Attachment 2:

Drug Free Communities Support Program National Evaluation Plan

Drug Free Communities Support Program National Evaluation Plan

Final
December 27, 2010

Report prepared for the:

White House Office of National Drug Control Policy
Drug Free Communities Support Program

by:

ICF International
9300 Lee Highway
Fairfax, VA 22031
www.icfi.com

under contract number BPD-NDC-09-C1-0003

Table of Contents

Drug Free Communities Support Program
National Evaluation Plan

Final: December 27, 2010

The analysis plan outlined in this document is designed to provide ONDCP with strong evidence and useful results tailored to the needs of various stakeholder groups (i.e., SAMHSA, DFC grantees, community partners, etc.). Our approach will ensure that we not only continue to strengthen ONDCP's Government Performance Results Act (GPRA) and Program Assessment Rating Tool (PART) reports, but also continue to provide results that can be used by coalitions to enhance their operations and capacity, and ultimately, improve their performance in reducing community-level youth substance use rates.

The scope of the evaluation described in this attachment is specific to Drug Free Communities (DFC). This analysis plan is not intended to be simply the product of our initial planning efforts; rather, it will become a "living, breathing document" which will be used as a point of reference throughout the five-year evaluation contract. The maintenance of this plan will ensure that all major decisions concerning the analysis are stored in a single location. It will also ensure that new staff on the contract will quickly overcome any learning curve and become fully engaged in this effort as soon as possible.

1. Introduction to the Drug Free Communities (DFC) Program

The Federal government launched a major effort to prevent youth drug use by appropriating funds in 1997 for the Drug-Free Communities Act. That financial commitment has continued for more than a decade, and in Fiscal Year 2009, nearly 726 community coalitions across 50 States, the District of Columbia, the U.S. Virgin Islands, American Samoa, Palau, and Puerto Rico received grants to improve their substance abuse prevention strategies. With bipartisan support from Congress, the DFC Support Program provides community coalitions with up to $125,000 annually, with a maximum of $625,000 over five years with a maximum of 10 years. The Office of National Drug Control Policy (ONDCP), in partnership with the Substance Abuse and Mental Health Services Administration (SAMHSA), funded 161 new grants in August 2009, with a goal to extend these long-term coalition efforts. ONDCP funded 169 new grants in fiscal year 2010.

Through these grants, coalitions increase collaboration among 12 sectors in a community to target the needs of youth, their families, and the community as a whole.[1] The goals of these community coalitions are to: (1) increase collaboration among community agencies; (2) reduce risk factors and increase protective factors for youth; and (3) reduce substance use among youth.

---

[1] The 12 sectors include (1) Youth [persons <= 18 years of age], (2) Parents, (3) Business community, (4) Media, (5) Schools, (6) Youth-serving organizations, (7) Law enforcement agencies, (8) Religious or fraternal organizations, (9) Civic and volunteer groups, (10) Healthcare professionals, (11) State, local or tribal agencies with expertise in the field of substance abuse, and (12) Other organizations involved in reducing substance abuse.

Evaluation of DFC remains a significant priority, as the coalition movement has played a role in creating safer communities and can be a model for other public health issues outside of substance abuse prevention. While past evaluation studies have established a foundation for research and analysis, we believe that in the plan outlined in this document, we have the capacity to take research evidence to the next level. As we will demonstrate in this document, we will be able to exploit the natural variation in coalition operations to determine not only which coalition strategies are working, but also why they are working, how they are working, and in what situations they are working.

Evaluation Background

For two decades, communities have expanded efforts to address social problems through collective action. Based on the belief that new financial support enables a locality to assemble stakeholders; assess needs; enhance and strengthen the community's prevention service infrastructure; improve immediate outcomes; and reduce levels of substance use, DFC-funded coalitions have been able to implement strategies that have been supported by prior research.[2] Research also shows that effective coalitions are holistic and comprehensive; flexible and responsive; build a sense of community; and provide a vehicle for community empowerment.[3] Yet, there remain many challenges to evaluating them. Specific interventions vary from coalition to coalition, and the context within which interventions are implemented is dynamic. As a result, conventional evaluation models involving comparison sites are difficult to implement.[4]

Three major features of our evaluation approach allow us to expand upon previous analysis to include a far greater range of hypotheses concerning the coalition characteristics that contribute to stronger

---

DETERMINING "WHAT WORKS"

In this evaluation, we want to go beyond defining "success" at the coalition level. Each of the more than 700 DFC coalitions has unique strengths, and by understanding those strengths on a more granular level, we can provide more prescriptive guidance to coalitions on how to improve their operations and to ultimately achieve their objectives.

DFC coalitions aim to reduce substance use in their communities, as measured in this evaluation by past 30-day use. Although this is the core outcome measure of the evaluation, there are other key outcomes that logically precede – and follow – reductions in substance use. For example, precedents to reducing substance use include changes in attitudes (reflected in perceptions of risk and parental disapproval), environmental changes in the community (e.g., better lighting in areas where drug dealing takes place), and information sharing/education, which are all the products of complex processes where community partners come together to solve problems. We also expect reductions in substance use to have ancillary impacts, such as reductions in fatal crashes, reductions in crime, and even better academic performance. Although DFC coalitions have a clear goal to reduce substance use, our inquiries into "what works" will focus on all aspects of coalition functions and outcomes, including community collaboration, environmental strategies, changes in attitudes, reductions in substance use, and other resultant outcomes from reductions in substance use.

Operationalizing the concept of "success" in coalition activities will involve a two-step process. First, we will identify coalitions that had consistent, positive movements over time on outcomes of interest. We will then explore in depth the processes that would logically cause those movements to take place. By triangulating quantitative data and qualitative data gathered from coalition staff, we can have confidence in the attribution between process and outcomes.

---

[2] Brounstein, P. & Zweig, J. (1999). *Understanding Substance Abuse Prevention Toward the 21st Century: A Primer on Effective Programs*. Washington, DC: U.S. Department of Health and Human Services.

[3] Wolf, T. (2001). Community Coalition Building – Contemporary Practice and Research: Introduction. *American Journal of Community Psychology*, *29*(2), 165-172.

[4] Gruenewald, P.J. (1997). Analysis Approaches to Community Evaluation. *Evaluation Review*, *21*(2), 209-230.

outputs, stronger coalition outcomes, and ultimately stronger community outcomes.

- First, our approach will systematically deconstruct more encompassing measures (e.g., maturation stages) into specific constructs that are more clearly related to strategies and functions that coalitions must perform, and that define their capacity. This will provide measures of multiple coalition characteristics that may differentiate real world coalitions, may be important to producing effective coalitions, and may operate differently across different settings and in different coalition systems.

- Second, our approach uses the natural variation approach, in which we constantly look and test for differences in coalition organization, function, procedure, management strategy, and intent which may provide concrete lessons on how to construct effective coalitions in diverse settings.

- Third, our design and analysis uses a multi-method approach in which different "sub-studies" within the large DFC National Evaluation project umbrella can provide unique opportunities to contribute to project lessons. For example, the case study component of our design and analysis through the use of site visits will provide strong opportunity to implement many of the analyses identified in the discussion of the logic model that follows in the next section of this paper. The rich data attained during site visits to coalitions, combined with our process and outcome data, will serve this purpose. Over time, as the site visit data set grows in size, these rich measures will produce a valuable analytic database.

By better understanding the DFC Program and its mechanisms for contributing to positive change, the National Evaluation can deliver an effective, efficient, and sensitive set of analyses that will meet the needs of the program at the highest level while also advancing prevention science.

DFC National Evaluation Logic Model

At its first meeting in April 2010, the DFC National Evaluation Technical Advisory Group (TAG) identified the need for revision of the "legacy" logic model prepared by the previous evaluator. A Logic Model Workgroup was established and charged with producing a revised model that provides a concise depiction of the coalition characteristics and outcomes that will be measured and tested in the national evaluation. The TAG directed the Workgroup to develop a model that communicates well with grantees, and provides a context for explicating evaluation procedures and purposes.

The Workgroup held its first meeting by telephone conference in July 2010. In the following two months, the committee (1) developed a draft model, (2) reviewed literature and other documents, (3) mapped model elements against proposed national evaluation data, (4) obtained feedback from grantees through focus groups at the CADCA Mid-year Training Institute in Phoenix, (5) developed and revised several iterations of the model, and (6) produced the recommended logic model shown in Exhibit 1. The National Evaluation logic model has six major features, described below, that define the broad coalition intent, capacity, and rationale that will be described and analyzed in the National Evaluation.

Theory of Change. The DFC National Evaluation Logic Model begins with a broad theory of change that focuses the evaluation on clarifying those capacities that define well functioning coalitions. This theory of change is intended to provide a shared vision of the overarching questions the National Evaluation will address, and the kinds of lessons it will produce.

Community Context & History. The ability to understand and build on particular community needs and capacities is fundamental to the effectiveness of community coalitions. The National Evaluation will assess the influence of context in identifying problems and objectives, building capacity, selecting and implementing interventions, and achieving success.

Coalition Structure & Processes. Existing research and practice highlights the importance of coalition structures and processes for building and maintaining organizational capacity. The National Evaluation will describe and test variation in DFC coalition structures and processes, and how these influence capacity to achieve outcomes. The logic model specifies three categories of structure and process for inclusion in evaluation description and analysis:

- *Member Capacity*. Coalition members include both organizations and individuals. Selecting and supporting individual and organizational competencies are central issues in building capacity. The National Evaluation will identify how coalitions support and maintain specific competencies, and which competencies contribute most to capacity in the experience of DFC coalitions.

- *Coalition Structure*. Coalitions differ in organizational structures such as degree of emphasis on sectoral agency or grassroots membership, leadership and committee structures, and formalization. The logic model guides identification of major structural differences or typologies in DFC coalitions, and assessment of their differential contributions to capacity and effectiveness.

- *Coalition Processes*. Existing research and practice has placed significant attention on the importance of procedures for developing coalition capacity (e.g., implementation of SAMHSA's Strategic Prevention Framework). Identifying how coalitions differ in these processes, and how that affects capacity, effectiveness, and sustainability is important to understanding how to strengthen coalition functioning.

Coalition Strategies & Activities. One of the strengths of coalitions is that they can focus on mobilizing multiple community sectors for comprehensive strategies aimed at community-wide change. The logic model identifies the role of the National Evaluation in describing and assessing different types and mixes of strategies and activities across coalitions. As depicted in the model, this evaluation task will include at least the following categories of strategies and activities.

- *Information & Support*. Coalition efforts to educate the community, build awareness, and strengthen support are a foundation for action. Identifying how coalitions do this, and the degree to which different approaches are successful, is an important evaluation activity.

- *Enhancing Skills*. This includes activities such as workshops and other programs (mentoring programs, conflict management training, programs to improve communication and decision

making) designed to develop skills and competencies among youth, parents, teachers, and/or families to prevent substance use.

- *Policies / Environmental Change*. Environmental change strategies include policies designed to reduce access; increase enforcement of laws; change physical design to reduce risk or enhance protection; mobilize neighborhoods and parents to change social norms and practices concerning substance use; and support policies that promote opportunities and access for positive youth activity and support. Understanding the different emphases coalitions adopt, and the ways in which they impact community conditions and outcomes, are important to understanding coalition success.

- *Programs & Services*. Coalitions also may promote and support programs and services that help community members strengthen families through improved parenting; that provide increased opportunity and access to protective experiences for youth; and that strengthen community capacity to meet the needs of youth at high risk for substance use and related consequences.

Community & Population-Level Outcomes. The ultimate goals of DFC coalitions are to reduce population-level rates of substance use in the community, particularly among youth; to reduce related consequences; and to improve community health and well-being. The National Evaluation logic model represents the intended outcomes of coalitions in two major clusters: (1) core measures data, which are gathered by local coalitions, and (2) archival data (UCR, FARS), which will be synthesized by the National Evaluation team. These data will be utilized to assess the impact of DFC activities on the community environment and on substance use and related behaviors.

- *Community Environment*. Coalition strategies often focus on changing local community conditions that needs assessment and community knowledge identify as root causes of community substance use and related consequences. These community conditions may include population awareness, norms and attitudes; system capacity and policies; or the presence of sustainable opportunities and accomplishments that protect against substance use and other negative behaviors.

- *Behavioral Consequences*. Coalition strategies are also intended to change population-level indicators of behavior, and substance use prevalence in particular. Coalition strategies are also expected to produce improvements in educational involvement and attainment; improvements in health and well-being; improvements in social consequences related to substance use; and reductions in criminal activity associated with substance use.

Line Logic. The National Evaluation Logic Model includes arrows representing the anticipated sequence of influence in the model. If changes occur in an indicator before the arrow, the model represents that this will influence change in the model component after the arrow. For the National Evaluation Logic Model, the arrows represent expected relations to be tested and understood: How strong is the influence? Under what conditions does it occur?

In summary, the National Evaluation Logic Model is intended to summarize the coalition characteristics that will be measured and assessed by the National Evaluation team. The model depicts characteristics of coalitions that will be described as they present themselves, not prescriptive recommendations for

assessing coalition performance. This model is intended to guide an evaluation process through which we can learn from the grounded experience of the DFC coalitions who know their communities best. The model uses past research and coalition experience to provide focus on those coalition characteristics that we believe are important to well functioning and successful coalitions. The data we gather will tell us how community coalitions implement these characteristics, what works for them, and under what conditions. In this sense the model is an evolving tool -- building on the past to improve learning from the present and to create evidence-based lessons for coalitions in the future.

Exhibit 1. DFC National Evaluation Logic Model

Theory of Change: Well functioning community coalitions can stage and sustain a comprehensive set of interventions that mitigate the local conditions that make substance use more likely.

Coalition Structure & Processes → **Implementation of Strategies & Activities** → **Community & Population-Level Outcomes**

Member Capacity
(Acquisition & support of individual & organization member competencies)

Coalition Structure
(*Membership, leadership, gover-nance, organization structures***)**

Coalition Processes
(Assessment, planning, collaboration, evaluation, decision making, & inclusiveness)

Information & Support
(Providing information & support for prevention awareness, policy, resources & activities)

Policies / Environmental Change
(Policies & environmental barriers & opportunities that support prevention)

Programs & Services
(Support of programs & services providing positive skills and opportunities)

Community Environment
- *Awareness & norms*
- *Systems & policies (e.g., schools, juvenile justice)*
- *Sustainable opportunities & accomplishments*

Behavior & Consequences
- *Substance use*
- *Educational engagement & attainment*
- *Health, social & criminal consequences*

- Readiness
- Ability to mobilize
- Defines capacity from which the coalition builds

- Shapes coalition objectives
- Shapes strategies & activities
- Determines how objectives & strategies can be implemented.

