy-in for PM.

When to Use:  Apply this code when text refers to individual perspectives on the use of PM for budgeting (including its value for budgeting, fears about its use), current use for budgeting, formulas for budgeting.

When NOT to Use:

Coding Rules:          Double code if appropriate to "Use – Punitive" or "Use – Misuse"

Example:      "No science base, no consistency of definitions or anything like that, so you are going to have teaching to the test, I think you see it in our data right now even though it's not even, it's low stakes now, because they can just, some of them just say we're not even going to report it you know.  So they teach to the test, they say you want us to convene in an hour? Well we'll do it, we'll figure out a way to do it."

 "Everybody was afraid.  Oh my God, if they do this, they're going to take money away from us if we don't perform well"

**Code Number:** 047

**Code Name:** Purpose: Program Improvement

**Brief Description:** Performance measures as means for program improvement

**Long Description:** Performance measurement is viewed as an important monitoring tool that can identify problems early, track trends over time, etc. – all of which can inform the improvement of the program (through program adjustments, new strategies, etc.)

**When to Use:** Apply this code when the text refers to issues of program improvement and the use of PM for program improvement.

**When NOT to Use:** Do NOT use for other purposes (accountability, budgeting), unless the text reflects two or more ideas

**Coding Rules:** Double code with other purposes if text addresses accountability and/or budgeting as well as program improvement (see example below).

**Example:** "It's, all the other things in between, like accountability, seriously responding to a measure when you're not meeting it, looking for alternatives on how to meet it, how to meet that measure How to keep improving, that's the nature of the whole thing."

**Code Number:** 048

**Code Name:** Quality Assurance Methods

**Brief Description:** Methods used for quality assurance of PM data/system

**Long Description:** Quality assurance is an important component of a performance measurement system. Various strategies or methods can be used for quality assurance such as computer edit programs, periodic audits, peer review, data validation efforts, and involvement of workgroups, etc.

**When to Use:** Apply this code when text describes QA methods or strategies or groups working on QA efforts

**When NOT to Use:**

**Coding Rules:**

Example:     "So the calls actually quickly morphed into not just quality assurance of the data reported, but technical assistance on the measurement, on what was expected in terms of measurement as well."

Code Number:                049

Code Name:             Quality Assurance Benefits

Brief Description:     Benefits to performance measurement of QA efforts

Long Description:     Quality assurance is an important component of a performance measurement system. Quality assurance efforts are undertaken to improve the collection and reporting of data or to identify potential problems/issues in the PM system. Benefits are typically reflected in improved data quality or PM systems.

When to Use:  Apply to text that reflects perceived benefits to the PM data or system based on QA efforts

When NOT to Use:    Do NOT use for QA Methods

Coding Rules: Use "Quality Assurance – Methods" if text references specific QA strategies

Example:        "And then from the [QA] calls we revised the measures slightly and put out the guidance for the following year, you know, for the following budget period."

"We actually found [from the validation study] that the national data was very, I mean, very good, for the most part. There are problems with particular variables, and a lot of that, I think, has to do with, frankly, CDC's lack of guidance."

Code Number:             050

Code Name:             Stakeholders Buy-In Importance

Brief Description:     Importance of stakeholder buy-in in developing PM

Long Description:     The literature suggests that the support and involvement of stakeholders in the development and implementation of PM is critical to its success. Stakeholders are relevant at every level (CDC management, grantees, local level implementers)

When to Use:  Apply this code when text addresses the value and importance of stakeholder support, involvement, buy-in to the development and implementation of PM

When NOT to Use:     Do NOT use if text is describing examples of good/successful buy-in, addressing challenges to buy-in, describing names or types of stakeholders,  or strategies to build support

Coding Rules: Use other 'Stakeholder" codes for related text that doesn't meet definition above

Example:        "It's about change management, buy in, acceptability.  All of those, you know, important aspects. It's the softer side of the work we do but it's the most important."

Code Number:          051

Code Name:            Stakeholder Buy-In Strategies

Brief Description:    Strategies used to nurture buy-in for PM from key stakeholders

Long Description:     The literature suggests that the support and involvement of stakeholders in the development and implementation of PM is critical to its success. Various strategies may be employed to nurture or build buy-in from key stakeholders including engaging them in the development process, providing training on the PM, etc. Stakeholders are relevant at every level (CDC management, grantees, local level implementers)

When to Use:  Apply this code when text describes approaches to building stakeholder support for PM

When NOT to Use:     Do NOT use if text is describing examples of good/successful buy-in, addressing challenges to buy-in, describing names or types of stakeholders, or the importance of stakeholders to the development/implementation of PM

Coding Rules: Use other 'Stakeholder" codes for related text that doesn't meet definition above

Example:         "I think we considered these six low hanging fruit and we thought that they would too, sort of.  (Laughing)  So, you know, there were fewer complaints but still some but, you know, I think that it was partially to get their buy in too and to hopefully see that they, that they would agree that, that, yes, these are reasonable measures"

Code Number:          052

Code Name:            Stakeholders Buy-In Good

Brief Description:    Examples of good/successful buy-in from stakeholders

Long Description:     The literature suggests that the support and involvement of stakeholders in the development and implementation of PM is critical to its success.

When to Use:  Apply this code to text that describes "good" buy-in on the part of stakeholders. Stakeholders could be at any level, including CDC (e.g., mgmt).

When NOT to Use:    Do NOT use if text is describing strategies to achieve buy-in, addressing challenges to buy-in, describing names or types of stakeholders, or the importance of stakeholders to the development/implementation of PM

Coding Rules: Use other 'Stakeholder" codes for related text that doesn't meet definition above

Example:        "I mean it was at that point I think, you know, I don't want to say it was a sea change but there was something of that, there was something that said it no longer was 'our' thing we're foisting on the program, but it became something that had real program ownership"

 "Most project areas, when you talk with the STD program directors, are in favor of performance measures.  They just may not like the specifics on the ones [measures] that we have."

Code Number:          053

Code Name:            Stakeholders Buy-In Challenges

Brief Description:     Challenges to achieving buy-in from key stakeholders

Long Description:      The literature suggests that the support and involvement of stakeholders in the development and implementation of PM is critical to its success.

When to Use:  Apply to text reflecting challenges to achieving stakeholder buy-in (all levels of stakeholders)

When NOT to Use:    Do NOT use if text is describing strategies to achieve buy-in, addressing examples of successful buy-in, describing names or types of stakeholders, or the importance of stakeholders to the development/implementation of PM

Coding Rules: Use other 'Stakeholder" codes for related text that doesn't meet definition above

Example:        "One of the things I think we've learned here is some of the program consultants probably shouldn't go.  And we're dealing with that."

"Especially in project areas where you have so many different people involved in the performance measure, and they may not even know there's a performance measure that you're dealing with like IPP.  A lot of the IPP clinics may not understand that part of the data that they're providing relates back to performance measure."

Code Number:          054

Code Name:            Stakeholders – Names/Types

Brief Description:     Specific Names/Types of Stakeholders

Long Description:      Identity of relevant stakeholders to a program's PM system

When to Use:  Apply to text that identifies relevant stakeholders (e.g., CDC mgmt, CDC program consultants, national organizations, Congress, grantees, local level implementers)

When NOT to Use:     Do NOT use if text is describing strategies to achieve buy-in, addressing examples of successful buy-in, describing challenges to buy-in, or the importance of stakeholders to the development/implementation of PM

Coding Rules: Use other 'Stakeholder" codes for related text that doesn't meet definition above

Example:     "Well, you know, there's a lot of politics around it and, you know, we have the National Coalition of STD Directors and we have a process where we get input on those performance measures and then there's, after we've drafted proposed new ones, there's a period that we're in right now for, for commenting and, and feedback."

"And HHS, ASPR, yeah."

Code Number:          055

Code Name:            Use – Technical Assistance/Tools

Brief Description:     Technical assistance and tools meant to support the use of PM

Long Description:      Performance measures should be used in order to achieve their benefit. The literature suggests technical assistance and training are important to provide to users of the PM system. Other tools or strategies may also support use (e.g., dissemination of data reports to state and local levels)

When to Use:  Apply code when text describes TA efforts (e.g., PM guides, TA calls, training) and other tools to support the use of PM – or the lack thereof

When NOT to Use:

Coding Rules:

Example:      "And I think there's been a lack of really good technical assistance from CDC on this project, so there's a little bit of frustration that"

"We [IMS] have 7 technical consultants, but we actually prefer to have 8." & "Once the measure were, were developed, then we could come along behind and say, OK, you've got, like, 12 measures and I can, I can write a report that out of our, our system we can produce seven of them."

████████████████████████████████████████

Code Number:          056

Code Name:            Use Grantees Use (or Non-Use) /Capacity for PM

Brief Description:     Grantees use of PM data and challenges to their use

Long Description:      Grantees' capacity or motivation to use PM data may differ given their variability

When to Use:  Apply code when text describes grantees' use of PM, their lack of use, their capacity to use PM, and challenges to their use

When NOT to Use:

Coding Rules:

Example: "They may be something that, you know, find it's just another data collection exercise the CDC wants, they'll collect whatever data, but they may not actually be using the performance measures to kind of, you know, direct their program and improve on that -identify weaknesses, or something like that."

Code Number:          057

Code Name:            Use – Misuse / Punitive Use

Brief Description:      Punitive use of PM or other misuse of PM data

Long Description:      Concerns about punitive use of PM or other misuse may inhibit buy-in for the PM system.

When to Use:  Apply code when text refers to concerns or experiences of punitive use or misuse of PM data; when text describes a lack of trust on the part of grantees or others on how the PM data will be used

When NOT to Use:

Coding Rules: Double code to "Purpose – Budgeting" if appropriate

Example:       "And, and, you know, and again, I mean, just getting back to this whole, whole idea that we're trying to be punitive to them, that we're looking for every reason to ding them rather than to help them."

---

Code Number:          058

Code Name:            Use – Program Consultant Role

Brief Description:     Program consultant's role in PM

Long Description:      Program consultants have an on-going working relationship with grantees and may be critical 'change agents" in the adoption and use of PM

When to Use:  Apply code when text refers to the role of Program Consultants in the PM system – whether positive or negative – Their role could involve promoting adoption, supporting use, providing technical assistance, etc.

When NOT to Use:

Coding Rules: Double code with the "Buy-In" family of codes if relevant

Example:       "So, I mean, maybe they're [program consultants] put in, well, they are put in a, a delicate position where they kind of have to be the enforcer."

"A lot of that performance measure kind of implementation lands on us as program [consultants], because we have to go out and implement it to the grantees."

Code Number:         059

Code Name:           Use Political

Brief Description:    Use of PM for political purposes

Long Description:     PM may be used for political purposes – to argue for more funding, to defend against funding cuts, to support policy changes, to back up GPRA/PART, for inclusion in reports that are used as political tools, etc.

When to Use:  Apply to text that reflects political use of PM

When NOT to Use:

Coding Rules: Double code to "Use – Misuse" if appropriate

Example:        "Well, [use PM for] bragging rights, first and foremost [laughing].  I think, you know, we use it as a way, in my grant applications, whether it's to CDC or outside entities, that, you know, in order to be a quality program this is what, this is your documentation that supports you, you have a quality program."

"And then, you know, state and federal legislatures are going to have to make a decision about what's important, but right now we don't have any sort of national data to say, you know, we need more resources."

Code Number:         060

Code Name:           Great Quote

Brief Description:    Really great quote, regardless of topic

Long Description:

When to Use:  Use for any text identified as a terrific quote that may be particularly illustrative of an issue/idea

When NOT to Use:

Coding Rules: Double code with relevant code

Example:

Code Number:          061

Code Name:                   Design: Data Sources, Collection, Reporting, & Management

Brief Description:      Issues of data sources, collection, reporting, or data management related to the design of the PM system

Long Description:      Data sources, collection, reporting, and management systems are central to the design of a PM system. In networked context, issues of capacity and potential burden must be considered in developing these systems. In public health, the collection of data may involve extensive data lags which may affect the utility of the data.

When to Use:  Apply code when text refers to issues of the data system as it relates to the overall design of the PM system (e.g., concerns about data lags, burden of data collection requirements, use of existing data mgmt systems).

When NOT to Use:    Do NOT use when text refers to issues of data quality, data validity, or QA methods and benefits.

Coding Rules: If text refers to issues of data quality and validity, code as "Measurement: Data Quality / Validity Problems".

Example:        "So I don't know if we're going to be requiring measurement at the local level in the near future."

"So, I mean, she can tell you a lot of, you know, the frustrations, the good, the bad, and ugly about the Chronicle."