- Defines priority improvements (health & social consequences, substance use, changes in community conditions).

Community Context and History

The SAMHSA Strategic Prevention Framework

The SAMHSA Strategic Prevention Framework (SPF) is an integral component of the logic, data collection, and analysis proposed here. There are several reasons for this:

- *Continuity*. The SPF guidance for assessing and implementing coalition planning, decision making, and performance improvement has been a part of the DFC initiative since its inception. The SPF is part of the training and guidance that grantees receive through participating in the initiative; it is incorporated into the data collection system and current tools used by the DFC Program; and is fundamental to the analysis and reporting on the DFC Program that has been completed to date (e.g., the maturation typology is based upon data that has the SPF stages incorporated into it). Given its central role in the development of the DFC initiative and evaluation efforts to date, maintaining the SPF as a key organizing concept for the evaluation is a high priority.

- *Important Step in Prevention Progress*. SAMHSA's introduction of the SPF was an important advancement for the prevention field. For more than two decades, the Center for Substance Abuse Prevention (CSAP) and other public, non-profit, and private organizations had sponsored research and evaluation concerning the causes and consequences of substance use and how to prevent it. This research has developed important knowledge concerning the epidemiology of initiation of substance use (e.g., risk and protective factors); important information concerning the social and individual consequences of use, including recent advances in understanding the biological impacts that contribute (e.g., recent research on the adolescent brain); and a growing body of knowledge on the effectiveness of specific policies, programs, and practices that have made the promotion of evidence-based practices possible. Research has also highlighted the importance of planning, implementation, and the use of data in promoting effective prevention initiatives, and has included important work on the role of community-based coalitions in promoting and sustaining community improvements. Notably, in the early 1990's, SAMHSA launched the ambitious Community Partnerships Program through which hundreds of coalitions were funded and evaluated across the country. This initiative demonstrated the potential of community coalitions. It also demonstrated some of the most important challenges they faced. The Community Partnerships Program emphasized empowering local community members to identify and develop solutions to their community's problems. Many partnerships faced challenges in effectively organizing and planning, and often struggled to identify strategies and actions that would help them achieve their goals. While the Community Partnerships Program produced notable successes that have had sustained impacts in the communities, it also produced important lessons on the tools and assistance that communities often need to plan and implement strategies using the growing knowledge about how to successfully meet real needs in their communities.[5] Importantly, SAMHSA's SPF incorporates many of the lessons learned from earlier research on community coalitions, and integrates them with the growing

---

[5] Phillips, J.L. & Springer, J.F. (1997). Implementation of community interventions: Ten lessons from the community partnerships program. *The Secretary's Handbook.* Washington, DC: SAMHSA, Proceedings of the Secretary's Convening.

knowledge and capacity in prevention. It focuses attention on a planning, implementation, and evaluation process that addresses many of the needs that challenged coalitions.

- **SPF Contributions to Coalition Functions**. Importantly, we will maintain and enhance our use of the SPF in organizing data collection and analysis to more thoroughly test its contributions to address community coalition needs. By addressing the lessons learned from studying coalitions, the SPF framework makes the following important contributions:

  o *Process Components*. The SPF identifies seven contributing factors for planning and managing effective prevention strategies: assessment, capacity, planning, implementation, evaluation, cultural competence, and sustainability. Each of these components requires an identifiable set of competencies and skills, and each makes an important, identifiable contribution to effective planning and management. While these components are not new in the discussion of prevention, they have often been treated in a fragmented manner with insufficient attention to their complementary roles. They provide a comprehensive reference for coalition practitioners and researchers. In addition, they provide a common language to the substance abuse prevention field for the purposes of community mobilization and planning.

  o *Inter-relationship of Components*. These seven components are often presented as five steps with cultural competence and sustainability as overarching considerations that pervade all five. We will refer to the first five components as steps at times in this plan because there is an approximate chronological logic to them. For example, assessment of need and capacity logically informs capacity building and strategic planning; evaluation information is logically produced and used after plans have been implemented and results observed. However, the SPF framework makes it clear that the five steps should be conceptualized as a continuous cycle. Indeed, specific coalitions may be strong in some steps when they enter the National Evaluation, and activities relative to the five steps occur simultaneously. Furthermore, cultural competence and sustainability are criteria for consideration in each of these steps, and cannot be attained fully if they are not incorporated into each step. For example, one foundation for cultural competence must be information on the differential needs of diverse community populations. This must be a consideration in assessment of needs and existing capacity. As another example, research on sustainability has shown that developing workgroups involved in service delivery (not only planning) across agencies and organizations is crucial to sustaining (e.g., institutionalizing) innovations in activities and services. This requires consideration at planning and implementation stages, and cannot be handled as a separate process, or simply as a resource issue.

  o *Systems Perspective*. In addition, the SPF includes a systems perspective, emphasizing that these components produce results only when they work together in a single system. While coalitions will have different emphases and skill levels across the five SPF steps, and may accomplish them differently, it is critical that they all be addressed with reasonable effectiveness and balance. One of the issues confronting Community Partnership coalitions was an overemphasis on the planning function, with little clear connection to

the activities that were eventually implemented. Similarly, a coalition cannot stay focused on real problems and progress without addressing the assessment and evaluation components of the SPF process. As noted above, sustainability and cultural competence will not be effectively achieved unless they are considered as characteristics of the entire system with implications for each component, and for system processes.

o   *Evidence-Based Practice.* The SPF creates a place and process for evidence-based practice to be incorporated into coalition planning. One of the persistent challenges to prevention practitioners in selecting and implementing evidence-based practices has been matching and adapting them to meet identified needs, environmental characteristics, population characteristics, and implementation capacity. Cultural competence is a particularly important consideration in SPF's emphasis on linking these decisions to assessment, capacity building, planning, and evaluation. The adoption and maintenance of evidence-based practices and continuous quality improvement will contribute to longevity of positive outcomes, and institutionalization of effective policies, programs, and practices. SAMHSA has facilitated the adoption and adaptation of evidence-based practices by coalitions, which often must adapt to resource constraints and conditions in the community context, by providing more flexible guidelines that coalitions can use to make decisions about identifying policies, programs, or practices as evidence-based, identifying core components, and making appropriate adaptations. A final implication of SPF's focus on systems is that evidence-based practice applies to organizational processes and capacity as well as to strategies and practices for interventions. For example, evidence on sustainability demonstrates that it requires resolving multiple concrete challenges, such as recruiting and maintaining participation, integrating and institutionalizing specific work relationships in the coalition system, maintaining long-term outcomes, and the acquisition of resources.

o   *Importance of Context.* The SPF framework and its focus on continuous, data-sensitive decision making places an emphasis on the importance of context, and the importance of continuous monitoring and consideration of local conditions. These community contexts are diverse, taking on distinct configurations across communities. For example, cultural competence may focus on racial/ethnic populations in one community, but focus on religious, socio-economic, or regional cultural values, beliefs and behaviors in others. The systems concept, and the focus on communities, makes context an important factor in prevention planning, decisions, and activities, as well as the measures and analyses used in the National Evaluation.

o   *Data-Based Decision Making.* Data-based decision making reflects the same orientation to empirically and scientifically-based policy and practice as the promotion of evidence-based practices, but with an important difference. While evidence-based practices use the accumulated, science-based knowledge produced through research and evaluation, the challenge is to apply this knowledge to a specific, current circumstance. Data-based decision making strives for real-time (or close to real-time) production of data that monitors needs, management of inputs and outputs, and attainment of outcome objectives to inform policy and management decisions by a coalition. The challenge is to

produce accurate and reliable information that is relevant and can be used by decision makers. This means creating decision systems (processes) that make information available and set expectations for its use. Thus, the SPF process sensitizes practitioners to the importance of both the availability and the use of quality information. Consistent with the systemic perspective of the SPF, the capacity for both producing and using data must be developed. Evaluation within the SPF is an essential component of process and performance monitoring and feedback to inform decisions about improvement of coalition strategy.

- **_SPF Challenges and Knowledge Development Opportunities_**. In addition to providing a framework for applying the research, evaluation, and experiential lessons generated in past studies and applications of prevention, the SPF provides an important guide for improving knowledge and practice. It also provides a guide for identifying the challenges that need resolution to achieve further improvement in research-based knowledge and its application in prevention. Accordingly, our research design, data collection, and analysis plan will provide tests of current hypotheses about how implementation of SPF constructs will improve coalition functioning and community outcomes. We will do this by (a) developing improved, practical measures of the manner and strength of coalition implementation of each step, and the way in which the outputs of each stage contributed to other components of the process (e.g., to other steps, to sustainability, to cultural competence); (b) articulate specific hypotheses implicit in guidance about how to implement SPF steps, and expectations about their benefits on capacity and outcomes; (c) test these hypotheses using our process and outcome measures and the mixed-method analysis plan defined in Section 4 of this plan; (d) refine and strengthen theory about SPF application and impacts; and (e) test revised hypotheses concerning application of the SPF to improve each of the five component areas, and ultimately strengthen attainment of community outcomes. Examples of areas of opportunity for knowledge expansion include:

    o _Are Expectations about SPF Best Practices Empirically Confirmed?_ As elaborated later in this plan, the current measures of SPF accomplishments are based on expectations about what constitutes strength in each area. The empirical evidence about how specific structural and procedural characteristics contribute to each component is limited, and these specifics are critical to providing relevant guidance to practitioners. Our analysis is designed to deconstruct the current conceptualization of each SPF step and assess more precisely and concretely how coalitions organize and act to successfully fulfill the necessary functions of each SPF step.

    o _How Can SPF Practices Be More Objectively Measured?_ Currently, measures of progress in core components are largely based on the perceptions of an informed observer in each local coalition. This reliance on a single observer has potential for measurement error (random and non-random) that may compromise the sensitivity and value of the SPF construct in our analyses. We will assess this potential with analyses of consistency, discrimination, and correlation, and recommend feasible improvements to establish more objective behavioral and event indicators.

o *Do SPF Best Practices Differ by Setting?* Current knowledge concerning the SPF assumes significant homogeneity in the markers of accomplishment in each of the components. The diversity of coalitions and their settings is emphasized in past DFC reports, but it is not empirically specified, nor are the implications for approaches to successfully meet SPF functions. Our design and analysis provides tests of these conditions.

o *How Do SPF Best Practices Contribute to Maturation?* Research on coalition functioning generally, and on DFC coalitions specifically, has identified coalition maturation – the systematic building of capacity that will improve effectiveness in attaining outcomes – as a central concept in developing knowledge and best practices with respect to community coalitions. While the maturation concept is widely used, there is little consensus on exactly what coalition structures, procedures, or activities are most important to defining it, or exactly how it contributes to effectiveness in achieving outcomes. The evaluation of earlier cohorts of DFC coalitions developed a measure of coalition maturation called the Coalition Classification Tool (CCT) that provided criteria for assigning coalitions to levels of maturation. The SPF components are central to this measure, which assesses maturity in part by how well respondents report capacity in the SPF components. However, analysis demonstrated that these classifications could not be empirically validated with the statistical models that were used. Furthermore, there is no specific, deconstructed analysis of how these components contribute to maturation in different areas. Our design and analysis addresses this important specification of empirical knowledge by (a) developing better measures of the empirical manifestations of capacity in each step, and (b) testing their relationship to capacity and community outcomes.

o *What Components/Dimensions Are Core?* The SPF is a complex construct with multiple domains (components and crosscutting concepts) and dimensions within each domain. As with any strategy or program, the potential identification of core elements is of great value by specifying what functions and achievements are most important to output and outcome objectives. We will review existing work on these domains and dimensions (e.g., Renee Boothroyd's literature review on core competencies [15] and essential processes [12] within the SPF) and use this input in the development of deconstructed measures that will help identify the relative contribution of SPF domains and dimensions to coalition capacity and outcome effectiveness.

o *How Do SPF Steps Work Together as a Cohesive System?* As a final example, one of the important contributions of SPF noted above is that it provides a systems perspective. However, little of the evaluation or research on the SPF to date concerns the specification of how this system works. What makes an implementation of the SPF system coherent? What makes outputs (e.g., a needs assessment) from one SPF step useful for decisions that enhance output in another step, and ultimately in community outcomes?

In sum, we see the SPF as a central organizing feature of our evaluation. It has the great strength of being widely applied in current prevention research, evaluation, policy, and practice. This currency will strengthen the relevance of the DFC National Evaluation to practitioners in the field. It also provides a strong and organized guide for advancing knowledge concerning community-based prevention planning,

policy, and practice. Subsequent sections of this analysis plan provide more detail concerning how our multi-method design and analysis will provide evidence concerning the evaluation issues and questions identified above.

2. Evaluation Framework

Our approach to this evaluation is to move ONDCP to a progressively stronger evidence base for GPRA and PART reporting, while identifying best practices and providing more *practical* results for the field. The previous contractor's work has contributed to this progression in the past five years, particularly concerning the standardization of outcomes across grantees. In the next five years, this plan will further improve the validity and reliability of this evidence.

Exhibit 2 presents a top-line overview of our study design. This exhibit describes the (1) process of moving evidence to practice, (2) key stakeholder questions at each evaluation stage, (3) proposed study components that address each step in the process, and (4) proposed evaluation products that provide high-quality performance reporting, rigorous evaluation reports, and practical products that can aid in the replication of evidence-based practices.

> **DECONSTRUCTING EVIDENCE**
>
> Numerous independent efforts have been undertaken to identify evidence-based research on social programs, such as the U.S. Department of Education's What Works Clearinghouse, RAND's Promising Practices Network, and SAMHSA's National Registry of Evidence-based Programs and Practices (NREPP). Typically, when evidence of a program's effectiveness is found, an intervention/program will get a "seal of approval" from these entities. In order for research to be truly useful to practitioners, however, we must deconstruct the evidence and identify core elements or practices that lead to better outcomes. Put another way, *communities cannot pursue effective replication strategies by simply knowing which coalitions are working; rather, they must know why they work, how they work, and in what situations they work.*

EXHIBIT 2. PROCESS OF ASSESSING DFC EVIDENCE
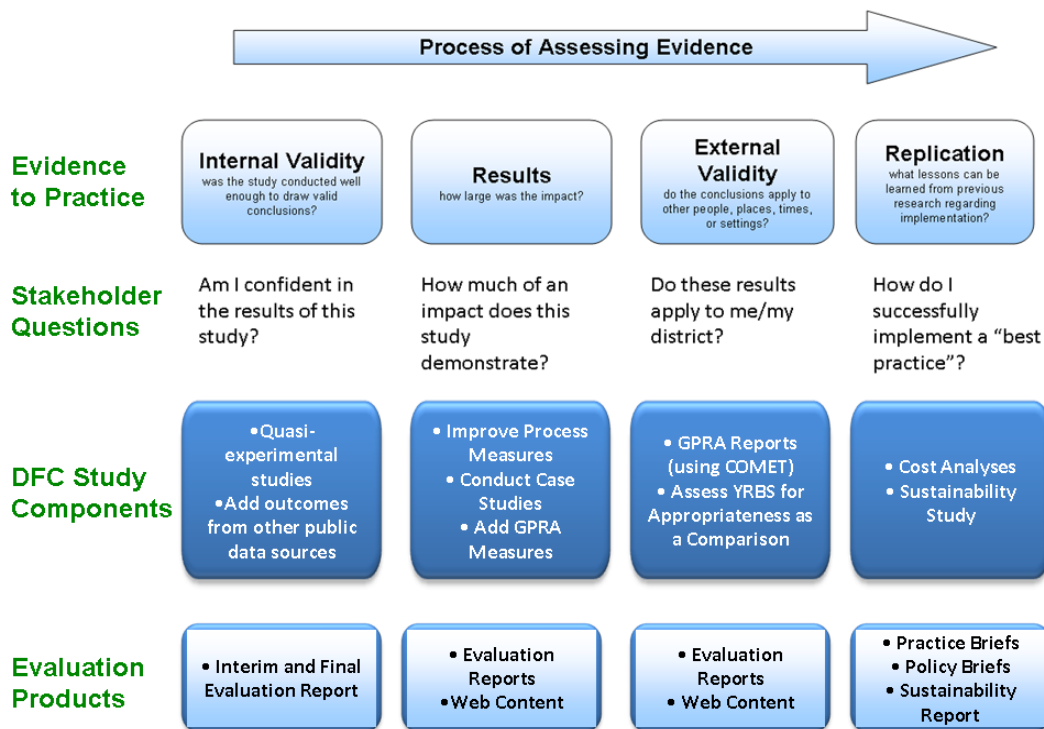
Moving Evidence to Practice

When a consumer of research–such as a legislator, school board member, or coalition member–reviews the effectiveness of various strategies, he or she is likely focused on a single question:

Will this strategy work in my community?

Many efforts currently underway do not go beyond assessing the quality of a study and its results, leaving the consumer to determine whether a strategy could be successfully replicated (and effective) in its community.[6]

While the assessment of whether a study's results are generalizable is a qualitative judgment best left to staff on the "front lines," there is still a great need to present information in a manner to help stakeholders make these types of judgments. In this evaluation, we will strive to provide a sufficient level of detail so every decision maker at the Federal, State, and local level can decide upon his or her best course of action to pursue in addressing prevention efforts.

---

[6] In 2007, the What Works Clearinghouse developed an "extent of evidence" rating to capture information on whether research has been tested in multiple settings (the first step in assessing external validity). A rating of "moderate to large" requires at least two studies and two schools across studies, and a total sample size across studies of at least 350 students, or 14 classrooms. Otherwise, the extent of evidence is rated as "small". This rating is not considered in the What Works Clearinghouse's evidence standards, however.

The top of Exhibit 2 presents a simplified framework for assessing evidence from both a researcher's and a consumer's point of view. The first step in the process is to assess the research evidence on a particular intervention and determine whether the study used scientific methods that could generate valid conclusions. Researchers call this "internal validity" and from a consumer's perspective, the implication of internal validity is to determine whether the results of the study are believable in the first place. Our evaluation design is intended to have sub-studies with strong internal validity; namely, the quasi-experimental comparison group studies. Other analyses that do not employ a comparison group design (e.g., GPRA reports) will have somewhat lower internal validity.

Next, both researchers and consumers have to **determine whether the study produced effects that are meaningful**. The degree to which results are meaningful can be determined in two major ways. The first focuses on the magnitude of the effect. Many researchers use the concept of statistical significance to determine whether a particular study had an effect; however, this is not necessarily a good idea since statistical significance is heavily dependent upon sample size, and because the most rigorous research (e.g., randomized studies) is often too costly to implement on a large scale. For example, with a study sample of 10,000 students, even small effects will be statistically significant while a sample of 50 students will rarely produce statistically significant results, even with relatively large effects. Ultimately, what matters most is not statistical significance alone, but rather, whether the size of an effect is meaningful in a practical sense. Researchers use the concept of effect size to determine this. In this evaluation, we will present evidence based on effect sizes, which tells us not only if results are meaningful, but also how results compare on different outcomes (e.g., we can put all results on a common scale so we can assess whether a given coalition had, for example, stronger results on perceptions of risk than they did on age of onset).[7] The second way in which results can be determined to be meaningful is *the established strength of the relationship on an observed outcome to associated, and often distal, outcomes*. For example, the relationship of binge drinking to future consequences is much stronger than that of simple 30-day use. Any behavioral measure of use has a stronger relation to consequences than perceived health risk. We will incorporate the growing research evidence on the strength of outcome indicators as predictors of consequences, as well as the prevalence of the behaviors, into our measurement of results.

If studies are conducted well and if results are meaningful, the next step is to determine whether the results are generalizable. Researchers call this "external validity" and it is an important factor to consider in the adoption of any program. *Consumers of research will likely want to know whether findings can be applied to their communities* and it is important to present sufficient context for State and local decision-makers to make informed choices on which strategies to adopt.

Finally, once a piece of research is found to be believable, to provide meaningful effects, and to be applicable to populations or settings of interest, thought must be given to replicating (and potentially adapting) that program. Realistically speaking, localities adopt strategies/initiatives and then tailor those strategies to address their risk factors, youth, and budgets. This, oftentimes, renders research outcomes at a strategy level obsolete, since the specific parameters of the intervention have been changed. What matters more is knowing which elements of a particular strategy to keep and which ones can afford to be modified. That is the level of detail we will strive to provide through our analyses.

---

[7] The presentation of effect sizes and statistical significance is not a mutually exclusive choice. In our analyses, we will present both metrics; however, our emphasis will be on effect sizes.

Evaluation Questions

Our ultimate goal in this next phase of the evaluation will be to implement a set of innovative, scientifically-based methods that will be both accepted by the research community and intuitive to non-researchers. Undoubtedly, this evaluation will need to address the concerns and needs of a number of stakeholders. Stakeholders in ONDCP, SAMHSA, and in the research and prevention communities will be focused on questions of internal validity and results of the DFC program, while stakeholders in the field will be focused on the identification of best practices and their replication. Our study design and analysis plan focuses on research questions that are relevant to key stakeholder groups, and, in collaboration with ONDCP, we will develop participative processes through which we can gather input from stakeholder groups concerning priority questions. One of our major emphases over the next five years will be to strengthen our design and analytic capability to answer relevant questions in ways that provide useful guidance based on the most rigorous and precise analysis possible within this initiative. Exhibit 3 provides a summary of select stakeholder groups (column 1), selected examples of research questions relevant to each stakeholder group (column 2), the products through which findings and lessons can be communicated to each group in a useful way (column 3), and a preliminary identification of analysis methods that our design will support, and that will provide answers to the questions (column 4).

EXHIBIT 3. KEY EVALUATION QUESTIONS, PRODUCTS, AND ANALYTIC METHODS RELEVANT FOR EACH STAKEHOLDER GROUP			
Stakeholder	Key Research Questions	Products	Methods
ONDCP and SAMHSA	• Do DFC coalitions have positive outcomes on the core measures? Are they improving over time?[8] • In what other ways are DFC coalitions having an impact?	• GPRA and PART reporting assistance • Interim & final evaluation reports	• Change analysis for GPRA/core measures/other outcomes/grantee outputs and outcomes in aggregate & by comparison groups • Comparison across maturation stages
Leaders of Coalitions	• What are the key ingredients to successful collaboration between community partners? • What potential pitfalls exist in the implementation process and how can they be avoided? • Are coalitions enhancing the prevention system? • Do coalitions make other systems receptive to the implementation of prevention science via a data-driven planning model? • What specific initiatives or strategies should be implemented to keep youth drug and alcohol free? • What specific practices are backed up by evidence? • What practices should be replicated in all coalitions and which practices are useful in specific contexts?	• Practice briefs • Interim & final evaluation report • Policy briefs • Website content	• Code & profile coalition assessments using qualitative data from site visits. • Profile and correlate process variables from site visits, replicate & confirm as site visits accumulate across years • Exploratory analysis using enhanced process measures • Confirmatory analysis of cross-site visit findings through enhanced process measures on all coalitions • Assessment of perceived value of strategies in site visits • Bivariate (exploratory) correlation, and multi-variety relation of practice measures with outcomes using coded site visit data in cross-site analysis, multi-variety (confirmatory) analysis • Correlate coalition strength measure (e.g., CCT and others) with community resources / other measures
Schools	• What specific initiatives should be implemented within schools to keep students drug and alcohol free? • Do coalitions have long-term effects on student achievement?	• Practice briefs (including cost information & a guide for replicating best practices)	• Assessment of perceived effectiveness of initiatives in site visits • Change analysis for GPRA /core measures re: substance use comparing coalitions differing in type and intensity of in-school initiatives • Mining of public use data (e.g., schoolmatters.com)
Local Governments	• What policies should be implemented to keep students and adults drug free? • What are the most cost effective ways to reduce substance use in communities?	• Interim & final evaluation report (executive summary for policymakers) • Policy briefs (best practices) • Practice briefs (including cost information)	• Assessment of perceived effectiveness of policies in site visits • Change analysis for GPRA/core measures re: substance use comparing coalitions differing in type and intensity of policies • Simple assessment and analysis of cost of policies related to effectiveness in site visit communities

---

[8] New core measures have been proposed for this evaluation. As part of our plan to implement these core measures, we will ask grantees to provide data from prior years on these measures, if practicable. In other words, in the collection of new core measures, we will try to immediately accumulate historical data to assess trends over time.

EXHIBIT 3. KEY EVALUATION QUESTIONS, PRODUCTS, AND ANALYTIC METHODS RELEVANT FOR EACH STAKEHOLDER GROUP			
Stakeholder	**Key Research Questions**	**Products**	**Methods**
Law Enforcement	• Are comprehensive community initiatives reducing the negative effects of alcohol and other drug use among youth (e.g., reductions in DUI/drugged driving)? Among adults?	• Website content • Practice briefs	• Multivariate assessment of relation between coalition strategies, implementation strength & community outcomes • Exploratory site visit, cross-site analysis, confirmatory multivariate analysis in comparison samples & full sample • Mining of public use data (e.g., UCR, FARS)
Social Service Agencies	• What resources do students need to stay alcohol and drug free? • What types of messages are most effective in the prevention of substance use?	• Practice briefs	• Assessment of perceived effectiveness of policies in site visits • Testing through site visit sample cross-site analysis
Judicial Agencies	• What policies have been most effective in reducing recidivism?	• Policy briefs	• Assessment of perceived effectiveness of initiatives in site visits

Exhibit 3 is a simple guide to the careful planning and implementation that will characterize our implementation of the evaluation design. A brief elaboration will demonstrate this point.

- *Relevance and Utility.* A major part of making findings useful is ensuring that the intended stakeholders see the findings as "actionable" – that they provide implications and lessons that can be acted on and implemented in the real world context of a coalition. Second, it is important that intended audiences believe that the guidance will "work" – that it will have the intended effect. Involving key stakeholders in the identification of relevant questions is critical to meeting these criteria for utility. We will work with ONDCP to create multiple channels of participation for stakeholders, focusing on the issue of relevant and useful questions and information. Avenues may include sessions at appropriate conferences, webinars, web-based suggestion boxes, or Internet surveys.

- *Products.* Research findings will not be used unless they are effectively disseminated and understandable. We will produce a variety of products designed to effectively convey information to the stakeholder groups, and will constantly monitor and improve the quality of these products through feedback from intended consumers.

- *Place in the Logic Model.* Our comprehensive research and analysis design is necessary to gain the rich perspective that is important to understanding the many systemic and situational factors that must be understood to improve coalition effectiveness. Our experience has taught us that fragmentation of the design and in analysis (i.e., conducting specific analyses in relative isolation from findings and information throughout the study) works counter to the comprehensive and integrated intent of the overall design. Our implementation process will consistently ensure that we place each specific analysis in the larger study context and consider influences and implications in the full study context. Later sections of the analysis plan will make this integrated use of our multiple methods more specific.