"… the data is fairly old by the time the program uses it to award money.  It's at least two or three years old.  Because you always have to give enough time for everything to have happened.  And then for the reports to come from IMS and so forth.  So, there are times when we know that things are really turned around but they're being punished for behavior from three years ago."

Code Number:        062

Code Name:          Program Characteristics: Wicked / Complex Problems

Brief Description:   Descriptions or references to "wicked problems" confronting public health

Long Description:    The literature has coined the term "wicked problems" to refer to complex social problems that are typically influenced by multiple factors and require multidisciplinary/multi-sector approaches to effectively address them.

When to Use:  Apply code when text describes the public health problem being addressed by the program (OSH, COTPER) – When text may refer to the lack of or existence of a strong science base for the problem.

When NOT to Use:    Do NOT use this code when text is addressing issues of MEASUREMENT for complex issues

Coding Rules: If text addresses issues of Measurement – look to "Measurement: Challenges" or "Measurement: Hard/ Impossible to Measure"

Example:        "And then in terms of another major division commitment is to reduce "disparities" OK.  What does that mean?"

        "Joe and I were trying to figure out, what is, how do we measure preparedness? And, you know, for me, the, the term preparedness sprung up with no real definition around it.  I mean, are we saying that before we coined that term we were unprepared? Or anything that's, that we know about that's preventable that we're not preventing, does that mean we're not prepared?"

Code Number:        063

Code Name:          Use: CDC Use / Non-Use

Brief Description:   Use or Non-Use of PM data by CDC

Long Description:    CDC leads the PM development process but may or may not effectively use the data collected for management or other purposes.

When to Use:  Apply code when text describes use or non-use of PM data by CDC

When NOT to Use:

Coding Rules:

Example:      "I've been extraordinarily impressed with the fact that not only do we collect the data but we actually use it, which in a lot of places, even at CDC, you know either is not collected, or is not collected well, or if it's collected nobody ever looks at it, or uses it, and I think that we have the full spectrum, so our Division has a lot to be proud in that regard, we manage with data."

"Interviewer.  OK.  OK.  And then just, one of the last questions is about how you're using the data right now.  Interviewee.  We're not [laughing].  Really, to be honest."

███████████████████████████

Code Number:      064

Code Name:      Purpose: Tension between Accountability & Program Improvement

Brief Description:      Perceived contradictory purposes for PM of accountability and program improvement

Long Description:      Some suggest that it is difficult or impossible to develop PM systems that serve both the purpose of accountability and program improvement. The literature has suggested that systems for each of these would require different types of measures, etc.

When to Use:  Apply code to text describing tension between the two purposes of accountability and program improvement

When NOT to Use:    Do NOT use code if text is just addressing accountability or just addressing program improvement…or if it's addressing both but not insinuating any tension between the two

Coding Rules: If text is addressing accountability: "Purpose: Accountability (fiscal, performance)"; If text is addressing program improvement: "Purpose: Program Improvement."

Example:      "Ideally, see, that's, this is the constant battle we have internally between accountability and program improvement for measure, for measures and I think, you

know, because of the legislation that we have in place, our number one requirement right now is to provide measures of accountability."

# APPENDIX I

# SUMMARY OF DESCRIPTIVE DATA BY CASE-PROGRAM CHARACTERISTICS

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| **PROGRAM CHARACTERISTICS** | | | |
| **Organizational Context** | | | |
| Originally funded in 1999; PHEP established in 2002 | Program established in 1957; one of oldest at CDC | Program established by law in 1990 | Initial efforts in early 1990s; NTCP implemented in 1999 |
| 62 grantees representing states, cities, and territories | 65 grantees representing states, cities, and territories | 68 grantees representing states, tribes, and territories (and D.C.) | 58 grantees representing states and territories (and D.C.) |
| Largest grantee program at CDC | Largest program within DSTDP | Largest program within DCPC | Largest program within OSH |
| Extensive staff turnover in all parts of DSLR | Staff often start as CDC field staff working in states; long-term tenure | Turnover frequent among program consultants | More recent turnover throughout OSH |
| Context of constant change and political volatility | Institutionalized culture where change can be difficult | Data-driven program culture | Science-based program culture |
| Politics compromises science | Grantees have had extensive autonomy and have grown resistant to CDC oversight | Legislated program requirements | Faces political opponent in the tobacco industry |
| **Program Goals** | | | |
| Evolving goals and expanding scope over time to "all hazards" | Lacking clear goals and direction from CDC | Consensus on program goals among CDC and grantees | Consensus on four program goals among CDC, grantees, and other national partners |
| No consensus on defining public health "preparedness"; limited science base | Focus on syphilis, gonorrhea, and Chlamydia | Narrow focus on women screened through the NBCCEDP only | Focus on smoking initiation, smoking cessation, second-hand smoke, and reducing disparities |
| Complexity: Program outcomes for "preparing" and "preventing" difficult to define | Complexity: 25 different STDs, shifting disease burden and epidemiology, prevalence differences across grantees, changing health care delivery patterns | Program outcomes reflect clinical outcomes for women served through the NBCCEDP | Program outcomes reflect a comprehensive tobacco control effort inclusive of educational, clinical, regulatory, economic, and social strategies implemented by multiple agencies and supported by diverse funding sources |
| Outcomes dependent on integrated and seamless emergency response at the local, state, and federal levels; collaboration essential | Grantees support a range of program activities implemented at the local level by diverse agencies | Service delivery occurs primarily at the local level | Service delivery occurs at local and grantee level; collaboration and coalition building emphasized |

| PHEP | CSPS | NBCCEDP | NTCP |
|------|------|---------|------|
| **Stage of Program Development** | | | |
| New; blazing new paths, in development | Mature | Mature | Mature |
| CDC recognizes that it must meet states "where they are" given stage of program development | Programs must stay responsive to changes in the health care environment and in epidemiology | State grantees have been funded for 11 or more years | Strong tobacco control community |
| OMEB working to develop a conceptual framework for PHEP | Institutionalized culture difficult to shift | Well established service delivery infrastructure with local providers | Evidence-based logic models developed for 3 of 4 program goals |
| **Program Budget** | | | |
| $700 million FY 2008 | $104 million FY 2008 | $157 million in FY 2008 | $63 million in FY 2008 |
| Median award $8.9 million | Median award $1.1 million | Median award $2.1 million | Median award $XX million |
| Declining federal funding | Flat funding | Flat funding | Flat funding |
| PAHPA (2006) requires performance based budgeting beginning in FY 2009 | HIV/AIDS gets greater percentage of resources | State legislatures contribute resources in some states; advocates are a source for resources as well | State contributions from excise taxes and settlement funds dropping or disappearing completely given state budget crisis |
| Multiple sources of preparedness funds to states from federal government (CDC, DHS, ASPR) | STD not viewed as high priority | Focus on achieving efficiencies given flat funding | 90% of funding for tobacco control are from excise taxes and settlement funds rather than CDC |
| **Stakeholders** | | | |
| Congress, GAO, OMB, President | Congress, OMB | Congress, OMB, GAO | Congress, OMB |
| High powered national stakeholders – DHS, HHS, ASPR, FEMA | | National advocates – Susan B. Komen for the Cure, ACS | National advocates – ACS, ALA, AHA, CTFK, Legacy Foundation) |
| ASTHO's Directors of Public Health Preparedness, NACCHO, CSTE, APHL | NCSD | NACDD, NACCHO, NBCCEDP Council of Program Directors | NACDD |
| CDC Office of the Director; other CDC Centers | Division of HIV/AIDS Prevention, CDC, Division of Cancer Prevention and Control, Division of | CDC's WISEWOMAN™ and CCC programs | Other CDC programs, especially in the NCCDPHP, that share tobacco use as a risk factor (e.g., heart |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| | Reproductive Health | | disease, diabetes, cancer) |
| Grantees | Grantees | Grantees | Grantees |
| **Political Environment** | | | |
| Dominant theme – public visibility, perceived threat of hazards, pressures to "get states prepared" | Stable | Politically visible program given attention to breast cancer – CDC Director reports on related GPRA measure at quarterly HHS briefing | OSH active in CDC D.C. office |
| Political influences of other federal agencies (DoD, FEMA, HHS, ASPR). Top-down, compliance orientation – PAHPA, HSPD-21 | | Policy differences exist across states, especially in regard to mammography screening for younger women aged 40-49 | Political context in individual states can have important programmatic implications for individual NTCP grantee programs |
| Interferes with CDC's control over program implementation | | | Adversary represented by the tobacco industry |
| Grantees have political influence | Grantees have political influence | Grantees have political influence | Grantees have political influence |

**APPENDIX J**

**SUMMARY OF DESCRIPTIVE DATA BY CASE
NETWORK CHARACTERISTICS**

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| **NETWORK CHARACTERISTICS** | | | |
| **Network Structure: Vertical Relationships** | | | |
| 62 grantees; thousands of local-level partners | 65 grantees; hundreds of local-level partners | 68 grantees; 22,000 local-level providers | 58 grantees; some local-level partners |
| Congress – HHS – CDC – grantees, local-level | Congress – HHS – CDC – grantees, regional (sometimes), local | Congress – HHS – CDC – grantees, regional (sometimes), local | Congress – HHS – CDC – grantees, local (if resources permit) |
| Most vertical relationships formalized via funding | Most vertical relationships formalized via funding; dominant dimension | Most vertical relationships formalized via funding; dominant dimension | Most vertical relationships formalized via funding |
| Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory | Vertical structure varies within state, tribe, territory |
| Primarily intergovernmental throughout vertical chain | Regional and local-level partners include: local public health agencies, CBOs, family planning clinics | Regional and local-level partners include: public health agencies, CBOs, private providers, community health centers, family planning clinics | Local-level partners include local health agencies, CBOs (usually funded with non-CDC funds) |
| Program activities at state and local levels | Most program activities at local level | Service delivery at local level | CDC funds primarily support grantee infrastructure; program activities at state and local-level |
| CDC accountable for performance of network partners 1-2 steps removed | CDC accountable for performance of network partners 2-3 steps removed | CDC accountable for performance of network partners 2-3 steps removed | CDC accountable for performance of network partners steps removed |
| **Network Structure: Horizontal Relationships** | | | |
| Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local | Horizontal partners at all levels – federal, CDC, state, local |
| Typically informal and unfunded (e.g., at state-level Departments of education, transportation, emergency response; at local-level schools, first responders, commerce, transportation) | Typically informal and unfunded (e.g., jails, private providers, juvenile detention facilities) | Some funded (e.g., CBOs to conduct outreach and recruitment), others informal and unfunded (e.g., community health agency to assist with referral) | Typically informal and unfunded (e.g., at state-level – Department of education, other public health departments, advocacy organizations; at local-level – health care systems, schools) |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| Essential to achieving program goals – program outcomes are dependent on network efforts | Increasing importance of partnering with horizontal partners to access priority populations that do not seek STD services at public health clinics or are incarcerated | Support referral of priority populations to program services; provide advocacy | Horizontal partners are essential to achieving NTCP program goals – program outcomes are dependent on network efforts |
| Some horizontal partners are funded for preparedness efforts by other federal and state sources, but not for the public health component | Horizontal partners typically facilitate program integration and improved access to populations at risk for STDs | Horizontal partners typically extend program reach, increase access to priority populations, support program integration (e.g., WISEWOMAN™, and contribute to broader public education efforts that benefit all U.S. women | |
| Horizontal partners are integral to any emergency response | "Community paradigm" recognizes horizontal partners as means to expand influence and contribute to population-based effects | | |
| **Network Function: Authority and Control Within the Network** | | | |
| Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation | Control and authority compromised in decentralized implementation |
| As vertical decentralization increases, CDC's authority and control decreases | As vertical decentralization increases, CDC's authority and control decreases | As vertical decentralization increases, CDC's authority and control decreases | As vertical decentralization increases, CDC's authority and control decreases |
| No authority over unfunded, horizontal partners on which grantees are dependent | No authority over unfunded, horizontal partners – some of which grantees are dependent | In vertical chain, control and authority are facilitated by funding and management tools (e.g., MDEs), network relations | No authority over unfunded, horizontal partners on which grantees are dependent |
| Cooperative agreement provides some authority over grantees | Grant offered minimal authority over grantees; recent shift to cooperative agreement beginning in January 2009 | Cooperative agreement provides some authority over grantees | Cooperative agreement provides some authority over grantees, but not exercised |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| Limited control over local-level within vertical chain; grantees models of decentralization may hinder authority by grantee over local-level | CDC has not exercised much authority over grantees during the past 15 years | Dominance of vertical network in NBCCEDP facilitates authority and control over performance at all levels | Grantees are given extensive autonomy and have minimal reporting requirements |
| Building relationships to facilitate coordination is essential in preparedness; Developing relationships with horizontal partners at state and local-level is a critical part of a public health manager's role | Skills to negotiate partner relationships important, but viewed as weak among grantees and local-level staff | | |
| **Network Function: Shared Organizational Goals and Priorities Within the Network** | | | |
| Differing priorities across grantees; goal and mission conflicts with some horizontal partners | Differing priorities across grantees; goal and mission conflicts with some horizontal partners | Shared goals among network partners | Shared goals among network partners |
| Goal conflicts with vertical partners; Priorities of upstream partners (ASPR, HHS, DoD) privileged | Vertical network partners often have their own priorities | Some policy differences within vertical chain (e.g., screening ages for mammography) | Each organization involved in comprehensive tobacco control has their own agenda and perspective depending on the constituency they represent |
| Public health is new to the "emergency preparedness" arena and is challenged to "earn a place at the table" | Different disease burden (epidemiology) likely an influence on goal and priority differences | Some policy differences between CDC and national advocates | |
| Risk for different hazards (and the potential scale of a hazard) likely an influence on goal and priority differences | | | |
| **Network Function: Context, Capacity, and Resources** | | | |
| Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources | Extensive variability across grantees in capacity and resources |
| Variability in level of risk for and type of "hazard" | Variability in STD epidemiology (disease burden, populations affected) | Variability in demographic profile of priority population; cultural barriers | Variability in emphasis around goal areas; Political climate within states often influence program priorities |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| Program performance must be interpreted based on the unique context, capacity, and resources of the individual grantee | Program performance must be interpreted based on the unique context, capacity, and resources of the individual grantee | Program performance must be interpreted based on the unique context, capacity, and resources of the individual grantee | Program performance must be interpreted based on the unique context, capacity, and resources of the individual grantee |
| Variability in level of state contributions | Laws and regulations vary by state – affects authority relationships | Variability in level of state contributions | If include all resource contributions for tobacco control within a state or territory, there is huge variability across grantees |