- *Analysis Methods.* The final column in Exhibit 3 identifies the sub-components of the analysis that will specifically address each question, and the components of the data set that will be used. This identification of method is preliminary and will be further specified as data sources are assessed and input on research questions is incorporated. The important point in early planning is to identify the need for specificity, and anticipate ways of developing necessary measures and analysis strategies.
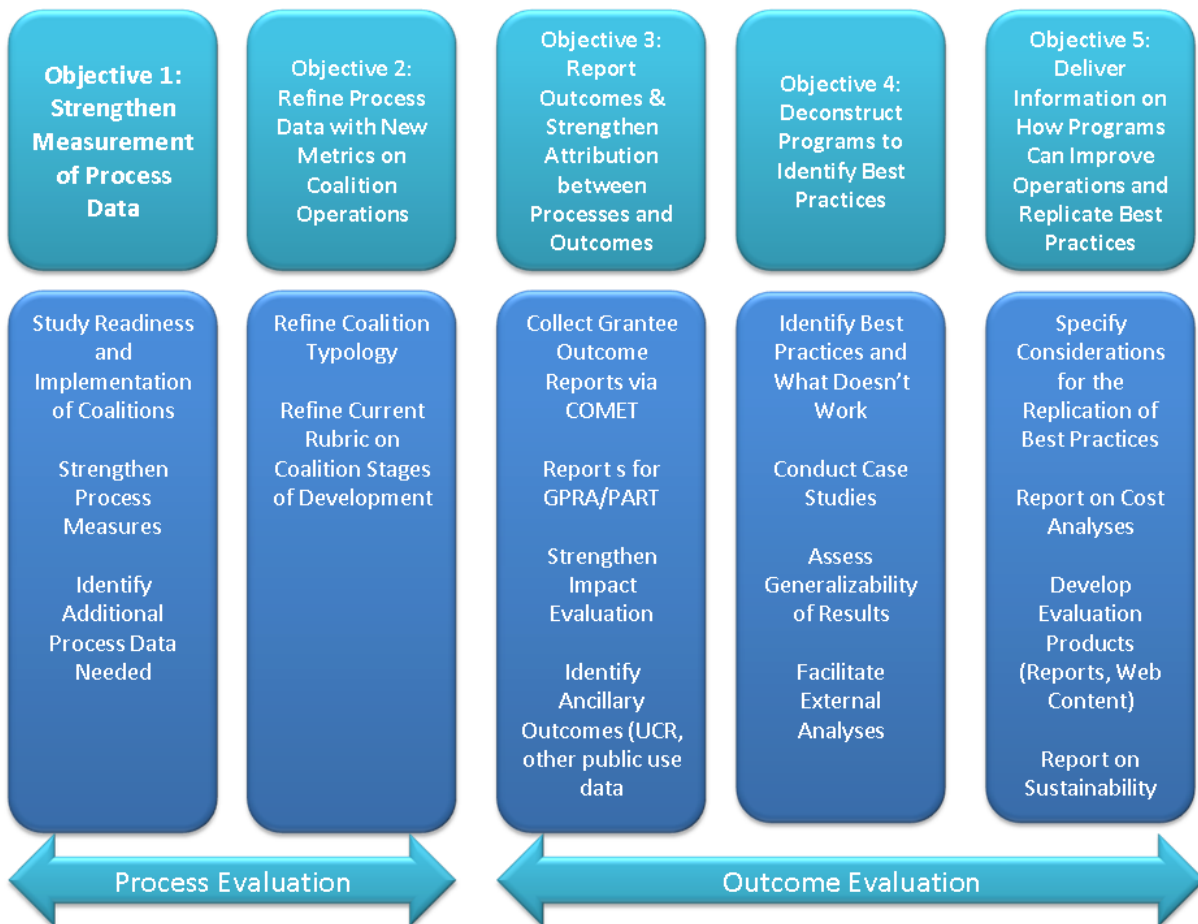
Evaluation Design

Exhibit 4 contains a simplified version of our evaluation plan. This exhibit demonstrates that there are five objectives (or stages) in the execution of this evaluation. They are:

- *Objective 1: Strengthen Process Measures.* As stressed throughout this plan, our analyses will greatly increase the level of detail with which coalition strategies are described and differentiated. Thus, details will support analyses of the degree to which different coalition structures, procedures, strategies, and implementation characteristics contribute to the achievement of the grantee and community outcomes identified in our evaluation logic model (Exhibit 1). Strengthening process measures is simply the necessary foundation for answering many of the stakeholder research questions previewed in Exhibit 3.

- *Objective 2: New Metrics on Coalition Operations.* The detailed measurement of process is necessary to accurately measure coalition structure, procedure, and activity. However, providing practical and generalizable guidance to coalition practitioners requires development of more encompassing metrics that characterize this detail in more general terms. These more general metrics "bundle" detailed measures into larger constructs that can guide planning, implementation, and capacity building across coalitions, and provide guidance to the settings, purposes, and populations to which they are most applicable. The Coalition Classification Tool (CCT) maturation categories are an example of such a multiple-item metric, and are the single major process measure in the previous evaluation. We believe that the CCT provides a single (yet important) dimension of coalition operations, and that other dimensions are needed (e.g., collaboration quality, types of coalition strategies, strength of implementation, capacity for SPF steps, cohesiveness, sustainability) to accurately encapsulate conceptually important process measures in analyses of outcomes. Our focus in this stage will be the development of new summary metrics – in addition to the maturation measure captured in the CCT – to further explain what is truly happening at the local level and how these other factors contribute to a coalition's effectiveness.

- *Objective 3: Outcomes and Attribution.* Simultaneous to the strengthening of process measures and coalition metrics, we will be strengthening community intermediate outcomes, substance use outcomes, and additional related outcomes (e.g., consequence data). Our design and analyses for demonstrating attribution of coalition effects on outcome measures has been strengthened through improved measurement and improved comparison design. The ability to explain the measurable coalition structures, strategies, and implementation characteristics that contribute to attaining outcomes has been provided by the strengthened process measures.

These metrics can be entered into multivariate models, which can identify contributing factors and specify them across different community settings, organizational contexts, and populations.

EXHIBIT 4. OBJECTIVES AND EVALUATION COMPONENTS

Objective 1: Strengthen Measurement of Process Data	Objective 2: Refine Process Data with New Metrics on Coalition Operations	Objective 3: Report Outcomes & Strengthen Attribution between Processes and Outcomes	Objective 4: Deconstruct Programs to Identify Best Practices	Objective 5: Deliver Information on How Programs Can Improve Operations and Replicate Best Practices
Study Readiness and Implementation of Coalitions Strengthen Process Measures Identify Additional Process Data Needed	Refine Coalition Typology Refine Current Rubric on Coalition Stages of Development	Collect Grantee Outcome Reports via COMET Report s for GPRA/PART Strengthen Impact Evaluation Identify Ancillary Outcomes (UCR, other public use data	Identify Best Practices and What Doesn't Work Conduct Case Studies Assess Generalizability of Results Facilitate External Analyses	Specify Considerations for the Replication of Best Practices Report on Cost Analyses Develop Evaluation Products (Reports, Web Content) Report on Sustainability

Process Evaluation ←→   Outcome Evaluation ←→

- *Objective 4: Identify Best Practices.* Our focused analyses of contributions to effectiveness by different coalition strategies, our case study analyses, and our cross-site comparisons for site visit coalitions will contribute to a strong ability to identify best practices and test the degree to which they can be generalized.

- *Objective 5: Deliver Useful Best Practice Guidelines.* The mixed-method richness of the analysis and interpretation provided by our evaluation design will support a variety of products, such as policy briefs and best practices briefs, to convey relevant, understandable, and useful lessons and best practices to policy makers, coalition practitioners, and other stakeholders and interested parties.

Objective 1: Strengthen Measurement of Process Data

As noted above, the strengthening of process data is fundamental to our approach to strengthening the attribution between process and outcome data (which can be extraordinarily difficult, especially at the community level). It also will provide tools to identify potential implementation challenges before they

happen. The ICF team will conduct an in-depth study of core processes that are being implemented and cross-reference key developments in the literature, our Technical Advisory Group's guidance, and our previous efforts working with coalitions to identify new process measures needed for collection. Our approach to strengthening process measurement involves the following steps.

Step 1. Literature Review. The development of stronger process measures will begin with a literature review. The team will develop parameters for the literature review, which will ensure that our efforts will stay focused on the core needs of the evaluation. In addition, the team will develop a structured abstract protocol that will ensure that the appropriate information is being collected. This literature review will be supplemented by past and current studies of coalition processes, many of which ICF team members have already conducted. For example, this will include:

- Various surveys that focus on comprehensive community initiatives in the prevention arena.

- State studies on SPF-SIG coalitions (e.g., Tennessee) have identified systemic features of coalition activity associated with stronger processes. For example, the presence of adult drug courts facilitates the use and effectiveness of environmental strategies related to DUI enforcement by providing systemic capacity to effectively address these cases and promote positive outcomes.

- Studies of coalitions focusing on integrating services (e.g., SAMHSA's Starting Early Starting Smart multi-site evaluation) have documented the importance of integrated line work groups for sustainability of innovation.

- The substantial literature (both refereed and fugitive) produced concerning SAMHSA's Community Partnerships Initiative. This is an important source because many of the focuses of the SPF and concepts of coalition functioning (e.g., various maturation or developmental models) were initially generated in the substantial research generated by this large and well-funded initiative. It will provide a strong basis for identifying both hypotheses concerning contributors to effectiveness and challenges commonly faced by coalitions.

Step 2. Review and Assessment of Current DFC National Evaluation Measures. Process measurement in the DFC National Evaluation to date has largely used data from two sources: (a) the Coalition Classification Tool (CCT) and (b) the Coalition Online Management and Evaluation Tool (COMET), a web-based performance monitoring system. Simultaneously with the literature review, we will conduct a thorough review and assessment of the current measures, their strengths and limitations, and make recommendations for their use, revision, or augmentation in the evaluation. Briefly, some of the relevant issues will include the following.

- *The Coalition Classification Tool.* The CCT is a tool that asks an informed coalition member to provide judgments about the coalition's performance or characteristics with regard to four functional areas: (1) Coalition development and management, (2) Coordination of prevention programs/services, (3) Environmental strategies, and (4) Intermediary of community support organization. Very little analysis of the data gathered through the CCT has been reported in published evaluation reports to date. Analyses of the association of maturation stages defined by the CCT to other judgments of coalition function, and, more importantly, analyses showing moderate association of coalition type to outcome effectiveness are the focus of what has been

reported. This relative paucity of analysis reporting concerning this core source of process data raises several points.

o The actual CCT maturation classifications are based on only a portion of the data gathered through the instrument. It uses six general rating items for each of the four functional areas identified above. The total measure includes 24 items. As noted, the CCT items are "global" which exacerbates several common problems with closed-ended key informant reports on organizational processes and status. First, the respondent is being asked to report on complex organizational processes from her individual perspective. Second, the item statements on which the respondent is to rate the organization's performance from novice to mastery are often very complex, including multiple items that appear to be duplicative of one another (or so closely related, they are duplicative for all intents and purposes). It is difficult to know the precise empirical reference upon which the response is based. In short, these items do not provide clear empirical referents, and thereby need to be augmented with additional analyses if they are to provide clear "actionable" lessons learned.

o There are many additional items in the CCT (in fact, more than 100), and many provide a more concrete empirical referent than the "general ratings" used in the CCT. For example, questions 3 and 4 in the CCT ask about characterizations of the organizational structure and status of the coalition. This is potentially important information concerning the diversity of coalition organization, and the need for assessing the homogeneity of best practice across different types. However, this information (and many other potentially useful items) has not been profiled or associated with CCT items, types, or the full range of items in the tool.

o Accordingly, exploratory analyses of the CCT, including correlations and dimensional and clustering analyses will be a primary analytic product in our evaluation plan.[9] This analysis will include the entire instrument, and not just the small number of items that are included in the maturation typology. The objectives will be (a) to better understand the profiling [e.g., similarities and differences, both univariate and multivariate] of coalitions that is important to understanding whether analyses can be meaningfully aggregated or must be disaggregated; (b) to assess the measurement quality of CCT items; and (c) to assess the degree to which CCT items, scale scores, or sub-scale scores correlate.

These analyses will provide substantial information on the degree to which CCT items form dimensions or clusters, the quality of items in terms of variation and contributing to these measurement dimensions, and the degree to which groups of items may form meaningful measures of strategies or capacities that vary across coalitions.

• *Coalition Online Management and Evaluation Tool Items.* COMET process data has received even less analytic attention than CCT in DFC National Evaluation analyses to date. The National

---

[9] This task has been completed and our results are fully described in the DFC National Evaluation's Systems, Measures, and Tools Plan, which has been published under separate cover.

Evaluation team has begun a process of sorting, organizing, and analyzing COMET data that is similar to that described above for the CCT items. The COMET data provides opportunity for several important types of exploration.

o   We will identify variables that can help profile the amount and nature of diversity in coalition characteristics, and the degree to which they may form distinct types of strategy, structure, or some other set of coalition structures, procedures, or implementation characteristics. For example, the reporting of strategies will support analysis of the frequency and distribution of strategies, whether strategies tend to co-occur to form distinct types, and what proportion of coalitions fall into distinct types.

o   We will organize reporting into longitudinal repetitions suitable for assessing change and development relevant to the grantee outputs and outcomes in the logic model. This may allow more precise tests of maturation based on strategies rather than perception, or tests of the sequencing of events (e.g., increases in quality of implementation following TA and training events).

o   We will match select data from COMET with CCT data and assess consistency (correlations) of items/measures hypothesized to change together. These may be alternate measures of the same construct, or measures that would be hypothesized to vary together or sequentially based upon the logic model and theories of coalition development or intervention effect. In this manner, analyses could be used to cross-check and validate findings or indicate areas for further exploration.

These analyses will greatly increase our ability to assess the measurement quality of existing measures, to identify those that may be replaced,[10] and to better understand the characteristics and diversity of coalitions.

- *Map Measures Against Elaborated (Internal) Logic Model.* Based on the literature review and enhanced understanding of the profile and diversity of DFC coalitions, we will revise and elaborate the logic model identifying conceptually important constructs that should be measured at each stage of the model. This task was recently completed by the Logic Model Workgroup and details of the internal logic model are presented in Exhibit 5. We will then map the potential measures from existing data onto these constructs, and identify needs for revision or addition of measures.

- *Recommend Revisions in Measures.* Following this cross-reference of the key processes and constructs identified in the literature with specific survey items from well-known prevention surveys (including our own and existing COMET data collection), the team will recommend the measures for future data collection. We will then present these findings to our COTR for review and comment, and where possible, we will propose that new measures be implemented in a checklist or matrix format (i.e., instead of moving to a new screen for each activity, checklists or matrices will be developed to capture all data on activities on a single screen).

---

[10] For the great majority of items, continuity with past analysis is not a critical concern since little past analysis has been publicly reported.

Before any new measures are approved, the ICF team will assess response burden for these additional questions, as well as expected response burden for the entire data collection effort. Specifically, we will assess not only the time it takes to complete all data collection requirements for DFC, but also the time it takes to transfer data from one source to another. Many coalitions have multiple data collection requirements from multiple funders, and our burden-reduction efforts must take these other requirements into consideration. By identifying strategies to streamline data collection and reporting efforts all funders (either through minimum data set or data export strategies), we will ensure that the totality of response burden is considered.

For DFC requirements specifically, we will aim to make any changes time-neutral in terms of response burden, as we want to ensure that prevention practitioners stay focused on their jobs and not on data collection requirements. We expect that making a good-faith effort to keep response burden down will also result in stronger buy-in for evaluation activities from practitioners.