# APPENDIX K

## SUMMARY OF DESCRIPTIVE DATA BY CASE -
## DESIGN CHARACTERISTICS OF THE PERFORMANCE MEASUREMENT
## SYSTEMS

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| **PERFORMANCE MEASUREMENT SYSTEM** | | | |
| **Purpose** | | | |
| Accountability "up" to OMB, Congress, and others "above" CDC | Accountability for performance and program improvement | Accountability to the women served, Congress, OMB, and the public | Accountability "up" to OMB, Congress, and others "above" CDC, and to state legislatures |
| Pressured to demonstrate that the nation is "prepared" | Adamant that performance measures will not be used for budgeting | Program improvement at all levels; a subset of the performance measures are part of a budgeting formula | Important to demonstrate accountability for CDC and other funding (e.g., excise tax revenues, MSA settlement funds) |
| Other sets of measures will be used for monitoring, program improvement, and budgeting | Performance measures intended to communicate program priorities | Performance measures intended to communicate program priorities | Performance measures intended to provide a "national picture" of comprehensive tobacco control efforts |
| The 3 GPRA/PART measures relate closely to six PHEP performance measures | Performance measures are aligned with GPRA/PART measures, although all but 1 of the GPRA/PART measures are population-based | Of the 3 GPRA/PART measures, 1 is specific to women served in the NBCCEDP and the other two are population based | Of the 3 GPRA/PART measures, one is a core indicator – all our population-level measures |
| **Level of Measurement** | | | |
| Grantee-level | Primarily local-level with a couple that are grantee-level related to timely and complete data submissions to CDC | Local-level | Primarily grantee but some local-level |
| Six measures considered "low hanging fruit" – that is, what most grantees can address and those measures for which data are available for most grantees | First group of performance measures were "GRAM" ("get right at 'em") measures, that is, what most grantees can address and those measures for which data are available for most grantees | Patient-level, clinical measures | A consideration in selecting core measures were those "low hanging fruit" – measures for which data are available for most grantees |
| **Types of Performance Measures** | | | |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| The current six measures are process measures reflecting "programmatic capabilities" | Process measures (e.g., proportion of females admittees to large juvenile detention facilities tested for chlamydia) | Process measures (e.g., percentage of screening mammograms provided to women aged 50 and older) | Program found that common process measures were impossible to identify given variation in program priorities and activities across grantees |
| Intention to develop intermediate measures in the future | Short-term outcome measures (e.g., among clients of STD clinics, the proportion of women with positive chlamydia tests that are treated within 14 and 30 days of the date of specimen collection) | Short-term outcome measures (e.g., percentage of abnormal screening results with time from screening test to final diagnosis less than 60 days) and Intermediate-outcome measures (e.g., percentage of women diagnosed with breast cancer with treatment started) | Core measures reflect outcomes at the immediate, intermediate, and long-term levels |
| Distal measures are unlikely to be included in future performance measures given that multiple factors contribute to them | Distal, population-level measures (e.g., changes in morbidity and mortality) viewed as not helpful for performance measures given that multiple factors contribute to them, they can take years to achieve, and CSPS resources cannot support a population-level impact | Distal, population-level measures (e.g., changes in morbidity and mortality) viewed as not helpful for performance measures given that multiple factors contribute to them, they can take years to achieve, and CSPS resources cannot support a population-level impact | Concerns about data availability for core indicators across all grantees given cost to participate in various surveys from which most data are derived |
| 3 GPRA/PART measures reflect 3 of the 6 performance measures aggregated at the national level | Population-based GPRA/PART measures viewed as unrealistic and not useful | Population-based GPRA/PART measures viewed as unrealistic and not useful | Population-based GPRA/PART measures viewed as unrealistic |
| **Use of Targets** | | | |
| National standards set by CDC for 4 of the 6 measures | Grantees set their own 3-year targets given their individual baseline. | National standards set by CDC for all 11 measures | Grantees set their own targets given their individual baseline. |

| PHEP | CSPS | NBCCEDP | NTCP |
|---|---|---|---|
| | DSTDP does not believe there is enough trend data to consider national indicators | Targets based on trend data for national program and policy decisions | Variability across grantees in terms of baseline data for the measures makes establishing national standards difficult. |
| **Quality Assurance Efforts** | | | |
| OMEB staff conducted individual telephone calls with grantees to address data quality issues for the first two data submissions | Different people with different skill levels enter data from local and state levels. | Different data systems across grantees used although CDC provides an optional software system for data management to grantees | Not reporting on core measures yet – still in development |
| Validation assessment conducted in 2007 identified major quality assurance issues (e.g., misunderstanding about the intent of the measures, uncertainty regarding the definition of the measures and how each is calculated, poor adherence to measurement protocols) | Data quality problems have been identified and documented including misunderstanding the measures, capacity issues affecting data collection and reporting from local-level, and differences in data management systems across grantees. Challenging for DSTDP to assess quality across so many grantees. | National validation study of MDE data was conducted and supported strong data quality for program. Some individual grantees conduct chart reviews and use other quality assurance practices to assess data quality. | In regard to broader set of key outcome indicators, data concerns relate to frequent "customization" of the measures rather than following measurement protocol and definition. |
| OMEB has provided guidance to grantees on the performance measures including a document with detailed definitions, measurement specifications, and data collection methods. | DSTDP has developed technical assistance materials and strategies: A guidance document on the performance measures with detailed information about the intent of each measure, definitions, measurement specifications, data sources; Performance measurement "Learning Tours" involving on-site review of data quality and provision of technical assistance | Extensive data quality system in place including standard edit programs to assess data quality, technical consultants who work for the NBCCEDP data contractor that work individually with grantees on data-related issues, semi-annual review of data quality as part of MDE data assessment. DCPC provides grantees a detailed MDE data dictionary CDC requires that grantees include a staff position for a data manager. | Much of the data supporting the core indicators will come from survey data and should be of high quality. A few of the core indicators require local-level data collection using data sampling and collection protocols that may be difficult for some grantees or may not be implemented consistently across grantees. |

# APPENDIX L

# MATRIX TO FACILITATE DEVELOPMENT OF CASE-SPECIFIC FINDINGS

| Research Question 1: How does networked public management affect the observability of program outputs and outcomes? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ-1** |
| **PHEP** | **Variability of "hazards" across network and rarity of hazards affects observability of specific outcomes**. Risk of hazard varies (e.g., hurricanes not likely to be observed in some parts of network). Common "observable" measures may be difficult to define in a network. Also feasibility of accomplishing some possible outcomes varies given grantee context (48-hr prophylactic distribution). Together, the complexity of preparedness and network structure leads to challenges in observing outcomes. | Relates to challenge of identifying common measures that are meaningful to all network partners. "Low hanging fruit" selected early on. | H |
| | **Network fragments accountability for outcomes**. Complexity of problem necessitates network response (at multiple – vertical – levels AND from multiple – horizontal – sectors). Outcomes can't be attributed to grantee alone. | Based on observability as perspective of attribution and accountability | H |
| | Political environment (top level federal partners in network) contradicts science-based approach to identifying or defining observable program outcomes. | PAHPA, HSPD-21 | Unique to PHEP |
| | **Complexity of preparedness is an important factor in challenging observability of program outcomes.** Complexity is reflected in the lack of a conceptual framework for PHEP which means OMEB cannot identify outcomes; lack of science base; expanding scope; need for multi-sector response at all levels of government; and difficulties to measure "prevention" "synergy" "collaboration.". PHEP measures represent process measures more than outcomes since outcomes are difficult to identify at this stage of the program's development. | What is it to "be prepared"? Unique political context which has contributed to an expansion of program scope and demands, "Are we prepared?" Important to discussion of interplay of *program complexity and networks* | M |

| Research Question 2: How does networked public management influence the use of performance measurement and the types of performance measures used? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ2** |
| **PHEP** | **Need to "meet states where they are."** Broad network led to choice of "low hanging fruit" to accommodate varied capacity, resources, data availability, and to respond to accountability demands from above. | Complexity, political context, and stage of program development also relevant | H |
| | **Political influence of network members affects types of measures selected**. Push-back from grantees if types of PM do not seem realistic, feasible, meaningful, or if grantees do not feel they have control over affecting performance for the measures. The network (especially grantees) requires a negotiation over types of measures selected. Policy requirements (political context) require use of performance measures and even specifies use of some measures (top-down network influence) by PHEP. Political pressure for accountability "up" | Policy tool may be a factor here given "cooperative" nature; HSPD-21, PAHPA – Unique to PHEP | H |
| | **Network influences level of measures selected**. For PHEP, difficult to include local-level outputs and outcomes because of network structure – local level network partners lack capacity to collect and report data at local level in network. Using state-level measures for now. | Newness (stage of development) of program may contribute to lack of data collection/reporting infrastructure | M |
| | **Network blurs who is responsible for what**. Hard to know WHO (which network member) is responsible for WHAT indicators given that outputs/outcomes often reflect work of many. | Fuzzy boundary and joint production problems | M |
| | Different network partners want PM to serve unique purposes (CDC wants PM for accountability; grantees want PM for program improvement) – Purpose affects types of measures selected | Tension between accountability and program improvement unique to PHEP | H |

| Research Question 3: How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 3** |
| **PHEP** | **As vertical and horizontal decentralization increase, DSLR's ability to influence performance on outcomes decreases.** CDC has little to no control over local level network partners; vertical structures vary within states; preparedness demands reliance on horizontal partners. Goal conflicts between vertical and horizontal partners further reduces control over outcomes. | Preparedness is achieved through joint production – contributions from varied levels of government and from various sectors | H |
| | **Grantees demand measures that reflect outputs and outcomes for which they have control – leads to process measures**. Expert groups added criteria related to "control" for measure selection. Grantees argue it is not fair to make them accountable for performance on measures for which they don't have control. Network may lead to choice of measures "closer to the work" of grantees (that they have control of and can affect and can be accountable for). May support selection of process measures. | | H |
| | **Institutional and funding arrangements facilitate authority and control within vertical network.** For PHEP, virtually no control over horizontal partners – but PHEP dependent on them. | Joint production of outcomes by vertical and horizontal partners | M |
| | **Network demands extensive stakeholder involvement in development of PM system – collaborative effort rather than a CDC effort alone**. Network partners influence design. Need extensive QA to assure data validity in network. Adoption incremental given number of grantees. | | |
| | Network demands sophisticated performance mgmt system to facilitate use | Data validity problems to date | M |

| Research Question 1: How does networked public management affect the observability of program outputs and outcomes? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 1** |
| **CSPS** | **Variability across network in STD epidemiology challenges observability of outcomes**. Outcomes may be observed in some parts of network and not in others (e.g., outcomes related to syphilis may not likely to be observed in some parts of network). Common "observable" and meaningful measures may be difficult to define in a network. | STD identified alternative measures for low morbidity areas, but some thought they were not priority measures | H |
| | Complexity issue Some longer-term effects may take years to detect, especially given minimal resources to affect them. Science base lacking for some measures and don't know how to measure some outcomes (PID). 25 different STDs, differing and shifting disease burden. | Important to discussion of interplay of *program complexity and networks* | M |
| | **Network structure fragments accountability**. Complexity of problem necessitates network response (at multiple – vertical – levels AND from multiple – horizontal – sectors). Outcomes can't be attributed to grantee alone. DSTDPs move to a 'community approach' shifts public health role to influencing partners (jails, private providers). | Based on observability as perspective of attribution and accountability. Important to discussion of interplay of *program complexity and networks* | H |
| | **Differences of opinion across networks of what outcomes should be observed**. Extensive network has led to differences in ideas about what key outputs and outcomes should be prioritized – There is a lack of consensus about priority outputs and outcomes needed to reduce long term morbidity and mortality related to STDs. | Participants suggest lack of leadership at CDC to define clear goals and priorities – Unique to DSTDP | L |
| | **Capacity and resource differentials across network limit ability collect and report data outcomes.** Outputs and outcomes will be observed several levels below CDC (local level) and across 100's of network partners. While the outputs and outcomes may be observable, there may be challenges related to data availability across network and/or ability to collect/report output and outcome data to grantee. | | M |

| | | | |
|---|---|---|---|
| **Research Question 2: How does networked public management influence the use of performance measurement and the types of performance measures used?** | | | |
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 2** |
| **CSPS** | **Choose "get right at 'em" measures**. Broad network led to choice of GRAM ("get right at 'em) measures to accommodate varied capacity, resources, data availability of grantees and to support grantee buy-in – DSTDP hopes for more sophisticated measures in future | Complexity and political context also relevant | H |
| | **Network influences level of measures selected**. Local level measures used given that most program activities occur at this level (2-3 steps below CDC) | Level of measurement | H |
| | **Political influence of network members affects types of measures selected.** Network members have political influence and resist if types of PM do not seem realistic, feasible, meaningful or consistent with their priorities | Types of measures – Some proposed 2009 measures rejected | H |
| | **Changes in context and network structure effects types of measures included**. Complexity: Changing patterns in health care seeking behavior (shift from pubic STD clinics to private providers) results in changes in network structure and need to move to broader, population-level PM | Complexity of public health problem interacting with network structure | |
| | **Variability in network.** Large network introduces data quality issues that affect use of PM data; use of PM to make comparisons across grantees not helpful given their unique context, resource differentials, etc. With broad network that includes 100s of local level partners, the adoption of PM and the use of PM data has been slow. | | L |
| **Research Question 3: How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement?** | | | |
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 3** |
| **CSPS** | **Authority and control over outcomes compromised given network structure**. Control weakens with greater vertical decentralization (local level partners) and out (no authority over horizontal, unfunded | Control over jails, juvenile centers, some family planning clinics difficult | H |

| | | | |
|---|---|---|---|
| | partners); varied state structures affect level of control; funding tools vary between grantees and local partners; laws and regulations differ in states. | | |
| | **Mission and goal conflicts lessen control and authority with local and horizontal partners.** | Jails have mission of public safety vs. public health | H |
| | **Pushback from grantees for PM where performance is reliant on partners over which they have little control.** More process/short-term outcomes (venue-specific) selected vs. population level outcomes. | Guidance acknowledges measures outside their control. | H |
| | **PM help communicate program priorities within a broad network. Cooperative agreement as means to strengthen authority in requiring reporting of PM.** | | |
| | **Network demands extensive stakeholder involvement in development of PM system – collaborative or negotiated effort rather than a CDC effort alone**. | Network partner influences PM design | |
| | **Network requires that flexibility needed in PM system.** Grantees set their own targets given variability in terms of disease burden, but also in resources, capacity, and context. | Also, some choice in measure selection for low morbidity areas (syphilis) | |
| | **Incremental approach needed given extensive network**. Time for adoption and buy-in to occur at all levels – even CDC. Champions in DSTDP. **GRAM measures first selected.** | Organizational culture resistant to PM | |

| Research Question 1: How does networked public management affect the observability of program outputs and outcomes? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 1** |
| **NBCCEDP** | **Nature of the program (service delivery) and decision to focus on outcomes only for women served through the program facilitates observability of program outputs and outcomes.** HOWEVER, some outcomes less observable in parts of network given community size (tribes and territories have smaller numbers of cervical cancer cases) | Unique to NBCCEDP | H |
| | **Nature of program, dominance of vertical dimension, and use of specific polity tools facilitate attribution of outcomes to program.** | Based on observability as perspective of attribution and accountability. Important to discussion of interplay of *program complexity and networks* | H |
| | **Some longer-term effects may take years to detect, especially given minimal resources to affect them** (NBCCEDP serves 1% of population). Science base lacking for some measures (re-screening rates) and don't know how to measure some outcomes (disparities). | Important to discussion of interplay of *program complexity and networks* | M |
| | **Outputs and outcomes will be observed several levels below CDC (local level) and across 1000's of network partners**. Outputs and outcomes are observable, in part, given large investment in data management system and staff capacity to collect/report data. | Offers contrast to CSPS and PHEP | M |
| | Observability of treatment initiation (outcome) obscured by policy factors within network (HIPAA) that require use of proxy | | L |

| Research Question 2: How does networked public management influence the use of performance measurement and the types of performance measures used? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 2** |
| **NBCCEDP** | **Purpose of PM, nature of program, and network all influence type of measures selected**. Types of measures influenced more by program goals/priorities and the nature of the NBCCEDP as a service delivery program – led to selection of PM related to clinical services. Selected types of measures for which programs can be held accountable and are actionable. | Recognition that population-level measures would not provide useful information in network context. | H |
| | **Political influence of network members affects types of measures selected.** Network members (grantees) push-back if types of PM do not seem realistic, feasible, meaningful or consistent with their priorities (issues with cervical measure). DCPC has revised measures based on their feedback. | Types of measures | H |
| | **Extent of network and its variability has effects on PM** lends comparisons across grantees not helpful given unique context, resource differentials; adoption of PM has taken time given size of network (time to create a data-driven program culture at all levels that supports submission of quality data and use of PM data); requires incremental implementation/use of PM over time. Large network introduces data quality issues that require extensive QA and TA system. | | L |

| Research Question 3: How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 3** |
| **NBCCEDP** | **Service delivery occurs 2-3 steps below CDC and some authority and control is compromised given network structure**. | CDC has little influence on local level providers | H |
| | **Funding relationships strengthen authority and control down vertical chain within NBCCEDP. Cooperative agreement used by DCPC to improve authority in requiring PM** Focus on program-funded screenings also facilitates authority and control. Dependence largely on vertical chain increases authority and control. | Funding tools important – partner relationships important | H |
| | **PM communicate program priorities within a broad network.** Strong consensus across network on program goals and priorities | | M |
| | **Network demands extensive stakeholder involvement in development of PM system – collaborative and negotiated effort rather than a CDC effort alone**. Push back on measures seen outside grantees control (cervical measure); revise measures seen as problematic. | | H |
| | **Network variability requires some flexibility needed in PM system.** | PM not calculated if small numbers | M |
| | **Longer term outcomes not viewed as fair to be held accountable for (GPRA measures) given network structure and complexity of problem.** | Morbidity and mortality influenced by other factors than screening alone | L |
| | **Incremental approach to PM implementation needed given extensive network** – time for adoption and buy-in to occur at all levels. | | M |
| | **PM design involves extensive performance management system that supports quality and use.** | Significant and sustained investment by CDC | H |
| Research Question 1: How does networked public management affect the observability of program outputs and outcomes? | | | |

| Case | Findings | Notes | Importance of Case to RQ 1 |
|------|----------|-------|----------------------------|
| **NTCP** | **Network structure fragments accountability.** Complexity of problem necessitates network response (at multiple – vertical – levels AND from multiple – horizontal – sectors). Outcomes can't be attributed to grantee alone (especially since CDC primarily supports infrastructure vs. implementation activities). OSH has accepted this as a given and aims to observe effects of *comprehensive tobacco control* efforts broadly. | Based on observability as perspective of attribution and accountability. Important to discussion of interplay of *program complexity and networks* | H |
| | **Variability across network in smoking prevalence challenges observability of outcomes**. **Variability across grantees in priorities (which goal area) affects observability of outputs and outcomes.** Science base strong (logic models) and has facilitated identification of observable program outputs and outcomes. HOWEVER, tobacco burden varies across grantees making some outcomes more or less observable across network. | Important to discussion of interplay of *program complexity and networks* | M |
| | Network has adopted four goals which facilitates identification of program outputs and outcomes. HOWEVER, the extensive NTCP network challenges ability to identify common outputs and outcomes given that grantees address different goal areas (due to resource limitations, political context, etc.) and that grantees implement different activities. | OSH was unable to identify common process measures given variability in which activities are implemented and how they are implemented | H |
| | **Capacity and resource differentials across network limit ability to collect and report data ("observe") outcomes.** While outputs and outcomes may be observable, there are challenges across network to 1) collect/report local-level valid data given resource and capacity issues and 2) data availability across network (expense of survey data). | | M |

| Research Question 2: How does networked public management influence the use of performance measurement and the types of performance measures used? | | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 2** |
| **NTCP** | **Select "low hanging fruit."** OSH selected PM for which most/all states have data – the "what's available" approach. | "Low hanging fruit" concept also observed in PHEP and CSPS | M |
| | **Consensus across network on broad goals, purpose of system, and strong science base facilitate decisions around types of measures included.** Agreement on four goal areas, outcomes for accountability, science base supporting outcomes at three levels. | Goal consensus – also exists in NBCCEDP | M |
| | **Variability across network in program activities influences type of measures used**. Outcomes versus process measures. | Common process measures possible for NBCCEDP, PHEP, CSPS | H |
| | **Variability across grantees in goal area priorities, program activities, and implementation approaches, influences choice of PM used.** | | H |
| | Extent of network and variability of network have led to data quality concerns for KOIS. Comparisons across grantees not helpful given unique context, resource differentials, etc. | Core indicators not implemented yet | L |

| | Research Question 3: How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? | | |
|---|---|---|---|
| **Case** | **Findings** | **Notes** | **Importance of Case to RQ 3** |
| **NTCP** | **Choice of outcome-level measures means grantees have little control over outcomes.** Integrated funding streams and broad comprehensive tobacco control approach leads to outcomes that reflect comprehensive tobacco control efforts. | Recognition that outcomes would only be affected through integrated, comprehensive approach. May support evaluation more than PM. | H |
| | **Challenge to identify a small, core set of indicators**. Grantees focus on different goal areas and implement different activities. | Couldn't identify common process measures | M |
| | **Decentralization lessens CDC control over local level activities.** CDC dollars typically support grantee infrastructure and may be a small part of overall resources in a given state. | Large contributions to tobacco control efforts from excise taxes and MSA funds | M |
| | **Flexibility needed in system** given grantees focus on different goals, different political context, and different activities. | Target setting, choice of KOIs | L |
| | **QA challenges will likely emerge with local-level core measures in decentralized context**. | Core measures not implemented yet – specialized protocols for data collection under development. | L |

# APPENDIX M

# MATRIX TO FACILITATE DEVELOPMENT OF CROSS-CASE FINDINGS

| Jennings & Haist (2004) Hypothesis: "The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed" (p185) | Research Question #1: How does networked pubic management affect the observability of program outputs and outcomes? [Jennings and Haist define observability defined as ability to measure outputs and outcomes and ability to attribute outputs and outcomes to the program] |
|---|---|

**Finding #1: Network pubic management fragments the PHEP program's accountability for results, creating challenges for performance measurement.**

| | |
|---|---|
| Effect on performance measurement system | Process measures selected that are more closely tied to the work of grantees; Grantee-level measures selected rather than local level measures that may be jointly produced |
| Broader implication for performance measurement | Problematic to include output and outcome measures that are jointly produced; Impossible to discern accountability or attribution for jointly produced outputs and outcomes, especially outcomes further along the results chain; Pragmatic challenges of data collection and management for jointly produced outputs and outcomes; Challenges to measure some constructs such as collaboration and synergy. |
| Case examples | Difficult to get partners to participate in public health-related preparedness activities (e.g., planning for pandemic flu); Grantees resistant to performance measures that they feel are unfair to be held accountable for because performance on the measure is outside their control. |
| Issues of case context (situationality) | Achieving outcomes for preparedness demands a coordinated response across levels of government and sectors at each level – vertical and horizontal relationships involved in a response muddy efforts to assign accountability. |
| Other supportive case examples | Network also fragments accountability in CSPS and NTCP; NTCP closest example to jointly produced outcomes. |
| Negative case example | NBCCEDP processes, short-term and intermediate outcomes are easier to observe because program involves service delivery and relate only to women served through the program – accountability is more easily maintained within the network |
| | |