EXHIBIT 5. INTERNAL DFC NATIONAL EVALUATION LOGIC MODEL

CONTEXT	COALITION STRUCTURE AND PROCESSES	STRATEGIES AND ACTIVITIES	COALITION CAPACITY	COALITION EFFECTIVENESS
Community • Sense of community / Community social organization • Community readiness for change • ATOD needs • ATOD prevention resources • Related initiatives • Socio-economic context • Critical events • Stability **Coalition** • Organizational history • Past collaboration involvement • Funding history • Past accomplishment • Institutional / grassroots focus • Critical Events	**Member Competency** • Member coalition involvement • Member relationships • Skills and knowledge **Structure** • Formalized structures • Degree of centralization • Membership breadth and diversity • Leadership structure • Institution / grassroots mix • Workgroup/committee organization **Processes** • Formalized procedures • Task and role clarity • Supports for member involvement • Leadership style • Communication • Conflict resolution • Shared decision making • Information / expertise input and use • Strategic planning framework (SPF) • Use data to inform decisions • External relations / resource access	**Coalition Role** • Membership and Focus • Direct implementation • Advocacy • Initiation / support • Member contributions • External partnering **Programmatic Capacity** • Clear, realistic objectives • Unique, innovative services • Need driven • Cultural competence • Use of evidence-based practice **Strategy / Activity Mix** • Information / Education • Coordination • Environmental policy / enforcement • Environmental assets and opportunities • Direct service programs **Coalition & Community Outputs** • Coalition outputs • Population targets/ penetration • Volume (dose) • Comprehensiveness	**Coherence** • Shared vision • Shared solutions • Inclusiveness • Broad involvement • Commitment /Satisfaction **Coalition Climate** • Open • Trusting • Values diversity • Shared input **Positive External Relations** • Non-member agencies • Activity focus • Access community resources • Community role / recognition **Capacity Building Effort** • Outreach / recruitment • Access to Training / TA • Build member, coalition, and programmatic capacity **Continuous Improvement** • Data based & Systematic • Explicit decision points (e.g., annual action plans) • Responsive / adaptive	**Community/ System Outcomes** **Norms & Awareness** • ATOD risk perception / acceptance • Media **Systems & Policy Change** • Collaboration • Service integration • Public / business /education policies **Sustainable Accomplishments** • Systems • Services **Community Behavioral Outcomes** **Substance Use Prevalence** • Core use measures (revised) **Contributors & Consequences** • Educational performance • Health & wellness • Crime • Family • Social • Economic

Discussion: Examples of Potential Additional Process Variables

From our review of existing data collection protocols, we can predict that the likely outcome of this effort will be a focus on measures of implementation quality. While the current evaluation has focused process measures on a coalition's stage of development, this only tells part of the story. Even more important (and more difficult to measure) is the quality of the coalition's collaboration and outreach efforts. We expect that adding dimensions to our understanding of coalition processes will put us in a more favorable position to present useful outcomes to ONDCP, its Federal partners, prevention practitioners, and other stakeholders in this evaluation. By defining quality processes, we will also be in a better position to help ONDCP and other stakeholders provide guidance and assistance to DFC grantees, as well as recommend new criteria for the grants award process.

Collaboration Quality

While the evaluation field has not fully investigated how certain collaborative variables and dynamics lead to successful coalitions, it is not due to a lack of measures. In 2004, Granner and Sharpe identified more than 140 different measures of scales of collaborative variables through a literature and web-based search. We can identify adequate measures from Granner and Sharpe's (2004) review, but ICF also has established measures and has been very successful at tailoring them to specific community change initiatives in the past. In determining the addition of a new measure, the first and most important issue is relevance: is the variable truly meaningful to the majority of DFC communities and is it relevant or integral for evaluation purposes? In terms of psychometrics, a major issue is reliability (the consistency of a measure), and a common rule of thumb is that measures should have a reliability of .70 or higher (Nunnally & Bernstein, 1994). On the other hand, another aspect that drives reliability is the number of items –more items results in higher reliability, but also increased burden on respondents. To provide some context as to what new process measures may strengthen the evaluation, below we briefly summarize some initial or preliminary thoughts that could be the starting point of our collaborative discussion about *what process measures can add value to the evaluation and inform practice.*[11] The National Evaluation team will follow up the submission of this evaluation plan with appendices that document, at the variable/item level, recommendations as to whether items should be retained, augmented, or replaced in order to improve upon process measurement and provide the team with the ability to link quality process measures to quality outcome measures.

Grantee Readiness for DFC

The field of community change and comprehensive community initiatives has long stressed the importance of *"readiness for change"* as a major variable distinguishing successful from unsuccessful coalition efforts (Donnemeyer et al., 1997; Engstrom et al., 2002). While many past measures were qualitative, extremely time consuming, and reserved for initiatives with a small number of participating communities, we have developed quantitative readiness measures that tap into the essential components of readiness and capacity for change (e.g., knowledge, support, expertise, leadership, and resource availability) at both the *collaborative* and *community* levels. In our research efforts, we found

---

[11] Please see our accompanying document on Systems, Measures, and Tools for a complete listing of proposed variables and items.

it extremely useful in identifying and demonstrating the variability that exists among communities in terms of readiness at the beginning of a community change initiatives, as well as documenting the varying trajectories of communities throughout the initiative.[12] This is also the type of measure that provides an early warning sign that some communities may need technical assistance in order to move forward toward their goals of substance use reduction and positive community change. In past research, we also used other quantitative and qualitative data to provide more depth and context in explaining communities' various trajectories in terms of their readiness for change. Finally, ICF has tailored readiness measures to the unique aspects of initiatives – each change effort is unique and while there are common elements to the concept of readiness, there are also unique aspects that need to be captured. In terms of analyses and alignment with the theory of change, grantee readiness for DFC implementation could be modeled as an important "input" variable – something communities bring with them at the start of the initiative. The measure we developed (and can adapt) for DFC collaborative readiness has 11 items (α=.86) while the community readiness measure has 5 items (α=.71).

Shared Vision and Cohesion

Another variable that has been empirically associated with collaborative effectiveness is the importance of establishing a shared vision, which is often one of the first steps in the strategic planning process, as well as a cohesive group that is able to articulate that vision with common language. Based on a search of the extant literature, we previously developed a measure with a small number of items which had adequate reliability (5 items, α=.87).

Perceived Effectiveness

Given the difficulty linking collaborative strategies and efforts directly to community and individual change, much of the past research on collaborative functioning has asked respondents about whether or not they perceived that their efforts made a difference in the issue they were addressing in their community. While it provides perceptual data, such a measure is one more link in the chain tying coalition efforts to community-level and individual-level change. Similar to the current items on collective self efficacy in the CCT, belief in the possibility of change due to coalition efforts is required before action and possible corresponding outcomes will result (Foster-Fishman, Cantillon, Pierce, & Van Egeren, 2007). However, in reviewing the CCT, we believe perceived effectiveness items, tailored to the goals of the DFC Program (e.g., increase protective factors, decrease risk factors, reduce youth substance use rates, increase collaborative capacity) would provide ONDCP with more valid and tailored information regarding coalition efforts. We have created such output measures in the past, tailoring items to the specific aspects of the change initiative, and established high reliability rates.

Sustainability

Another option is to include a more comprehensive measure of sustainability, particularly since it is one of the major elements of the SPF, yet not fully captured in the current CCT. Sustainability items are currently spread out in the CCT and the most direct item simply asks if the coalition chair thinks the

---

[12] For the Children's Bureau Systems of Care Initiative, one of the volumes in ICF's final report assessed how readiness could be conceptualized and measured at varying levels and linked to organizational outcomes (*Volume VI: Readiness for Systems Change: Implementing Systems of Care in Child Welfare*).

collaborative will be around in ten years, along with a checklist of six items if they do not believe the collaborative will still be in existence. Sustainability is an important issue for DFC, and in order to assess this construct comprehensively we previously created a lengthy measure that looked at a number of areas of sustainability (e.g., sustainability of family involvement in the initiative). One of these components was sustainability of interagency collaboration with informal collaborative efforts, as well as the sustainability of a guiding collaborative body in the community post-funding (5 items, α= .74). We plan to work collaboratively with all stakeholders to augment and tailor this sustainability measure to DFC Programmatic efforts and accurately capture what elements communities will be able to sustain once Federal funding ends.

Inter-Organizational Coordination and Systems Change

Along with perceived effectiveness, we view inter-organizational coordination and systems change as two critical outcome variables that are currently not fully captured within the CCT. One of the main goals for DFC is to increase collaboration among community-based agencies, organizations, leaders, and residents. We feel these measures assess (a) whether or not collaboration and coordination has increased and (b) whether this coordination has resulted in meaningful community change corresponding to DFC goals. The inter-organizational measure is short yet reliable (4 items, α=.89) and has been utilized in a number of community-based collaborative change efforts (Allen, 2005; Nowell, 2009).

Contextual Variables

Another missing component from the current version of the CCT is a lack of neighborhood contextual assessments. Given the varied settings of DFC coalitions, we believe more needs to be done to capture the complexity of context, in order to correspond to DFC's logic model and inform analyses. We propose adding variables that could greatly enhance the relevance of analyses as practitioners would be able to answer one of the most important questions regarding generalizability – "Would this work in my community?" Community social organization and collective efficacy are two variables that have been utilized extensively in community-based research and also reflect identified protective factors at the community level.

Finally, we will not limit our assessment of quality to measurement indices, but look for alternative methods to compile this critical information. For instance, quality can also be captured through the use of innovative methodologies that have rarely been applied to coalitions and community change efforts, such as social network analysis (Cross, Dickman, Newman-Gonchar, & Fagan, 2009; Nowell, 2009). This analytic technique allows researchers to understand not only which agencies and individuals interact to a greater degree, but also provide other characteristics, such as the depth of their collaboration and whether or not they are linked to the same players, in a similar deep and meaningful way (e.g., network density). To capture the multi-dimensional nature and quality of collaborative relationships, a number of indicators in network analysis can be included, such as communication frequency, responsiveness to concerns, trust in follow-through, legitimacy, and shared philosophy (Nowell, 2009). This type of analysis provides the necessary data to differentiate what leads to various collaborative outcomes (e.g., coordination outcomes versus community and systems change outcomes). This type of analysis may help identify whether the overall quality and depth of partnerships among key players, community

members, and community-based organizations has reached a tipping point to produce meaningful community change. To decrease overall burden on DFC communities, we propose using this methodological approach with our case study communities[13] using a valued-tie roster questionnaire process (Wasserman & Faust, 1994).

As discussed, we see changes – whether the addition or deletion of measures – to be a collaborative process among ONDCP, the evaluation team, and the Technical Advisory Group (TAG), which includes two DFC funded-coalition members. However, given the vast number of measures, we will initially offer some targeted constructs that could add value to analyses based on our past experience and review of the current CCT. Also, before any new measures are approved, the ICF team will assess response burden for these additional questions, as well as expected response burden for the entire data collection effort. We will aim **to reduce the current response burden**, as we want to ensure that prevention practitioners stay focused on their jobs and not on data collection requirements. We expect that making a good-faith effort to keep response burden down will also result in stronger buy-in for evaluation activities from coalition staff.[14] Thus, while we plan on strengthening process data and attribution between process and outputs and outcomes, overall, we believe we can reduce burden for the staff of DFC coalitions.

In our experience, the assessment and planning phases are precisely where most grantees experience the greatest challenges. By providing measures to detect challenges before they become problems (e.g., lack of cooperation among coalition members), the ICF team will be able to provide evaluation data that ONDCP and its partners can use from the outset of each grant. Details on administering this readiness for change and implementation measure will be a point of discussion in the vetting process for the final measures.

Finally, the team will assess the need for the identification of "critical incidents" that could slow down or even stifle coalition development, such as (1) when changes in leadership took place, (2) when key partnerships were formed or fell apart, and (3) when major initiatives were implemented. Documentation of these incidents will serve two purposes:

1. By understanding what critical incidents took place, the evaluation team can provide context for each year's evaluation results (e.g., "during this year, 29 coalition directorships changed hands").

2. By documenting when these changes took place, we can model the effect of each type of critical incident on outcomes. For example, if major initiatives were implemented, did they have an effect on outcomes, and how long did it take for the incident to influence outcomes?

Together, these efforts will strengthen our evaluation results, and will allow ONDCP to share lessons learned with new grantees. In other words, we expect this effort to result in improved program administration, as well as an improved evaluation.

---

[13] Beginning in year 2, nine communities per year will be selected for case studies. By the end of year 5, we will have data from 36 case study communities.

[14] Response burden is also a data quality issue. We need to ensure that grantees are not so burdened as to be given an incentive to under-report the full extent of their efforts.

Objective 2: Refine Process Data with New Metrics on Coalition Operations

Following the refinement of process measures, we will be in a better position to develop summary metrics on coalition operations. This effort will involve both a review of the CCT, as well as the development of new dimensions to describe coalition operations and functioning. These new metrics can be used as covariates in outcome analyses or they can be descriptive metrics that stand on their own. These metrics are critical because they provide the path to testing more parsimonious, understandable, and powerful models of how coalitions operate and improve. They organize the many indicators of activity into principles or strategies for success.

The Coalition Classification Tool

Early in the study of coalitions promoting public health outcomes, researchers created models of coalition development and maturation (Florin et al., 1993; 2000). This stage-based developmental model is similar to other stage-based models, including the Community Coalition Action Theory (Butterfoss & Kegler, 2002) and the stage-based typology that has largely driven DFC National Evaluation efforts to date. Overall, our past experiences have taught us that coalition development and capacity building is certainly stage-based, and that more "mature" coalitions tend to perform more effectively, but there is not one seminal theory that captures all the complexities of coalitions. The important practical knowledge related to stage-based development will include guidance on why more mature organizations are more effective, and what strategies or actions coalitions can take to reach and sustain higher stages. There is little consensus on answers to these questions, especially when coalitions encounter problems and cycle back a stage due to the continuous challenges that confront complex change efforts.

The existing CCT is in the tradition of these stage-based developmental models. It shares their basic underlying concepts, and to date it has not advanced answers to the why or how questions common to most stage-based typologies. It does, however, differ from prior measurement of stage-based development in its emphasis on basing assessment of stages on the degree to which coalitions have achieved capacity (as perceived by a key informant) in each of the five steps of the SPF across four functional areas. This incorporation of the SPF suggests the possibility of developing more useful guidance concerning what coalitions need to do to move forward. We will carefully assess the CCT to determine if closer analysis of resulting data, or slight revision to the instrument, may help achieve this contribution, which is necessary to the generation of lessons and useful information that is central to our analysis objectives.