**Finding #2: Network consensus on goals, a strong evidence base, and extensive survey data facilitate identification of outcome measures for comprehensive tobacco control.**

| | |
|---|---|
| Effect on performance measurement system | Network consensus on four goals is part of what facilitates identification of 120 key outcome measures for NTCP from which core indicators have been selected. Strong science base allows focus on outcomes. |
| Broader implication for performance measurement | Network consensus on goals facilitates identification of common performance measures even with variation in processes across grantees – accountability assigned to efforts of entire comprehensive tobacco control program, not just CDC-funded efforts (conscious decision to accept limitations related to attribution). |
| Case examples | Four program goals with three evidence-based logic models facilitated the identification of 120 KOIs using rigorous, panel process; More challenging to identify small set of core indicators (that all grantees will report) from that larger set because grantees focus on different |

| | |
|---|---|
| | program goals and few have resources to address all four. |
| Issues of case context (situationality) | National partners in tobacco control and all grantees have adopted four common goals for tobacco control; Strong evidence base after years of well-funded research; Extensive survey data available at national level from several sources; performance measurement needed for accountability "up" and to defend resources at state and federal levels; CDC interest to provide a "national snapshot" of tobacco control. |
| Other supportive case examples | Network consensus on program goals and activities in the NBCCEDP facilitates identification of common outputs and outcomes. |
| Negative case example | Lack of consensus on goals (more an issue that the program is not well understood yet – a developmental issue), weak science base, no survey or surveillance data challenge PHEP to identify outcome measures; Weak agreement on program goals (including weak leadership on goals and priorities at CDC) challenges identification of outcomes for CSPS, although logic models help. |
| | |
| **Finding #3: The network implementation structure of the NTCP leads to a joint production of outcomes and shared accountability** | |
| Effect on performance measurement system | Decision to focus on outcomes and to openly recognize that outcomes are jointly produced with shared accountability for outcomes among all those participating in comprehensive tobacco control work. |
| Broader implication for performance measurement | Accountability is fragmented in a network structure that involves a dominant horizontal dimension; Impossible to decipher who is accountable for what. |
| Case examples | KOIs and core indicators represent outcomes for "comprehensive tobacco control" efforts, not CDC-funded NTCP; Explicit examples available for each program goal. |
| Issues of case context (situationality) | OSH funds for NTCP primarily support *infrastructure costs* for grantees, not program delivery; CDC funds are often only a fraction of state funds available (excise taxes and MSA); Strategy of comprehensive tobacco control recognizes importance of multiple types and levels of interventions. |
| Other supportive case examples | PHEP reliant on jointly produced outcomes to ensure "preparedness;" CSPS increasingly reliant on networked response involving private physicians, jails, etc. in order to achieve population-level effects inherent in their "community perspective." |
| Negative case example | NBCCEDP outputs, short-, and intermediate outcomes observable because all programs do the same thing (service delivery) and focus on same goal; Accountability more easily maintained within vertical network. |

| Jennings & Haist (2004) Hypothesis: "The extent to which performance measures are used and the types of measures used will depend on the degree to which outputs and outcomes can be observed" (p185) | Research Question #2: How does networked public management influence the use of performance measurement and the types of performance measures used? |
|---|---|
| **Finding #1: In the case of the PHEP program, performance measurement is a "political, social, and scientific" process.** | |
| Effect on performance measurement system | Performance measurement development process is negotiated and incremental; Political initiatives shape the performance measurement system; Lack of science makes system vulnerable to political influence. |
| Broader implication for performance measurement | Multiple factors influence the design of performance measurement system; Network structure implies many levels of stakeholders involved in development and implementation process; performance measurement development is incremental and negotiated process adoption will take time. |
| Case examples | Multiple performance measurement sets developed for different purposes for PHEP; Evolving mission and expanding scope reflecting political influence; Policy initiatives HSPD-21, PAHPA big factors affecting PHEP performance measurement; Extensive stakeholder involvement in process via Evaluation Workgroup, expert groups, meetings with federal stakeholders in D.C. |
| Issues of case context (situationality) | PHEP in highly political context; Lack of science base; Extensive network interested in performance measures and pressure for performance measures as means to show that "we're prepared" grantees, ASPR, Preparedness Directors, HHS, DoD, DHS, etc. |
| Other supportive case examples | PHEP more or less unique in the extent of political influence; NBCCEPD has political influences on some measures – i.e., to allow flexibility in measure on providing mammograms to women ages 40-49. |
| Negative case example | Less political influence in CSPS and NTCP, although grantees as key stakeholders have political power and influence. |
| | |
| **Finding #2: Given the CSPS network context, performance measurement is a negotiated and incremental process.** | |
| Effect on performance measurement system | Process to develop performance measures has been slow and negotiated; Adoption has been challenging which has affected use and data quality (lack of endorsement by DSTDP management and program consultants; resistance to performance measurement by grantees). |
| Broader implication for performance measurement | Network structure implies many levels of stakeholders involved in development and implementation process; performance measurement development is incremental and negotiated process adoption will take time; For CSPS, adoption is required at hundreds of local-level sites; To facilitate adoption in a network, extensive management system is needed to support data quality, use, buy-in. |

| Case examples | CSPS is not yet a data-driven culture; Grantees are resistant to requirements; Program consultants sometimes entering data for grantees; Process of workgroups negotiating measures, piloting, vetting processes; Poor data quality at this time, although improving; Grant has given much autonomy to grantees – move to cooperative agreement with new five-year award in January 2009. |
|---|---|
| Issues of case context (situationality) | Organizational culture issues Much autonomy and few requirements for grantees over the years; Not particularly a data-driven organizational culture; Long-standing program with institutionalized organizational culture that is difficult to shift. |
| Other supportive case examples | NBCCEDP and PHEP also incremental and negotiated performance measurement development processes. NBCCEDP has incrementally expanded the use of their measures – most recently to budgeting. |
| Negative case example | N/A |

| | |
|---|---|
| **Finding #3: The variability across the PHEP program's vertical network significantly shapes the design of its performance measurement system.** [Variability: risk and scale of potential event/hazard; Program context; Program priorities and activities: Capacity and resources; Data availability and sophistication of data collection systems]. | |
| Effect on performance measurement system | DSLR selected "low hanging fruit" at start; Can't compare grantees' performance; Data quality issues; Selection of grantee-level measures vs. local-level measures; Difficult to identify common measures. |
| Broader implication for performance measurement | Network variability in vertical chain – in terms of characteristics of public health problem (risk/scale of hazard), in grantee organizations (resources, capacity, context, data availability); Overall, variability effects all aspects of the performance measurement system – choice of measures, level of measurement, design of system (data quality mechanisms, flexibility in system). |
| Case examples | Capacity and data availability issues led to selection of grantee-level measures even though "all preparedness is local;" Extensive quality assurance problems with individual telephone calls to grantees needed; Selected "low hanging fruit" at start viewed as ones acceptable by grantees and for which data was available across the grantees; Decision not to compare grantee performance given their unique contexts. |
| Issues of case context (situationality) | Huge variation in resource levels across grantees that affect capacity for performance measurement, including data collection; Differences in risk levels, potential hazards, and scale of event lead to different priorities and activities across grantees; Limited control over local-level although they get 50% of grantee funds. |
| Other supportive case examples | Three of four cases selected "low hanging fruit" at start in order to accommodate variability across grantees, especially in regard to capacity and data availability; None of the four cases compare performance across grantees; Quality assurance big issue for all four cases; Flexibility in performance measurement system is built into 3 cases. |
| Negative case example | Common set of performance measures possible for NBCCEDP given that service delivery is same across grantees and program priorities are tied to those services; Local-level measures possible given outcomes tied to local-level service delivery; Capacity developed over time for |

| | data collection/reporting – large commitment of CDC resources to build that capacity and support data quality. |
|---|---|
| | |

| **Finding #4: Extensive variability across grantees influences the design of the CSPS performance measurement system.** [Variability in epidemiology disease burden, populations affected; Geographic context; Capacity and resources; Program priorities and activities; Data availability and sophistication of data collection systems]. | |
|---|---|
| Effect on performance measurement system | Difficult to identify common set of measures meaningful to all grantees; Selected GRAM (easiest) at start; Need for flexibility in system for grantees; Can't compare grantee performance; Data quality issues. |
| Broader implication for performance measurement | Network variability in vertical chain: Variability in terms of characteristics of problem (STD burden; populations affected); Variability in terms of grantee organizations (geog context, resources, capacity; data availability) DO THEY HAVE DIFFERENT IMPLICATIONS? In total variability effects all aspects of the PM system choice of measures, level of measurement, design of system (data quality mechanisms, flexibility in system. |
| Case examples | Some states with little syphilis, so syphilis-related measures not really relevant; Added choice of measures for low morbidity areas (but some staff disagreed with approach); Allow grantees to set their own baseline and targets; Selected GRAM measures; Decision not to compare grantees since "apples and oranges." |
| Issues of case context (situationality) | 25 different STDs and CDC focuses on 3; Epidemiology of these 3 STDs varies by grantee in terms of disease burden and populations affected; Some states have huge geographic area and part-time staff given small resources while others have lots of DIS and CDC assignees. |
| Other supportive case examples | All four cases are confronted with this issue of variability – NBCCEDP's variability seems to have the least affect on their performance measures given it is a service delivery system. However, even the NBCCEDP has had to accommodate this variability in their system by not calculating measures if a small "n," not comparing grantees given variable context, etc. |
| Negative case example | N/A |
| | |

**Finding #5: The NBCCEDP network requires that DCPC make significant resource investments to build a comprehensive performance management system in order to ensure data quality and the use of performance measurement data at multiple levels.**

| | |
|---|---|
| Effect on performance measurement system | Large financial investment in performance management and data monitoring system for NBCCEDP; Strong data quality; Data reporting requirements in cooperative agreement. |
| Broader implication for performance measurement | Implementing performance measurement in a large network context requires extensive resources and commitment to build adequate performance management system to support data quality and data use at all levels. |
| Case examples | NBCCEPD has supported a data contractor since the program's inception; Requirement of data management staff for all grantees; Training provided on-going; Extensive technical assistance for data management efforts; Monitoring cycle instituted to regularly review and discuss data with grantees; Software provision; Consistent managing with data. |
| Issues of case context (situationality) | DCPC has built a data-driven culture around the NBCCEDP from the start; DCPC has advantage of a service delivery program with the NBCCEDP that has made measurement of relevant processes and outcomes feasible. |
| Other supportive case examples | PHEP and NTCP do not have mature enough performance measurement systems to assess this aspect. |
| Negative case example | Use of performance data and data quality are limited in CSPS because DSTDP lacks a data-driven CSPS program culture; Lacks sophisticated performance management system to reinforce data use and data quality. |

| |
|---|
| |

**Finding #6: Network variability limits types and choice of NTCP core measures and their use**
[Variability: state political context; tobacco control priorities; program priorities and activities; capacity and resources].

| | |
|---|---|
| Effect on performance measurement system | Could not identify common process measures; Difficult to select common outcome measures because grantees focus on different goals "low hanging fruit" a factor for selection since data is not available for all grantees; Use of performance data may be limited to evaluation given focus on outcomes; Grantees select KOIs and reporting is optional (flexibility); Reporting of core indicators is not a requirement of new 2009 FOA. |
| Broader implication for performance measurement | Within vertical network, variability in grantee's resources, goal priorities, and context influences choice and types of performance measures. |
| Case examples | Only outcome measures are used because grantees implement activities (even "best practices) in too many different ways to identify common process measures; Data availability and grantee capacity a consideration in selecting "low hanging fruit" (easiest); Data collection infrequent for surveys (every 2-5 years) which may limit use of performance data to evaluation rather than performance management/monitoring. |
| Issues of case context | Grantees have different focus on the four goal areas sometimes choice is influenced by political context; Although best practices & |

| | |
|---|---|
| (situationality) | community guide provide extensive information on evidence-based program activities for tobacco control, implementation happens in many different ways; Costly to conduct/participate in all the varied surveys (tobacco supplements, etc) so not every grantee will have data to calculate core measures. |
| Other supportive case examples | All cases support this issue of the effects of network variability on performance measurement |
| Negative case example | N/A |

| Jennings & Haist (2004) Hypothesis: "Measurement will be more common and will have greater impact when agencies have greater control over outcomes" (p185) | Research Question #3: How does networked public management affect CDC's control over outcomes and the subsequent design and perceived impact of performance measurement? |
|---|---|
| **Finding #1: Dependency on the PHEP program's network partners diminishes CDC and grantee control over performance.** | |
| Effect on performance measurement system | Process measures selected that are more closely tied to work of grantees and for which grantees feel they have greater control. |
| Broader implication for performance measurement | Issue of control effects types and choice of measures. |
| Case examples | PHEP has applied a criteria for selection of new performance measures titled, "under control of public health"; Concerns about "fairness"; Grantees reject measures viewed as too far outside their control. |
| Issues of case context (situationality) | PHEP outcomes are not yet well defined; Outcomes for preparedness are jointly produced; Goal conflicts with horizontal partners from other sectors. |
| Other supportive case examples | CSPS and NTCP – all involve dependencies on vertical and HORIZONTAL partners |
| Negative case example | NBCCEDP has greater dependence on vertical partners with whom some authority can be exerted through policy tools, network relationships, etc. |
| | |
| **Finding #2: Dependencies and goal conflicts with CSPS local-level, horizontal network partners compromises grantees' control over performance.** | |
| Effect on performance measurement system | Selection of process and immediate outcome measures and venue-specific measures that grantees feel they have more control over. |
| Broader implication for performance measurement | Problematic to include intermediate or long-term outcome measures that are jointly produced; Data availability and quality problematic when relying on unfunded, informal partners; Hard to discern accountability for longer term outcomes. |
| Case examples | Grantees oppose performance measures viewed as too far outside their control; Priority and goal conflicts with vertical and horizontal partners public health vs. public safety (corrections) – makes it difficult to identify incentives for unfunded, horizontal partners with other priorities; No formal authority over jails, physicians outside the funded STD clinics and family planning sites, juvenile detention; Dependent on jails, physicians, and others to participate in STD screening and treatment AND to provide data for performance measures; Grantees reject population-based measures other than for syphilis (2009 proposed measures); DSTDP think GPRA measures at population-level are unrealistic. |

| Issues of case context (situationality) | DSTDP advocating "community perspective" and greater focus on population effects, but recognizes that their limited resources are inadequate to have population effects on their own; Emphasis on engaging network partners like private providers, jails, juvenile centers in order to access high risk populations; Changes in broader health care access affecting program goals; Stubborn institutional culture resistant to change; Grant as funding mechanism does not support much control within vertical network and DSTDP has not demanded much accountability in past; Inadequate resources to have population-level effects on disease prevalence and incidence. |
|---|---|
| Other supportive case examples | PHEP, NTCP |
| Negative case example | NBCEDP has less dependencies given greater reliance on vertical partners. |

| | |
|---|---|

**Finding #3: Although the NBCCEDP network compromises control over program implementation, DCPC has designed its performance measurement system in ways that support CDC's and grantees' control over performances.**

| Effect on performance measurement system | DCPC designed the program in ways that enhance control: Narrow focus on women served, rather than broader, population-level focus; Selected performance measures aligned with program priorities and consistent with service delivery implementation that grantees can control process, short-term, and intermediate performance measures related to service delivery; DCPC has made revisions in measures to strengthen grantee's control over measured performance; Allows multiple uses of performance measurement data for accountability, program improvement, and budgeting. |
|---|---|
| Broader implication for performance measurement | Network consensus on program goals facilitates selection of performance measures; Reliance on vertical network with less dependence on horizontal increases control over performance and accountability is less fragmented; Grantees view performance measurement as meaningful, fair, valid, and relevant to priorities allowing data to be used for multiple purposes, including budgeting. |
| Case examples | Revised calculation of measure (cervical measure related to time to diagnosis) to assure grantee control; Performance measures tied to work of providers in vertical network; policy tools used strengthen control within vertical network. |
| Issues of case context (situationality) | Narrow focus on women served through NBCCEDP; Service delivery dependent primarily on vertical network; Strong consensus within network on program goals; Control over vertical network through policy tools, management practice, and network relationships; 22,000 local providers in system. |
| Other supportive case examples | Less grantee resistance to measures for which they have greater control – in CSPS, more support for measures that are dependent on work of vertical, funded partners; in PHEP, greater support for measures that grantees directly affect. |
| Negative case example | NTCP has made decision to include outcomes that are jointly produced – performance is dependent on comprehensive activities implemented by multiple partners and at multiple levels. |

| | |
|---|---|

| **Finding #4: Policy tools, management practice, and network partnerships enhance DCPC and grantee control over the NBCCEDP performance measures.** | |
|---|---|
| Effect on performance measurement system | DCPC and grantees feel greater control over performance on measures Reporting of performance measures is a requirement of cooperative agreement; Performance measures is included as a requirement in grantee contracts with providers (not clear how many grantee use performance-based contracts). |
| Broader implication for performance measurement | Policy tools, management practices, and partner relationships can strengthen authority within vertical network to give grantees more control over performance. |
| Case examples | CDC's use of MDE monitoring data to enhance control over local level through on-going monitoring of extensive program data (service delivery data); Cooperative agreement requirements for MDE reporting; States use of performance-based contracts or other reimbursement policies to influence performance; CDC's use of performance measures for budgeting increases importance of measures to grantees; Grantees' relationships with providers supports control as it facilitates greater influence with these partners. |
| Issues of case context (situationality) | Mature program with long-standing relationships with grantees and with local level providers; Some grantees willing to apply performance-based contracts using the NBCCEDP performance measures; Extensive and sophisticated data management/monitoring system (MDEs); Data-driven organizational culture endorsed at all levels of CDC management; Grantee buy-in on importance and value of performance measures for assuring accountability with women served in program. |
| Other supportive case examples | Less evidence across other cases in use of policy tools, etc. to strengthen control. |
| Negative case example | CSPS and NTCP do not exert much authority through their policy tools (grant, cooperative agreement). |

# APPENDIX N

## SUMMARY OF INDIVIDUAL CASE FINDINGS

**Summary of Case Findings**

**Public Health Preparedness Program (PHEP)**

1.  Dependency on the PHEP program's network partners diminishes CDC and grantee control over performance.

2.  Network public management fragments the PHEP program's accountability for results, creating challenges for performance measurement.

3.  In the case of the PHEP program, performance measurement is a "political, social, and scientific" process.

4.  The variability across the PHEP program's vertical network significantly shapes the design of its performance measurement system.

**Comprehensive STD Prevention Systems (CSPS)**

5.  Dependencies and goal conflicts with CSPS' local-level, horizontal network partners compromises grantees' control over performance.

6.  Given the CSPS network context, performance measurement is a negotiated and incremental process.

7.  Extensive variability across grantees influences the design of the CSPS' performance measurement system.

**National Breast and Cervical Cancer Early Detection Program (NBCCEDP)**

8.  Although the NBCCEDP network compromises control over program implementation, DCPC has designed its performance measurement system in ways that support CDC's and grantees' control over performances.

9.  Policy tools, management practice, and network partnerships enhance CDC and grantee control over the NBCCEDP performance measures.

10. The NBCCEDP network requires that DCPC make significant resource investments to build a comprehensive performance management system in order to ensure data quality and the use of performance measurement data at multiple levels.

**National Tobacco Control Program (NTCP)**

1. Network consensus on goals, a strong evidence base, and extensive survey data facilitate identification of outcome measures for comprehensive tobacco control.

2. The network implementation structure of the NTCP leads to a joint production of outcomes and shared accountability.

3. Network variability limits types and choice of NTCP core measures and their use.

# REFERENCES

Abramson, Mark A., Jonathan D. Breul, and John M. Kamensky. 2006. *Six trends transforming government*. Washington D.C.: IBM Center for The Business of Government.

Affholter, D. P. 1994. Outcome monitoring. In *Handbook of practical program evaluation*, ed. K. Newcomer: 96-118. San Francisco: Jossey-Bass.

Agranoff, Robert. 2003. *Leveraging networks: A guide for public managers working across organizations*. Washington D.C.: IBM Endowment for The Business of Government.

_____. 2005. Managing collaborative performance: Changing the boundaries of the state? *Public Performance and Management Review* 29, no. 1: 18-45.

Agranoff, Robert. 2005. Managing collaborative performance: Changing the boundaries of the state? *Public Performance and Management Review* 29, no. 1: 18-45.

Agranoff, Robert and Michael McGuire. 2001a. American federalism and the search for models of management. *Public Administration Review* 61, no. 6: 671-681.

_____. 2001b. Big questions in public network management research. *Journal of Public Administration Research and Theory* 11, no. 3: 295-326.

Agranoff, Robert and Beryl Radin. 1991. The comparative case study approach in public administration. *Research in Public Administration* 1: 203-231.

Allison, Graham. 1971. *Essence of decision: Explaining the Cuban missile crisis*. New York: Longman.

Anastas, Jeane W. 2004. Quality in qualitative evaluation: Issues and possible answers. *Research on Social Work Practice* 14, no. 1: 57-65.

Barberis, Peter. 1998. The new public management and a new accountability. *Public Administration* 76, no. Autumn: 451-470.

Bardach, Eugene. 1977. *The implementation game*. Cambridge, Mass: MIT Press.

Barry, Michael A. 2000. How can performance standards enhance accountability for public health? *Journal of Public Health Management and Practice* 6, no. 5: 78-84.

Behn, Robert D. 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* 63, no. 5: 586-606.

Berman, Evan and XiaoHu Wang. 2000. Performance measurement in U.S. Counties: Capacity for reform. *Public Administration Review* 60, no. 5: 409-420.

Berman, P. 1978. The study of macro- and micro-implementation. *Public Policy* 26, no. 2: 155-184.

Blalock, Ann B. and Burt S. Barnow. 2001. Is the new obsession with performance management masking the truth about social programs? In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe:485-517. Albany, NY: Rockefeller Institute Press.

Bogdan, Robert C. and Sari K. Biklen. 2007. *Qualitative research for education: An introduction to theories and methods*. Boston: Allyn and Bacon.

Bouckaert, Geert and B. Guy Peters. 2002. Performance measurement and management: The Achilles heel in administrative modernization. *Public Performance and Management Review* 25, no. 4: 359-362.

Boyte, Harry C. 2005. Reframing democracy: Governance, civic agency, and politics. *Public Administration Review* 65, no. 5: 536-546.

Brass, Clinton T. 2004. *The bush administration's program assessment rating tool (PART)*. Washington D.C.: Congressional Research Service Library of Congress, RL32663.

Broom, C., S. Caudle, E. T. Jennings, and K. Newcomer. 2002. Meeting the challenges of performance oriented government. In *Meeting the challenges of performance oriented government*, ed. K. Newcomer, E.T. Jennings, C. Broom and A. Lomax. Washington, DC: American Society for Public Administration/Center for Accountability and Performance.

Brower, Ralph S., Mitchel Y. Abolafia, and Jered B. Carr. 2000. On improving qualitative methods in public administration research. *Administration and Society* 32, no. 4: 363-397.

Cavalluzzo, Ken S. and Christopher D. Ittner. 2004. Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society* 29: 243-267.

Centers for Disease Control and Prevention. 2005. *Key outcomes indicators for evaluating comprehensive tobacco control programs*. Atlanta, Georgia: U.S. Department of Health and Human Services.

Charmaz, Kathy. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks: Sage.

Coplin, William D., Astrid E. Merget, and Carolyn Bourdeaux. 2002. The professional researcher as change agent in the government-performance movement. *Public Administration Review* 62, no. 6: 699-711.

Creswell, John W. 2007. *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks: Sage.

Creswell, John W. and Dana L. Miller. 2000. Determining validity in qualitative inquiry. *Theory into Practice* 39, no. 3: 124-130.

Crotty, Michael. 2003. *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks: Sage.

Dahlgren, G. and M. Whitehead. 1991. *Policies and strategies to promote social equity in health*. Stockholm: Institute for the Futures Studies.

de Leon, Peter and Linda de Leon. 2002. What ever happened to policy implementation? An alternative approach. *Journal of Public Administration Research and Theory* 4: 467-492.

Denzin, Norman K. and Yvonna S. Lincoln. 2000. Introduction: The discipline and practice of qualitative research. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln:1-28. Thousand Oaks: Sage.

Derose, Stephen F., Steven M. Asch, Jonathan E. Fielding, and Mark A. Schuster. 2003. Developing quality indicators for local health departments: Experience in Los Angeles county. *American Journal of Preventive Medicine* 25, no. 4: 347-357.