We expect to suggest some minor refinements to the CCT; however, we do understand that any changes to this metric may risk the comparability of a coalition's stages of development across years, so certain core metrics will be identified and retained. We also intend to deconstruct the coalition typology into the coalition attributes that contribute to classification, so their separate relation to outcome measures can be assessed. Therefore, we will exercise caution to ensure that our changes do not compromise, but rather facilitate, future analyses.

In summary, we believe that the CCT stage-development model provides a single (yet important) metric characterizing coalition operations. However, it has several potential limitations. As noted in the prior

section, its measurement base is potentially weak because of the reliance on single key informant judgments concerning a broad range of complex strategies, and other factors that have not been sufficiently analyzed and reported to date. Some of the issues that we will address regarding the CCT include:

- *Conceptual Deconstruction of the Global Measurement.* While consistently applying the SPF, the items in the CCT are complex, based largely upon unarticulated hypotheses. For example, many of the empirical referents that are in the item statements are based on the expectation that more formal organization skills or technical skills are indicators of maturation. We will treat these assumptions as hypotheses to be tested. Our experience with coalitions, and lessons from the Community Partnerships and SPF-SIG initiatives, suggest that too much formalism can impede some coalitions in some settings. We will carefully assess the past literature on coalitions and identify those dimensions of strategy and action that are hypothesized to promote maturation, and identify those that have been empirically demonstrated in particular. We will map these onto the CCT measure to identify plausible ways of testing these hypotheses or of modifying the measure to better reflect past lessons.

- *Empirical Deconstruction of the CCT.* The National Evaluation team will conduct exploratory analyses to identify attributes that might feed into a more robust typology of sites. Using dimensional (e.g., factor analysis) and/or cluster analysis (which employs statistical methods to maximize variation between clusters and minimize variation within clusters), our team of analysts will determine whether coalitions can be grouped by a specific set of characteristics. If so, these typologies can also be used as covariates in future analyses. These are exploratory methods which must be assessed for their ability to produce intuitive results. If our clusters are unstable (i.e., if cluster membership changes greatly by adding or subtracting a seemingly minor variable), we will search for other methods to group grantees. For example, in our evaluation of Communities In Schools (CIS), our cluster analyses were not fruitful, so we moved to a measure of fidelity to the CIS model. This implementation metric was used in subsequent outcome analyses, and eventually provided the basis for policy change that defined a minimum set of operations needed for CIS sites to pass national standards.

  Some attempts at confirmatory analysis of the CCT have been performed and reported by the prior contractor of the DFC evaluation. Several distinct models were used to confirm the extent to which coalitions judged by experts to be at one of the four stages of the model could be verified as empirically distinct.[15] Five methods were applied, and none of them achieved a meaningful empirical difference in clustering of CCT items between coalitions judged to be at different stages. In summary, the confirmatory analyses could not identify consistently different configurations of operational characteristics at each developmental stage.

  Our analysis approach will be exploratory rather than confirmatory. We will explore the internal relations of CCT items, and see if they correspond to conceptual interpretation guided by prior research. To date, the National Evaluation team has conducted a simple factor analysis of the 24 items in the CCT staged-development measure. The result was suggestive. Five distinct factors

---

[15] Battelle Institute, *Development of a Classification Rule for the Drug Free Community Evaluation.* Internal project report.

emerged in the analysis. Four of them clearly were aligned according to the four functional areas (e.g., those respondents who rated any one of the SPF capacity items at a certain level of mastery tended to assess all other levels of SPF capacity in that area similarly). A fifth factor loaded most heavily on evaluation capacity items across two functional areas, suggesting that evaluation capacity may be seen as a more distinct activity that develops independent of a specific functional area.

- *External Validation of the CCT.* Another step in our deconstruction of the CCT measure will be associating or correlating scale scores, functional scores, SPF step scores, and individual items with external items or multiple item measures within the CCT, and extracted from COMET, as feasible. This analysis will answer important questions about what referents outside the SPF items may be influencing respondent ratings. In other words, are there outside referents that will help interpret how respondents see greater mastery in the functional or improved capacity in a SPF step?

These analyses will inform collaborative decisions about CCT use and possible revisions as agreed upon by the TAG and ONDCP.

Additional Metrics

We will use our refined typology characteristics, as well as the Stages of Development rubric, to create other overarching metrics that describe the intensity, maturity, and quality of a coalition's operations. By keying in on quality metrics, we will be in a better position to understand why some coalitions are successful and some are not. The metrics upon which we will focus will emphasize measurement of the systemic characteristics of coalitions. A preliminary set of examples includes:

- *Comprehensiveness.* Our experience with coalitions indicates that they vary significantly in the comprehensiveness of the intervention strategies they coordinate, support, or collaborate on. The comprehensiveness of strategy may be a key coalition characteristic linking to characteristics of the community setting, SPF step strategies, other process characteristics, and outcomes.

- *Integration.* Increasing access to multiple components of an intervention strategy has been a documented contributor to community change, and an objective of coalition activity. Measuring the degree to which strategy components are delivered to target groups or places in an integrated system may be an important coalition metric for explaining success.

- *Intensity.* Coalitions differ in the intensity of their intervention exposure in the community. Intensity, or the frequency and amount of intervention exposure, may vary by amount of exposure and by scope of audience. We will assess current DFC National Evaluation data to test the potential value of an intensity (i.e., dosage) measure, and approaches to strengthening a measure if it has analytic potential.

Objective 3: Report Outcomes and Strengthen Attribution between Processes and Outcomes

Our primary focus in this evaluation will be on *helping ONDCP respond to GPRA and PART requirements, as well as strengthening attribution between processes and outcomes*. We therefore must ensure that our evaluation plan has elements of both performance measurement and rigorous evaluation methodology. Specifically, performance measurement must involve all grantees using existing COMET data, while outcome evaluations may involve a subset of grantees using additional rigorous data collection or methods. By bringing together evaluation results from a number of sources to produce findings tailored to different stakeholder audiences, we will maximize both the utility and innovation of our efforts.

We expect the GPRA and PART reporting process to be challenging, but relatively straightforward. We will work in concert with KIT Solutions, the contractor that manages COMET, to provide detailed guidance on any new data sources that are identified by grantees, and to ensure that the data are arriving as clean as possible. Given that ICF already has developed a relationship with KIT Solutions, we expect this coordination process to proceed smoothly. The National Evaluation team will also work directly with KIT Solutions to develop a protocol for the turnover of data from COMET on a periodic basis. The team will hold periodic meetings with KIT to provide feedback from grantees on the system and to develop new reports to meet grantee needs.

As mentioned previously, we will use a ***mixed-method evaluation design***, triangulating both qualitative and quantitative evidence from several efforts to form a meaningful, comprehensive set of evaluation results. Following is a description of the evaluation methodologies we plan to use:

Three quasi-experimental, comparison group designs. It is important for us to form a strong basis of evidence, not only to stay ahead of the curve for GPRA and PART reporting, but also to use in the identification of best practices. As with the National Registry of Evidence-based Programs and Practices (NREPP), the What Works Clearinghouse, or any other repository of best practices, we first need strong research to have confidence in our results. Then, we can make inferences regarding what works and what does not. In its current form, the GPRA and PART reporting would not pass What Works Clearinghouse standards because a credible comparison group is not being used (the old NREPP standards would also invalidate these reports, although they would be acceptable according to new NREPP standards). There are three possibilities for the formation of a comparison group that would ratchet up the level of evidence for the DFC Program and provide opportunities for the identification of best practices:

1. **Recruit a comparison group.** ICF could engage in discussions with new grantees to conduct a quasi-experimental study. We will consider working with a select number of volunteer coalitions in the latest cohort of grantees to identify neighboring jurisdictions that would, ideally, already be collecting data on the core measures.[16] If data are not already being collected on the core measures, the National Evaluation team would have school staff administer a short survey to students. We would then conduct a quasi-experimental study, using extant survey data as a pre-

---

[16] This strategy would require careful understanding of the counterfactual, or other programs/initiatives in place in the neighboring community. The evaluation would focus on what would have happened in the absence of DFC, and the nature of the evaluation questions would not be to determine whether DFC was more effective than having no coalition in place; rather, it would test whether DFC is better than a business-as-usual condition, which is a higher bar to cross, but also a more realistic inquiry into the effectiveness of this grant program.

test measure and follow up with schools to conduct a post-test on an annual basis. To increase participation, we would build in an incentive for each school, and assume that 20 comparison schools will participate in the study.[17] While this will be a relatively small scale study compared to the National Evaluation, it would nonetheless provide us a prime opportunity to ratchet up the strength of our evaluation design. As demonstrated in Exhibit 6, this design would be well-powered to produce meaningful outcomes. Assuming an intraclass correlation of .10, critical *p* value of .05, 200 students per school, and proportion of explained variance by the covariates of .49 (a standard assumption), we can achieve a minimum detectable effect size of .22. For reference, the What Works Clearinghouse defines a "substantively important" effect size to be .25 and above, so the design would be able to detect effects at that threshold level. We view this strategy as a long-shot, given the logistical considerations and the burden to comparison sites. While this is the least feasible of our three options, it is still worth consideration.

EXHIBIT 6. POWER ANALYSIS FOR SCHOOL-LEVEL COMPARISON GROUP DESIGN
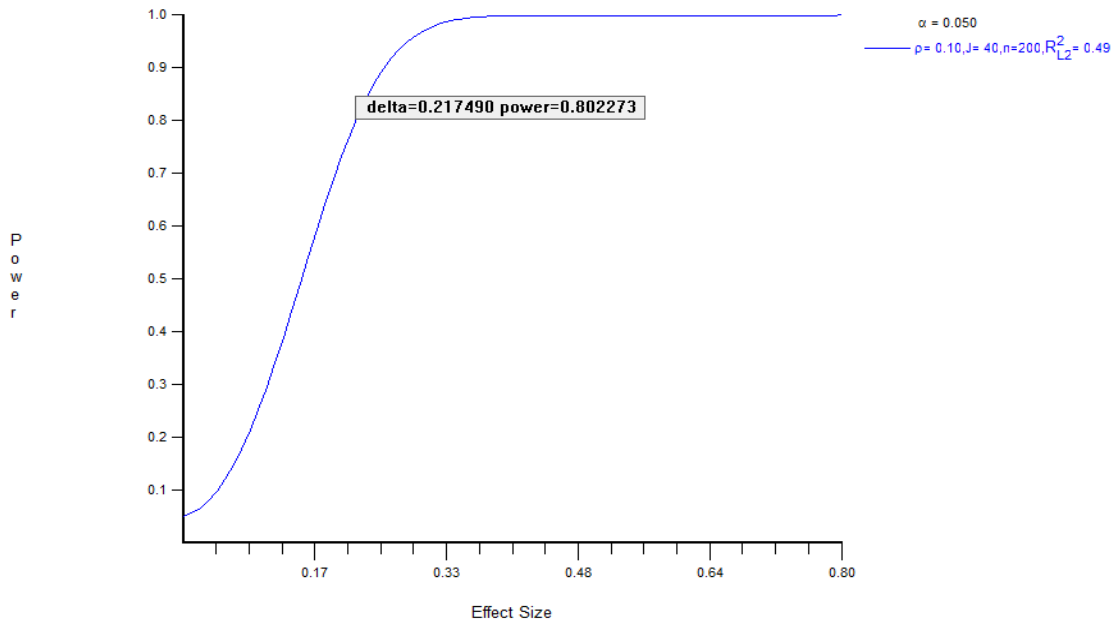
2. **Use advanced matching techniques to develop comparison communities in States where data on the core measures are publicly available.** Some States administer Pride or other surveys to all high schools. We will be able to form comparison communities in a select few States to study whether the core measures have changed across time. Assuming we could get data from Arkansas and New York, we could get more than 50 DFC communities (and many more non-DFC communities) for this study. Accessing Pride Survey data will require negotiations with individual

---

[17] We would like to consider offering participating schools an additional incentive: the development of training on an evidence-based practice of the site's choosing. This incentive would be congruent with the spirit of the evaluation's purpose and would provide a lasting "give back" to the school.

PROPENSITY SCORE MATCHING

Propensity score matching is considered by many researchers to be a cutting-edge technique for identifying comparison groups, and the next best alternative to randomized controlled trials. A propensity score is described by Rosenbaum and Rubin (1983) as "the conditional probability of assignment to a particular treatment given a vector of observed covariates" for the treatment of interest. In other words, rather than matching directly on multiple characteristics, propensity scores allow researchers to match on a large number of factors that are summarized in a single scalar summary variable.

Propensity score matching techniques have been widely used for constructing comparison groups in non-experimental designs. Propensity score matching generally requires the availability of baseline covariate information that is believed to have an important relation both with the groups' assignment to treatment and control conditions, and the outcome. The idea is to replace the collection of confounding variables with a function of these covariates, called the propensity score, which becomes the only confounding variable (Rubin, 1997a). The derived propensity scores can then serve either to aid directly matching of individuals case by case or to stratify individuals.

In the case where individual matching is used, for every individual in the intervention group a matching individual is found from among the comparison group sharing similar characteristics. Popular uses of the propensity score include one-to-one and one-to-many matches that are based on distance-metric methodologies, weighting schemes, or matching on ranges. Alternatively, with propensity score stratification, individuals are sorted into homogeneous subgroups with respect to these propensity scores, and compared in terms of several outcome measures. Based on Cochran's (1968) observation that subclassification in five strata is sufficient to remove at least 90% of the bias, Rosenbaum and Rubin (1983, 1984) recommended a two-phase procedure. In the first phase, a logistic regression is run using a set of covariates to derive a propensity score estimate for the treatment group. In the second phase, treatment subjects are stratified into homogeneous subgroups based on the quintiles of the distribution of the estimated propensity scores, and then matched to comparison subjects.

States, and accessing data from even a few States would provide enough sites for a solid quasi-experimental comparison group design.