Derthick, Martha. 1972. *New towns in town: Why a federal program failed*. Washington D.C.: Urban Institute.

Dixit, Avinash. 2002. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources* 37: 697-727.

Dubnick, Melvin. 2005. Accountability and the promise of performance. *Public Performance and Management Review* 28, no. 3: 376-417.

Elmore, R. 1978. Organizational models of social program implementation. *Public Policy* 26, no. 2: 185-228.

Ezzy, Douglas. 2002. *Qualitative analysis: Practice and innovation*. London: Routledge.

Fawcett, S. B. 1999. Some lessons on community organization and change. In *Reflections on community organization: Enduring themes and critical issues*, ed. J. Rothman. Itasca, IL: F. E. Peacock.

Feller, Irwin. 2002. Performance measurement redux. *American Journal of Evaluation* 23, no. 4: 435-452.

Forsythe, Dall W. 2001. Pitfalls in designing and implementing performance management systems. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe:519-551. Albany, NY: Rockefeller Institute Press.

Fossett, James, Thomas Gais, and Frank J. Thompson. 2001. Federalism and performance measurement: Health insurance, food stamps, and the take-up challenge. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe. Albany, NY: Rockefeller Institute Press.

Frederickson, David G. 2001. *The potential of the government performance and results act as a tool to manage third-party government*. Washington D.C.: The PricewaterhouseCoopers Endowment for the Business of Government.

_____. 2003. Performance measurement and third party government: A study of the implementation of the government performance and results act in five health agencies. Dissertation, Indiana University.

Frederickson, David G. and H. Frederickson, George. 2006. *Measuring the performance of the hollow state*. Washington D.C.: Georgetown University Press.

Frederickson, H., George and Kevin B. Smith. 2003. *The public administration theory primer: Essentials of public policy and administration*. Boulder: Westview Press.

Geertz, Clifford. 1973. *The interpretation of cultures: Selected essays*. New York: Basic Books.

Gilmour, John B. 2006. *Implementing OMB's program assessment rating tool (PART): Meeting the challenges of integrating budget and performance*. Washington D.C.: IBM Center for The Business of Government

Glaser, Barney G. and Anselm L. Stauss. 1967. *Discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glesne, Corrine and Alan Peshkin. 1992. *Becoming qualitative researchers*. White Plains, N.Y.: Longman.

Goddard, Maria and Russell Mannion. 2004. The role of horizontal and vertical approaches to performance measurement and improvement in the U.K. Public sector. *Public Performance and Management Review* 28, no. 1: 75-95.

Goerdel, Holly T. 2006. Taking initiative: Proactive management and organizational performance in networked environments. *Journal of Public Administration Research and Theory* 16, no. 3: 351-367.

Golden-Biddle, Karen and Karen Locke. 1993. Appealing work: An investigation of how ethnographic texts convince. *Organizational Science* 4, no. 4: 595-615.

Goldsmith, Stephen and William D. Eggers. 2004. *Governing by network*. Washington D.C.: Brookings Institute Press.

Gormley, William T., Jr. and David L. Weimer. 1999. *Organizational report cards*. Cambridge, Mass: Harvard University Press.

Government Performance and Results Act of 1993. Public Law 103-162.

Grizzle, Gloria A. 2002. Performance measurement and dysfunction: The dark side of quantifying work. *Public Performance and Management Review* 25, no. 4: 363-369.

Handler, A., M. Issel, and B. Turnock. 2001. A conceptual framework to measure performance of the public health system. *American Journal of Public Health* 91, no. 8: 1235-1239.

Harkreader, Steve A. and Gary T. Henry. 2000. Using performance measurement systems for assessing the merit and worth of reforms. *American Journal of Evaluation* 21, no. 2: 151-170.

Harmon, Michael M. and Richard T. Mayer. 1986. *Organization theory for public administration*. Burke, VA: Chatelaine Press, Inc.

Hatry, Harry P. 1997. Where the rubber meets the road: Performance measurement for state and local public agencies. *New Directions for Evaluation* 75, no. Fall: 31-44.

_____. 1999. *Performance measurement: Getting results.* Washington D.C.: Urban Institute Press.

_____. 2001. What types of performance information should be tracked? In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe. Albany, NY: Rockefeller Institute Press.

_____. 2002. Performance measurement: Fashions and fallacies. *Public Performance and Management Review* 25, no. 4: 352-358.

Hatry, Harry P., Elaine Morley, Shelli B. Rossman, and Joseph S. Wholey. 2004. *How federal programs use outcome information: Opportunities for federal managers*. IBM Center for the Business of Government.

Heclo, Hugh. 1978. Issue networks and the executive establishment. In *The new American political system*, ed. Anthony King:87-124. Washington D.C.: American Enterprise Institute.

Heinrich, Carolyn J. 1999. Do government bureaucrats make effective use of performance management information? *Journal of Public Administration Research and Theory* 9, no. 3: 363-393.

_____. 2002. Outcomes-based performance management in the public sector: Implications for government accountability and effectiveness. *Public Administration Review* 62, no. 6: 712-725.

Heinrich, Carolyn J., Carolyn J. Hill, and Laurence E. Lynn, Jr. 2004. Governance as an organizing theme for empirical research. In *The art of governance: Analyzing management and administration*, ed. Patricia A. Ingraham and Laurence E. Lynn, Jr.:3-19. Washington D.C.: Georgetown University Press.

Heinrich, Carolyn J. and Laurence E. Jr. Lynn, eds. 2000. *Governance and performance: New perspectives*. Washington D.C.: Georgetown University Press.

Hendricks, Michael. 2002. Outcome measurement in the nonprofit sector: Recent developments, incentives, and challenges. In *Meeting the challenges of performance oriented government*, ed. Kathryn E. Newcomer, Edward T. Jennings, Cheryle Broom and Allen Lomax. Washington D.C.: Center for Accountability and Performance of the American Society for Public Administration.

Hill, Carolyn J. and Laurence E. Jr. Lynn. 2005. Is hierarchical governance in decline? Evidence from empirical research. *Journal of Public Administration Research and Theory* 15, no. 2: 173-195.

Hood, Christopher and B. Guy Peters. 2004. The middle-aging of new public management: Into the age of paradox? *Journal of Public Administration Research and Theory* 14, no. 3: 267-282.

Hsieh, Hsiu-Fang and Sarah E. Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research* 15, no. 9: 1277-1288.

Ingraham, Patricia A. and Laurence E. Lynn, Jr., eds. 2004. *The art of governance: Analyzing management and administration*. Washington D.C.: Georgetown University Press.

Institute of Medicine of the National Academies. 1988. *The future of public health*. Washington D.C.: The National Academies Press.

_____. 2003. *The future of the public's health in the 21st century*. Washington D.C.: The National Academies Press.

Janesick, Valerie J. 2000. The choreography of qualitative research design: Minuets, improvisations, and crystallization. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln, 2nd:379-400. Thousand Oaks: Sage.

Jennings, Edward T. and Meg Patrick Haist. 2002. Does performance measurement perform? Paper for presentation, College Station, Texas.

Jennings, Edward T., Jr. and Meg Patrick Haist. 2004. Putting performance measurement in context. In *The art of governance: Analyzing management and administration*, ed. Patricia A. Ingraham and Laurence E. Lynn, Jr.:173-194. Washington D.C.: Georgetown University Press.

Julnes, Patria de Lancer and Marc Holzer. 2001. Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review* 61, no. 6: 693-708.

Kates, Jennifer, Katherine Marconi, and Thomas Jr. Mannle. 2001. Developing a performance management system for a federal public health program: The Ryan White Care Act titles i and ii. *Evaluation and Program Planning* 24: 145-155.

Keast, Robyn, Myrna Mandell, Kerry Brown, and Geoffrey Woolcock. 2004. Network structures: Working differently and changing expectations. *Public Administration Review* 64, no. 3: 363-371.

Kelly, Janet M. 2002. Why we should take performance measurement on faith (facts being hard to come by and not terribly important). *Public Performance and Management Review* 25, no. 4: 375-380.

Kettl, Donald F. 2002. *The transformation of governance: Public administration for twenty-first century America*. Baltimore: John Hopkins University Press.

_____. 2003. *The transformation of governance: Public administration for twenty-first century America*. Baltimore, MD: John Hopkins Press.

_____. 2005. *The next government of the united states: Challenges for performance in the 21st century*. Washington D.C.: IBM Center for The Business of Government.

Kravchuck, R. S. and R. W. Shack. 1996. Designing effective performance-measurement systems under the government performance and results act of 1993. *Public Administration Review* 56, no. 4: 348-358.

Krueger, Richard A. 1988. *Focus groups: A practical guide for applied research*. Newbury Park, CA: Sage.

Landrum, L. B. and S. L. Baker. 2004. Managing complex systems: Performance management in public health. *Journal of Public Health Management and Practice* 10, no. 1: 13-18.

Lasker, Rod D. and Elisa S. Weiss. 2003. Broadening participation in community problem solving: A multidisciplinary model to support collaborative practice and research. *Journal of Urban Health* 80: 14-47.

Lasker, Rod D., Elisa S. Weiss, and Rebecca Miller. 2001. Partnership synergy: A practical framework for studying and strengthening the collaborative advantage. *The Milbank Quarterly* 79, no. 2: 179-205.

Leonhardt, David. 2006. Rank colleges, but rank them right. *New York Times*, August 16, 2006, 1.

Lester, J. P. and M. L. Goggin. 1998. Back to the future: The rediscovery of implementation studies. *Policy Currents* 8, no. 3: 1-9.

Liamputtong, Pranee and Douglas Ezzy. 2005. *Qualitative research methods*. New York: Oxford University Press.

Light, Paul C. 1997. *The tides of reform: Making government work 1945-1995*. New Haven: Yale University Press.

Lincoln, Yvonna S. and Egon G. Guba. 1985. *Naturalistic inquiry*. Newbury Park: Sage.

Lofland, John and Lyn H. Lofland. 1995. *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont: Wadsworth.

Lynn, Laurence E. Jr., Carolyn J. Heinrich, and Carolyn J. Hill. 2000. Studying governance and public management: Challenges and prospects. *Journal of Public Administration Research and Theory* 10, no. 2: 233-262.

MacQueen, Kathleen M., Eleanor McLellan-Lemal, Kelly Bartholow, and Bobby Milstein. 2008. Team-based codebook development: Structure, process, and agreement. In *Handbook for Team-Based Qualitative Research*, ed. Greg Guest and Kathleen M. MacQueen. New York: Altamira Press.

Majone, G. and A. Wildavsky. 1995. Implementation as evolution. In *Public policy: The essential readings*, ed. S. Z. Theodoulou and M. A. Cahn. Englewood Cliffs, NY: Prentice Hall.

Managing for results: Using GPRA to assist congressional and executive branch decision-making. 1997. In *Committee on Government Reform and Oversight*. Washington D.C.: U.S. Government Accounting Office.

Mandell, Myrna and Robyn Keast. 2006. Evaluating network arrangements: Toward revised performance measures, Leuven, Belgium.

_____. 2006. A performing public sector: The second transatlantic dialogue, Leuven, Belgium.

Mark, Melvin M., Gary T. Henry, and G. Julnes. 2000. *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs.* San Francisco: Jossey-Bass.

Mathison, Sandra. 1988. Why triangulate? *Educational Researcher* 17, no. 2: 13-17.

Mauthner, Natasha S. and Andrea Ddoucet. 2003. Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology* 37, no. 3: 413-431.

Maxwell, J. A. 1992. Understanding and validity in qualitative research. *Harvard Educational Review* 62: 279-300.

McGuire, Michael. 2006. Intergovernmental management: A view from the bottom. *Public Administration Review* 66, no. 5: 677-679.

Medicine, Institute of. 1997. *Improving health in the community: A role for performance monitoring*. Edited by Institute of Medicine. Washington, D.C.: National Academy Press.

Melkers, Julia and Katherine Willoughby. 1998. The state of the states: Performance-based budgeting requirements in 47 out of 50. *Public Administration Review* 58, no. 1: 66-73.

_____. 2004. *Staying the course: The use of performance measurement in state governments*. Washington D.C.: IBM Center for The Business of Government.

_____. 2005. Models of performance measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public Administration Review* 65, no. 2: 180-190.

Merriam, Sharan B. 1995. What can you tell from an n of 1?: Issues of validity and reliability in qualitative research. *PAACE Journal of Lifelong Learning* 4: 51-60.