Propensity score matching is considered by many researchers to be the next best alternative to randomized controlled trials. This method can determine which comparison community is closest to a given DFC community across a large number of measures. Although we could only conduct this analysis in a few States, it would provide a strong level of evidence of effectiveness for the DFC Program.[18]

3. **Use the Uniform Crime Reports (UCR) combined with Census data to form comparison communities**. UCR data is available at the county level[19] from 1920 forward, and although this data source does not contain information on the core measures, it would allow us to expand our analyses to the unexamined portion of the DFC logic model (behavior & consequences). Using propensity score matching, the National Evaluation team will match DFC grantee communities to non-DFC communities on a number of variables gathered from the UCR and U.S. Census data, including DUIs, possession of illicit substances, single-headed households, percentage of population unemployed, percentage of population on public assistance, residential mobility

---

[18] The weakness of propensity score matching over randomized studies is that we can only develop matches using data at our disposal. While randomized studies would result in roughly equivalent groups on all observable and unobservable factors, quasi-experimental studies can only equate on observable factors.

[19] UCR data reporting is voluntary, so it may not be universally available for all DFC grantee communities. It is also available at the zip code level.

(transient neighborhood), and concentrated neighborhood poverty (above 40%). Using propensity score matching would provide a well-matched comparison group, which could then be tracked across time to assess whether and how DFC coalitions are changing, both before and after receiving a DFC grant. The primary advantage of this method is that the UCR data are publicly available, keeping data collection costs to a minimum. Although we will not have outcomes on the core measures through the UCR, other outcomes such as the rate of DUIs, possession violations, drug sale/manufacture arrests, liquor law violations, and drug abuse violations would be able to provide strong evidence on whether environmental approaches of DFC coalitions are effecting change at the community level. The UCR does have some limitations, however. Since data are reported by county, we can only study coalitions that report their catchment area at a county level. In our experience, DFC grantees tend to overstate their catchment area, so particular caution is needed to ensure that the results of this analysis are valid. Still, by providing a more robust picture of how coalitions are making a difference in communities, we hope to answer previously unaddressed parts of the DFC logic model and ultimately provide more opportunities for the identification of best practices.

Assess information from the respondents of YRBS, MTF, and NSDUH – among other sources – to determine whether DFC results can be aligned to National surveys. While the use of YRBS is an admirable first step in the development of a comparison group, it remains a primary weakness in the prior evaluation. More work is needed to determine whether the sample of students that responded to the various core measures surveys are comparable to the sample that responded to the YRBS. Moreover, because we do not know which specific communities responded to the YRBS, results in some states may be comprised primarily of DFC communities. In essence, *the previous evaluation can be described as a nonequivalent comparison group design*. In order to strengthen these results, a stronger analysis must be conducted to justify using YRBS as a comparison condition. The evaluation team will inquire about gaining access to restricted-use YRBS data in order to determine the feasibility of using YRBS as a comparison.

Identify other outcomes of the DFC National Evaluation. The core measures represent the proverbial tip of the iceberg in representing the true impact of Drug Free Communities grantees – and in addressing the core mission of the grant program. The outcome of environmental strategies pursued by DFC grantees should result in (1) limited access to substances, (2) a change in the culture and context within which decisions about substance use are made, and (3) a reduction in the prevalence of negative consequences associated with substance use (e.g., motor vehicle crashes, sexual assaults, etc.). We would also like to determine whether DFC grantees are having long-term impacts on educational outcomes. It would stand to reason that if DFC grantees are successful in reducing substance use – as initial indications support – then these successes could manifest themselves in a number of ways, including better academic performance or improved behavior.

Capture additional data on what other prevention programs or initiatives are present in DFC communities. The National Evaluation team understands the challenges inherent in evaluating community-level interventions, especially the difficult nature of attributing outcomes to the presence of a single intervention when, in fact, many prevention strategies may be underway simultaneously in a given community. The importance of this potential confound was confirmed in secondary analysis of data from SAMHSA's National Cross-site Evaluation of High Risk Youth Programs, in which adjusting for

differences in the strength of available prevention services across 48 communities increased the average community intervention effect size by .08. In order to strengthen the attribution of outcomes to the presence of DFC coalitions, ICF will undertake an assessment of other prevention initiatives within DFC communities, which is currently reported in COMET. We will also assess whether coalitions are filling out this part of COMET accurately and reliably. While we will keep this analysis at the top level, it will nonetheless provide opportunities to study coalitions while considering the intensity of other initiatives within the community.

Objective 4: Deconstruct Strategies to Identify Best Practices

Once we establish stronger linkages between processes and outcomes, ONDCP can use these results to inform the field about what is working and equally important, what is not. Our ultimate goal at this stage is to determine exactly what makes certain coalitions successful, and why.

There are two basic methods to determine which elements of a particular strategy are associated with success:

1. Construct a matrix of the types of approaches/services offered by each "successful" coalition, and identify the common elements of success among these programs.

2. Implement a "natural variation" design, which capitalizes on the diversity of coalition strategies. This design usually involves logistic regression, to determine which elements "successful" coalitions are more likely to employ compared to less successful coalitions. This methodology has the additional advantage of allowing us to control for key contextual variables, such as stage of development, urbanicity, or levels of use. This design was used in our National Evaluation of Communities In Schools.

We will employ both methods in this evaluation. Where we have a large number of coalitions in the analysis (e.g., in the core measures reports), we can implement a regression-based natural variation study. Where there is a higher level of evidence but a smaller sample (e.g., with our quasi-experimental studies), we will simply determine which key components were associated with stronger outcomes.

Before outcomes are finalized, we will conduct extensive qualitative research to determine whether our quantitative outcomes are supported by the practical experience of coalition staff. The National Evaluation team will identify sites that have innovative processes and positive outcomes (we are assuming 9 site visits per year in years 2-5). These sites will be selected for intensive case studies and will be used to share innovations in the field.

Case studies will provide the research team the opportunity to:

1. Conduct interviews with coalition leadership and determine how successful coalitions achieve positive outcomes.
2. Conduct interviews with coalition partners from a number of agencies to determine best practices in developing healthy collaborative relationships.

3. Conduct focus groups with young people to determine whether particular prevention strategies are resonating with them.

4. Determine what local evaluation data may be available for further research.

5. Determine whether the results of our evaluation are corroborated by the experiences of "front line" staff.

6. Conduct a social network analysis to determine how partners work together and develop a deeper understanding of intra- and inter-organizational relationships (i.e., depth and quality of relationships). Specifically, social network analysis methods will be used to study the interactions between each coalition agency/organization (i.e., to determine which organizations interact more and the nature of interactions) and to study network characteristics, such as centrality, clustering of the most highly interacting players, and gaps in interactions. An index of collaboration will be constructed to indicate strength of collaboration for any one agency or organization, allowing for exploration of the relationship between collaboration and numbers of participants in strategies, types of strategies, and community outcomes.

7. Ensure that our results will be packaged and disseminated in a manner that is useful to DFC grantees.

We feel this approach proves our intention to produce an evaluation and analysis plan that will help improve the performance of coalitions. We will supplement these practical findings on the mechanics of implementing successful strategies (e.g., information dissemination) with strong science in order to ensure that ONDCP and SAMHSA stay ahead of the curve in performance reporting.

Objective 5: Deliver Information on How Coalitions Can Improve Operations and Replicate Best Practices

In the final stage of this evaluation, we will provide coalitions with key information on how to improve operations and replicate best practices. This practical guidance will result in multiple products that can be used by the field to improve operations, and ultimately, improve performance on the core measures. The goal of this stage is to provide "give back" to both Federal and local stakeholders in order to ensure buy-in and to fulfill our end of the deal to provide practical evaluation results.

3. Data Collection and Management

Data collection and management are critical to supporting timely, accurate, and useful analysis in a large, multi-level, multi-site study such as the DFC National Evaluation. This section outlines the multiple data sources, both historical and planned, that are necessary to such a complex undertaking. The overall data set for this project has several unique characteristics that set the context for our data collection and management planning.

- The DFC National Evaluation is ongoing, and a large amount of data has been collected and is available for analysis. As we move forward into the next phase of this evaluation, we will build on this significant body of information to provide continuity and to increase the knowledge return from this valuable data set. Assessment and use of existing data has been discussed at several points in earlier sections.

- The National Evaluation team will be modifying current data collection and adding new measures while controlling burden through eliminating data that has not proven useful, and utilizing existing survey and archival data for the same purposes.

- The DFC National Evaluation database is complex, and draws from several distinct process and outcome data sources. An adequate and efficient overall data management plan must include procedures and markers (e.g., common site IDs) for integrating these data at with sites/coalitions as the unit of analysis. The system must also efficiently support the ability to create integrated analytic data sets in which data from different sources can be integrated for specific analyses.

This section will outline issues and procedures for collecting and managing these data and is organized by data source.

Data Sources

Current and planned sources are briefly summarized along with significant challenges to be addressed in our planning. They are organized by major data source (COMET, CCT, public use data, and case study data).

Response burden is a serious issue in any evaluation. After all, if a coalition is overburdened with data collection, they will lose focus on their core mission of reducing substance use and its consequences among youth. We also believe that additional response burden is only acceptable when it produces data that are manageable, measurable, and most importantly, meaningful. Prior to adding any new data collection, the National Evaluation team will first determine whether needed data are available through public use data files. Obtaining additional data from public use data files, such as the Uniform Crime Reports has two major advantages:

1. It reduces reporting burden on DFC grantees

2. It allows us to build in historical data on coalition effectiveness, which will provide a stronger basis of evidence for our evaluation.

In the absence of public use data, data needs will need to be addressed in the existing COMET system.

Coalition Online Management and Evaluation Tool

The Coalition Online Management and Evaluation Tool (COMET), developed and operated by Kit Solutions, is the primary source of data for the DFC National Evaluation. Data collected through COMET include the Coalition Classification Tool (CCT), the core measures (which are also used for GPRA reports), and process data.[20]

Core Measures

The main focus of this evaluation is on results from the core measures (i.e., 30-day use, perception of risk or harm, perception of parental disapproval, and average age of onset) for alcohol, tobacco, and marijuana. These three substances have previously been the only three on which the DFC National Evaluation collected data. The reasoning for this is that most of the DFC grantees reported these three substances as their greatest concern. The DFC Program guidelines do not force a coalition to plan and implement strategies that address alcohol, tobacco and marijuana. Coalitions can choose the drugs that are indicated as most serious within their own data collection, but they must report data on these three substances, in three grades, every other year.

A. **Past 30-Day Use**—The percentage of respondents who report using alcohol, tobacco, or marijuana at least ONCE in the past 30 days

B. **Age of Onset**—The age that respondents report first trying alcohol, tobacco, or marijuana

C. **Perception of Risk or Harm**—The percentage of respondents who report that regular use of alcohol, tobacco, or marijuana has *moderate risk* or *great risk*

   a. Regular use is defined for *alcohol as one or two drinks of an alcoholic beverage (beer, wine, liquor) nearly every day*.

   b. Regular use is defined for *tobacco as one or more packs of cigarettes a day*.

   c. Regular use is not defined for *marijuana*.

D. **Perception of Parental Disapproval**—The percentage of respondents who report their parents feel regular use of *alcohol* is *wrong* or *very wrong* and the percentage of respondents who report their parents feel ANY use of *cigarettes* or *marijuana* is *wrong* or *very wrong*

DFC grantees demonstrated their ability to collect and report on the core measures in their applications and are expected to upload data for these measures into COMET every two years. The grantees are requested to report data by school grade and gender. The preferred population is school-aged youth in grades 6 through 12. The provision of core measures data once every two years complicates analysis of

---

[20] CCT data are provided by grantees once a year; core measures data are provided once every two years; and process data are required every six months.

data. Since different cohorts of grantees enter data each year, the tracking of outcomes from year to year is impossible. While we would prefer core measures data to be submitted once a year, feedback from our TAG confirmed that this will be too much to ask.

> *Planned Revisions*: One of the major trends in substance use over the past 10 years has been the abuse of prescription drugs and other medications. The broad availability of prescription drugs and misperceptions about their dangers is an alarming combination. We feel that this trend should be tracked as carefully as possible, to the point where this should be a core measure. More specifically, we recommend incorporating the use of prescription drugs as a core substance, especially given that:

- New users of prescription drugs are relatively equal in number to new users of marijuana;

- Teens are abusing prescription drugs because they believe the myth that these drugs provide a medically safe high;

- Painkillers, such as OxyContin and Vicodin, are the most commonly abuse prescription drugs by teens.[21]

While the incorporation of painkillers will add burden to grantees, many existing surveys (e.g., Pride, Monitoring the Future, ADAS) already have questions on 30-day use of prescription drugs; therefore, we do not believe that the burden will be onerous.[22] The incorporation of painkillers into the core measures has other implications, however. For example, the National Evaluation team will have to revise all survey review documents to reflect this new core measure, and new GPRA measures will have to be developed. We feel that the additional burden will be worth tracking this threat to the well-being of our nation's youth.

The other major change to the core measures will involve a change to make the cores measures compliant with SAMHSA's National Outcome Measures (NOMS). The movement to the NOMS represents several major changes in the core measures:

- **30-Day Use**: The NOMS capture the number of days each substance was used in the past 30 days, instead of a simple yes/no framework used in the current core measures. The ability to track frequency of use allows the national evaluation team to determine whether DFC activities result in decreased use as well as higher abstinence rates.

- **Perception of Disapproval**: The NOMS focuses on friends' disapproval of substance use instead of parental disapproval. Because parental disapproval rates are already high, there is little opportunity to demonstrate growth on this measure. Friends' disapproval is not subject to ceiling effects and provides an opportunity to study peer influences on substance use.

- **Perception of Harm**: The NOMS question on perception of harm from alcohol focuses on binge drinking (drinking five or more drinks once or twice a week). The current core

---

[21] Office of National Drug Control Policy (2007). Teens and prescription drugs: An analysis of recent trends on the emerging threat. Washington, DC: Author.

[22] YRBS, however, does not include a question on painkillers.

measure for alcohol asks about regular use of alcohol (one or two drinks per day for five or more days per week), which has been publicized recently as being good for one's health. [23]

Although these changes in core measures will reduce burden for SPF/SIG grantees and other grantees using state surveys that are already NOMS compliant, there may be substantial burden in the transition process to these new core measures.

The National Evaluation team understands that new data reporting requirements at the individual, as well as community coalition levels, may be developed during the course of the evaluation. Our approach to incorporating new measures into an existing evaluation consists of carefully (1) assessing each proposed measure's contribution to understanding the initiative/program, (2) identifying the source of the data for the new measure, (3) testing the new measure to ensure it provides the information desired, and finally, (4) incorporating the new measure into any existing databases. ICF accomplishes each of these objectives through the use of logic models (to ensure the new measure fits and makes sense for the coalition), expert reviews by experts in the field, vetting to DFC coalition members, and consultation with ONDCP to ensure that any new measures address evaluation goals, as well as emerging new requirements.