_____. 1998. *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.

_____. 2009. *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass.

Mihm, J. Christopher. 2000. Managing for results: Continuing challenges to effective GPRA implementation. In *Committee on Government Reform*, ed. United States General Accounting Office. Washington D.C.: United States General Accounting Office.

Mihm, J. Christopher. 2001. Implementing GPRA: Progress and challenges. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe. Albany, NY: Rockefeller Institute Press.

Miles, Matthew B. and A. Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.

Milward, H. Brinton and Keith G. Provan. 1998. Measuring network structure. *Public Administration* 76, no. Summer: 387-407.

_____. 2000. Governing the hollow state. *Journal of Public Administration Research and Theory* 10, no. 2: 359-379.

_____. 2004. The public manager's guide to network management. In *Annual Meeting of the American Political Science Association*,. Chicago, Illinois.

_____. 2006. *A manager's guide to choosing and using collaborative networks*. Washington D.C.

Morgan, David L. 2007. Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research* 1, no. 1: 48-76.

Moynihan, D. P. and P. W. Ingraham. 2003. Look for the silver lining: When performance-based accountability systems work. *Journal of Public Administration Research and Theory* 13, no. 4: 469-490.

Murphy, Dara.L., Thomas Chapel, and Cynthia Clark. 2004. Moving diabetes care from science to practice: The evolution of the national diabetes prevention and control program. *Annals of Internal Medicine* 140, no. 11: 978-985.

Nathan, Richard P. 2001. Introduction. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe:3-15. Albany, NY: Rockefeller Institute Press.

Newcomer, Kathryn E. 1997. Using performance measurement to improve programs. *New Directions for Evaluation* 75, no. Fall: 5-14.

Osborne, David and Ted Gaebler. 1992. *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. New York: Penguin Books.

O'Toole, Laurence J., Jr. 1997. Treating networks seriously: Practical and research-based agendas in public administration. *Public Administration Review* 57, no. 1: 45-52.

O'Toole, Laurence J., Jr. and Kenneth Meier. 2004. Public management in intergovernmental networks: Matching structural networks and managerial networking. *Journal of Public Administration Research and Theory* 23, no. 1: 31-47.

Page, Stephen. 2004. Measuring accountability for results in interagency collaboratives. *Public Administration Review* 64, no. 5: 591-606.

Parsons, Wayne. 1995. *Public policy: An introduction to the theory and practice of policy analysis*. Northampton, MA: Edward Elgar.

Patton, Michael Quinn. 1999. Enhancing the quality and credibility of qualitative analysis. *Health Services Research* 34, no. 5 Part II: 1189-1208.

_____. 2002. *Qualitative research and evaluation methods*. Thousand Oaks: Sage.

Perrin, Burt. 1998. Effective use and misuse of performance measurement. *American Journal of Evaluation* 19, no. 3: 367-379.

_____. 2006. *Moving from outputs to outcomes: Practical advice from governments around the world*. The IBM Center for the Business of Government.

Perrin, Edward B. 2002. Some thoughts on outcomes research, quality improvement, and performance measurement. *Medical Care* 40, no. 6: III-89-III-91.

Perrin, Edward B., Jane S. Durch, and Susan M. Skillman, eds. 1999. *Health performance measurement in the public sector: Principles and policies for implementing an information network*. Edited by National Research Council. Washington, D.C.: National Academy Press.

Perrin, Edward B. and Jeffrey J. Koshel, eds. 1997. *Assessment of performance measures for public health, substance abuse, and mental health.* Edited by National Research Council. Washington, D.C.: National Academy Press.

Peshkin, Alan. 1993. The goodness of qualitative research. *Educational Researcher* 22, no. 2: 23-29.

_____. 1998. In search of subjectivity one's own. *Educational Researcher* 17, no. 7: 17-22.
_____. 2000. The nature of interpretation in qualitative research. *Educational Researcher* 29, no. 2: 5-9.

Peters, B. Guy. 2001. *The future of governing.* Lawrence: University Press of Kansas.

Peters, B. Guy and Jon Pierre. 1998. Governance without government? Rethinking public administration. *Journal of Public Administration Research and Theory* 8, no. 2: 223-243.

Pierre, Jon and B. Guy Peters. 2005. *Governing complex societies: Trajectories and scenarios.* New York: Palgrave MacMillan.

Poister, Theodore H. 2003. *Measuring performance in public and nonprofit organizations.* San Francisco: Jossey-Bass.

Poister, Theodore H. and Gregory D. Streib. 1999a. Performance measurement in municipal government: Assessing the state of the practice. *Public Administration Review* 59, no. 4: 325-335.

_____. 2005. Elements of strategic planning and management in municipal government: Status after two decades. *Public Administration Review* 65, no. 1: 45-56.

Pollitt, C. 1986. Beyond the managerial model: The case for broadening performance assessment in government and public services. *Financial Accountability and Management* 2, no. 3: 155-170.

Posner, Paul L. 2002. Accountability challenges of third-party government. In *The tools of government: A guide to new governance*, ed. Lester M. Salamon:552-564. New York: Oxford University Press.

Pressman, J. and A. Wildavsky. 1973. *Implementation.* Berkeley, CA: University of California Press.

Radin, Beryl A. 1998. The government performance and results act (GPRA): Hydra-headed monster or flexible management tool? *Public Administration Review* 58, no. 4: 307.

_____. 1998. Searching for government performance: The government performance and results act. *PS: Political Science & Politics* 31: 553-555.

_____. 2000a. The government performance and results act and the tradition of federal management reform: Square pegs in round holes? *Journal of Public Administration Research and Theory* 10, no. 1: 111-135.

_____. 2000b. Intergovernmental relationships and the federal performance movement. *Publius* 30, no. 12: 133-158.

_____. 2001. Intergovernmental relationships and the federal performance movement. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe. Albany, NY: Rockefeller Institute Press.

_____. 2003. A comparative approach to performance management: Contrasting the experience of Australia, New Zealand, and the United states. *International Journal of Public Administration* 26, no. 1: 1355-1376.

_____. 2006. *Challenging the performance movement*. Washington D.C.: Georgetown University Press.

Rapley, Timothy John. 2001. The art(fulness) of open-ended interviewing: Some considerations on analyzing interviews. *Qualitative Research* 1, no. 3: 303-323.

Richardson, Laurel. 2000. Writing: A method of inquiry. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln:923-948. Thousand Oaks: Sage.

Roper, W. L. and G. P. Mays. 2000. Performance measurement in public health: Conceptual and methodological issues in building the science base. *Journal of Public Health Management and Practice* 6, no. 5: 66-77.

Rossi, P. H., H. E. Freeman, and M. W. Lipsey. 1999. *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.

Rossman, Gretchen B. and Sharon F. Rallis. 2003. *Learning in the field: An introduction to qualitative research*. Thousand Oaks: Sage.

Salamon, Lester M. 2002. The new governance and the tools of public action: An introduction. In *The tools of government: A guide to the new governance*, ed. Lester Salamon, M.:1-47. New York: Oxford University Press.

Satterfield, D. W., D. L. Murphy, J. D. Essien, G. Hosey, M. Stankus, P. Hoffman, K. Beartusk, P. L. Mitchell, and A. Alfaro-Correa. 2004. Using the essential public

health services as strategic leverage to strengthen the public health response to diabetes. *Public Health Reports* 119: 311-321.

Scheirer, M. A. 1981. *Program implementation: The organizational context*. Beverly Hills, CA: Sage Publications.

Schick, Allen. 2001. Getting performance measures to measure up. In *Quicker better cheaper? Managing performance in American government*, ed. Dall W. Forsythe: 39-60. Albany, NY: Rockefeller Institute Press.

Schwandt, Thomas A. 2000. Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln:189-214. Thousand Oaks: Sage.

_____. 2001. *Dictionary of qualitative inquiry*. Thousand Oaks: Sage.

Shortell, S. M. 2002. Evaluating partnerships for community health improvement: Tracking the footprints. *Journal of Health Politics, Policy, and Law* 27, no. 1.

Smith, P. C. and M. Goddard. 2002. Performance management and operational research: A marriage made in heaven? *Journal of Operational Research Society* 53: 247-255.

Stake, Robert E. 1995. *The art of case study research*. Thousand Oaks: Sage.

_____. 2000. Case studies. In *Handbook of qualitative research*, ed. Norman K. Denzin and Yvonna S. Lincoln:435-454. Thousand Oaks: Sage.

_____. 2006. *Multiple case study analysis*. New York: Guilford Press.

Stewart, David W., Prem N. Shamdasani, and Dennis W. Rook. 2007. *Focus groups: Theory and practice*. Thousand Oaks: Sage.

Stoker, Gary. 1998. Governance as theory: Five propositions. *International Social Science Journal* 50, no. 1: 17-28.

Stone, Melissa M. and Ostrower Francie. 2007. Acting in the public interest? Another look at research on nonprofit governance. *Nonprofit and Voluntary Sector Quarterly* In press.

Straw, R. B. 1996. Evaluation of public health service programs under the government performance and results act. *Evaluation and the Health Professions* 19, no. 3: 394-404.

Task Force on Community Preventive Services. 2005. *The guide to community preventive services: What works to promote health?* Edited by Kate W. Harris. New York: Oxford University Press.

Turnock, B.J. 2000. Can public health performance standards improve the quality of public health practice? *Journal of Public Health Management and Practice* 6, no. 5: 19-25.

U.S. General Accounting Office. 1996. *Executive guide: Effectively implementing the Government Performance and Results Act. Washington D.C.: U.S. General Accounting Office, GAO/GGD-96-188.*

_____. 1997a. *Managing for results: Analytic challenges in measuring performance.* Washington D.C.: U.S. General Accounting Office, GAO/HEHS/GGD-97-138.

_____. 1997b. *Managing for results: Prospects for effective implementation of GPRA.* Washington D.C.: U.S. General Accounting Office, GAO/AIMD-97-113.

_____. 1997c. *Managing for results: Using GPRA to assist Congressional and Executive branch decision-making.* In Committee on Government Reform and Oversight. Washington D.C.: U.S. General Accounting Office.

_____. 1998a. *Managing for results: Measuring program results that are under limited federal control.* Washington D.C.: U.S. General Accounting Office, GAO/GGD-99-16.

_____. 1999. *Managing for results: Opportunities for continued improvements in agencies' performance plans.* Washington D.C.: U.S. General Accounting Office, GAO/AIMD-99-215.

_____. 2000. *Continuing challenges to effective GPRA implementation.* In Committee on Government Reform. Washington D.C.: U.S. General Accounting Office.

_____. 2004. *Results-oriented government: GPRA has established a solid foundation for achieving greater results.* Washington D.C.: U.S. General Accounting Office, GAO-04-38.

U.S. Government Accountability Office. 2005. *OMB's PART reviews increased agencies' attention to improving evidence of program results.* Washington D.C.: U.S. Government Accountability Office, GAO-06-67.

Ulin, Priscilla R., Elizabeth T. Robinson, and Elizabeth E. Tolley. 2005. *Qualitative methods in public health: A field guide for applied research.* San Francisco: Jossey-Bass.

Voets, Joris, Filip De Rynck, and Wouter Van Dooren. 2006. Managing and measuring collaborative performance, Leuven, Belgium.

Wandersman, A., R. M. Goodman, and F. Butterfoss. 1997. Understanding coalitions and how they operate. In *Community organizing and community building for health*, ed. M. Minkler. New Brunswick, NJ: Rutgers University Press.

Wholey, Joseph S. 1999. Performance based management: Responding to the challenges. *Public Productivity and Management Review* 22, no. 3: 288-307.

_____. 2002. Making results count in public and nonprofit organizations: Balancing performance with other values. In *Meeting the challenges of performance oriented government*, ed. Kathryn Newcomer, Edward T. Jennings, Cheryle Broom and Allen Lomax:13-36. Washington, D.C.: American Society for Public Administration, Center for Accountability and Performance.

_____. 2004. Evaluability assessment. In *Handbook of Practical Program Evaluation, 2ⁿᵈ edition*, ed. J.S. Wholey, H.P. Hatry, K.E. Newcomer. San Francisco: John Wiley & Sons, Inc.

Wholey, Joseph S. and Harry P. Hatry. 1992. The case for performance monitoring. *Public Administration Review* 52, no. 6: 604-610.

Wholey, Joseph S. and Kathryn E. Newcomer. 1997. Clarifying goals, reporting results. *New Directions for Evaluation* 75, no. Fall: 91-98.

Wilson, James Q. 1989. *Bureaucracy: What government agencies do and why they do it*: Basic Books.

Yin, Robert K. 2009. *Case study research: Design and methods, 4ᵀᴴ Edition*. Applied social research methods series, vol 5. Thousand Oaks: Sage.