Process Measures

Results on each element in the Strategic Prevention Framework – Assessment, Capacity, Planning, Implementation, and Evaluation – are collected through COMET to describe the range of strategies conducted by DFC coalitions. Most of the information presented is descriptive in nature, and further work will be needed to validate the content and quality of these data.

One of the key challenges in working with process measures is that grantees are not trained how to categorize some variables (e.g., implementation strategies). The National Evaluation team intends to conduct a validation study to determine whether certain variables need to be recoded. Further detail about our methods is included in the Analysis section (Section 4).

> *Planned Revisions*. The National Evaluation Team has reviewed COMET item-by-item and has made a determination whether each item should be kept, modified, or deleted – or whether new items were warranted. During this screening effort, which is outlined in detail in the Systems, Measures, and Tools document, the team operated on the assumption that if we cannot find a compelling reason to collect a specific piece of data, it should be dropped from the analysis.

---

[23] For a discussion of recent research related to the health benefits of moderate alcohol use, please see: http://www2.potsdam.edu/hansondj/AlcoholAndHealth.html. Much of this research does not take into consideration potential confounds, such as social characteristics of moderate drinkers. For example, moderate drinking may be indicative of an active social life, which is more likely among younger (i.e., more healthy) people, among people who do not have acute mental or physical disabilities, and among people who have a good sense of self-control (which may be correlated with other healthy behaviors).

Coalition Classification Tool (CCT)

The CCT instrument is completed by a representative from each coalition, usually the coalition's paid staff or evaluator. It contains a large number of process measures. However, the major use of this data to date has been for the CCT maturation categories, which were developed to assess a coalition's stage of development. Coalitions proceed through four stages of development (Establishing, Functioning, Maturing, and Sustaining). Coalition maturity is determined by scoring six questions in each of four key functional areas: (1) coalition development and management, (2) coordinating prevention programs/services, (3) implementing environmental strategies, and (4) serving as an intermediary support organization. Coalitions with average scores of 1.00-1.99 were classified as Establishing. Those with average scores of 2.00-2.99 were classified as Functioning. Coalitions with average scores of 3.00-3.99 were classified as Maturing. Those coalitions at the top level–Sustaining–had average scores between 4.00 and 5.00.

> *Planned Revisions*. As discussed in earlier sections of the plan, we will assess the quality and utility of the CCT stages of development classification and recommend changes that balance continuity, as well as increased measurement quality and utility. We will also explore potential use of additional CCT instrument items as process measures.

GPRA Measures

The Government Performance and Results Act (GPRA) was established by Congress in 1993 to engage Federal programs in strategic planning and performance measurement. Federal programs–including DFC–are required by the Office of Management and Budget (OMB) to establish goals, measure performance against those goals, and report results on an annual basis. The DFC currently reports on six GPRA measures:

1. Percent of coalitions that report a decrease in at least one targeted risk factor.

2. Percent of coalitions that report an increase in at least one targeted protective factor.

3. Percent of coalitions reporting at least 5% improvement in past 30-day alcohol, tobacco, and marijuana use in at least one grade.

4. Percent of coalitions that report positive change in youth perception of risk from alcohol, tobacco, or marijuana in at least two grades.

5. Percent of coalitions that report positive change in youth perception of parental disapproval of the use of alcohol, tobacco, or marijuana in at least two grades.

6. Percent of coalitions reporting positive change in age of initiation of alcohol, tobacco, and marijuana in at least one grade.

With the addition of prescription drugs incorporated into the GPRA measures, we feel that coalition performance will be more accurately assessed and valuable information will be available to understand this growing threat.

Public Use Data Files

Public use data files are a major new data source that the National Evaluation team will use to augment current outcome data. Historical data from public use data files can provide community level indicators before and after a coalition's implementation. The use of these public use data files will allow the team to identify comparison communities, which can then be used in rigorous quasi-experimental studies of DFC coalitions' effectiveness. By comparing DFC coalitions to non-coalition communities, we will get a sense of what *would have happened* in the absence of the DFC grant. Methodologically, this will be the most rigorous component of the DFC National Evaluation and will therefore provide the strongest evidence of the grant program's effectiveness.

Pride Surveys has collected data for years at the school level. It may be possible for the National Evaluation team to obtain these data for a limited number of states (e.g., Alabama, Arkansas, and New York are the most likely candidates). Even if we were able to obtain Pride data from only three states, the possibility of conducting quasi-experimental studies in these states (with three very different populations and challenges) would allow us to make unprecedented inferences about the effectiveness of DFC. The California Healthy Kids Survey is another potential source of local prevalence data.

Public use data can also help us determine whether DFC coalitions are changing health and behavioral outcomes, which are represented as long-term outcomes in the DFC logic model. Data from Uniform Crime Reports (UCR), the Fatal Analysis Reporting System (from NHTSA), and other Federal data archives contain consequence data relevant to coalition outcomes and are available at the county level. Whenever coalitions use their county as the catchment area, we could build in an interrupted time series analysis on outcomes such as:

a. Drug Abuse Violations (Total) – Violations of narcotic drug laws.
b. Drug Abuse Sale/ Manufacture – the sale and/ or manufacture of narcotics:
   - Opium/ Cocaine
   - Marijuana
   - Synthetic drugs (Demerol, methadone)
   - Other Sale – dangerous non-narcotic drugs (barbiturates, Benzedrine)
c. Drug Possession:
   - Opium/Cocaine
   - Marijuana
   - Synthetic narcotics
   - Other drugs
d. Curfew Violations
e. Runaways
f. DUIs
g. Liquor Law Violations
h. Drunkenness Charges (this varies greatly between jurisdictions as to whether it is a crime or not).
i. Alcohol-Related Fatal Accidents
j. Alcohol-Related Crashes

These other measures will provide a more robust view of the effectiveness of DFC coalitions. The National Hospital Discharge Survey also appears to have some utility for National Evaluation efforts.

Case Studies

Another important new data source will be qualitative and quantitative data generated through site visits to high performing coalitions. One way of identifying the sample would be to select CADCA's Milestones Award recipients that are DFC grantees and that demonstrated reductions in core measures.

These site visits will produce two valuable data sets. The first will be largely qualitative, using information gathered through interviews, focus groups, and brief surveys that will help attribute coalition processes to outcomes. Substance abuse prevention strategies – and especially environmental approaches – are notoriously difficult to attribute to positive outcomes since we are essentially modeling a non-event. The presence of numerous exogenous factors limits our ability to quantify outcomes with certainty; we also need qualitative data to truly understand what is happening and why. The National Evaluation team plans to conduct extensive on-site data collection in order to strengthen attribution of findings, and to collect data on key considerations in the replication of best practices. Strong measurement of setting, design, and implementation characteristics is crucial to maximizing the learning opportunities in a natural variation design.

The second use of site visit data will use the qualitative richness gathered through semi-structured interviews to code closed-ended measures for each site. This will produce comparable and detailed process variables that will support statistical analysis across sites. ICF team members have extensive experience with this type of coded site visit data, and will develop protocols and data quality assurance procedures.

Steps to Improve Data Quality

Data cleaning is an essential component of any evaluation, as any strong analysis must rest on quality data. In the DFC National Evaluation, we will undertake two concurrent strategies to improve the quality of data provided by grantees: (1) validate and refine data cleaning procedures, and (2) provide technical assistance to grantees.

Data Cleaning Procedures

Currently, the data entered by DFC grantees are cleaned at multiple points:

- A general cleaning process is conducted by Kit Solutions once data are entered into COMET. Cleaning procedures on the data include range checks and other standard techniques to ensure the quality of the data.

- Data are reviewed by SAMHSA project officers for completeness and accuracy. Once data are approved by SAMHSA, they are cleared for release to the National Evaluation team.[24]

---

[24] ICF will explore the option of providing training to SAMHSA project officers on the review of COMET data.

- A more in-depth cleaning process is conducted by the National Evaluation team. This cleaning process takes place in two steps:

  1. Raw data are cleaned and processed using structured query language (SQL) code, then appended to existing "raw" databases. Approximately 22,000 lines of SQL code are applied to this process, and most of the procedures involve logic checks within given databases. ICF has completed an initial review of this code, and the cleaning decisions appear to be in line with standard practice. A second round of review will be conducted in the second year of the contract to document all cleaning decisions and to ensure that the cleaning process is transparent to ONDCP.

  2. The raw data are processed to develop a set of "analysis" databases, which – as the name implies – are used for all analyses. Data cleaning procedures conducted at this step mainly involve logic checks both within and across databases.

- A final round of data cleaning is conducted within the analysis programs. For example, before data are analyzed, duplicate records are removed (duplicates are created when grantees update records from previous reporting periods).

By ensuring that our data are of the highest quality possible, we can have greater confidence in our findings. Given that DFC is implemented through the Executive Office of the President, and is attended to closely by members of Congress, we expect that this evaluation will be subject to scrutiny. Having confidence in our results is therefore of the utmost priority.

Facilitating Data Collection

Much of the data central to this evaluation is collected or provided by grantee organizations or individual respondents within organizations. As noted above, the National Evaluation team will be conducting extensive data quality and missing data bias analyses on existing data. We will identify major challenges in past data collection and develop responses and procedures that will help ameliorate these challenges. Issues will include the following.

Burden

Response burden is a serious issue in any evaluation. After all, if a coalition is overburdened with data collection, they will lose focus on their core mission of reducing substance abuse and its consequences among youth. We also believe that additional response burden is only acceptable when it produces data that are manageable, measurable, and most importantly, meaningful. Prior to adding any new data collection, the National Evaluation team will first determine whether needed data are available through public use data files. Obtaining additional data from public use data files such as the Uniform Crime Reports (UCR) has two major advantages:

1. It reduces reporting burden on DFC grantees

2. It allows us to build in historical data on coalition effectiveness, which will provide a stronger basis of evidence for our evaluation.

In the absence of public use data, data needs will need to be addressed in the existing COMET system.

Technical Assistance to Grantees

One of the key limitations in the current evaluation is that we do not know how grantees sampled youth for their surveys. Since the results of these surveys form the core findings for the DFC National Evaluation, we believe that additional steps need to be taken to ensure the validity of the sampling process, and by extension, the validity of our evaluation results.

Through proactive technical assistance to grantees, the National Evaluation team will provide detailed instructions on how to sample students for outcome surveys. We will work with in-house sampling statisticians and vet our plans to ONDCP before they are sent out to grantees. Our initial thinking is that we should give each coalition a target sample size for their outcome surveys based on the population of the catchment area. For example, given a target population of 10,000 youth, a coalition would need to survey 375 youth to obtain a margin of error of ±5%. We will also emphasize the importance of obtaining a representative sample, although this may be more difficult to codify since all coalitions have different target populations in different settings. Moreover, it is critical that the sampling frame remain consistent so we can accurately measure change across time.

By keeping in contact with grantees, we can also stay up to date on the latest developments in the field, and be in a trusted position to provide guidance on data entry, such as how to classify implementation strategies. We will enhance buy-in for evaluation activities through give-backs, such as policy briefs and practice briefs, and as coalitions see the return on their investment, we believe the result will be better evaluation data.

Data Storage and Protection

DFC data are housed on ICF's servers, and only the analysis team has authorized access to these data. The data collected as part of this evaluation are the property of ONDCP, and data will be handed back to ONDCP or destroyed at their request. In data reporting, the confidentiality of respondents will be protected, and cell sizes of less than 10 will not be reported to further protect respondents from identification. While we consider this a low-risk project from a human subjects protection perspective, we are nonetheless taking strong precautions to ensure that data are not mishandled or misused in any way.

4. Data Analysis

Analyses of the data will be conducted using methods that are appropriate for the design. Comparison group designs will be analyzed using analysis of covariance (ANCOVA), while interrupted time series analyses will employ repeated measures methods. All outcome analyses will include effect size calculations, which will allow us to compare outcomes across sites using a standardized metric. These effect sizes represent the magnitude of given outcomes, which we can then use in correlational studies to determine which prevention strategies are associated with success.

> **EFFECT SIZES**
>
> An effect size is a measure that describes the magnitude of the difference between two groups. Effect size is particularly valuable in best practices research because it represents a standard measure by which all outcomes can be assessed. For example, effect size allows us to compare the size of 30-day use, age of onset, perception of risk, and perception of parental disapproval on the same scale. Effect size is typically calculated by taking the difference in means between two groups and dividing that number by the pooled standard deviation.

Analysis of GPRA Measures

The Government Performance and Results Act (GPRA) was established by Congress in 1993 to facilitate strategic planning and performance measurement. Administered by the Office of Management and Budget (OMB), Federal programs – including DFC – must establish goals, measure program performance, and annually report their progress in meeting goals.

The following performance measures have been used to meet GPRA reporting requirements:

- Percent of coalitions that report a decrease in at least one targeted risk factor.

- Percent of coalitions that report an increase in at least one targeted protective factor.

- Percent of coalitions reporting at least 5% improvement in past 30-day alcohol, tobacco, and marijuana use in at least one grade.

- Percent of coalitions that report positive change in youth perception of risk from alcohol, tobacco, or marijuana in at least two grades.

- Percent of coalitions that report positive change in youth perception of parental disapproval of the use of alcohol, tobacco, or marijuana in at least two grades.

- Percent of coalitions reporting positive change in age of initiation of alcohol, tobacco, and marijuana in at least one grade.

DFC grantees report outcome data for GPRA on a bi-annual basis. GPRA performance measures are based on the current four core measures (30-day use, age of onset, perception of risk or harm, and perception of parental disapproval), and are collected using a variety of survey instruments. Grantees can select from a variety of pre-approved instruments or submit their instruments for approval by the National Evaluation team. Because grantees are only required to enter data on a bi-annual basis, different subsets of coalitions are represented in each performance year.

The evaluation team will assess these GPRA measures to determine their effectiveness in measuring coalition performance. Our initial assessment is that these measures need to be modified significantly. Summarizing results across multiple grades (e.g., positive change in perception of risk across two grades) is misleading because some coalitions report data from three grades (as required) and some coalitions report data on all seven grades (grades 6-12 inclusive). The coalitions that report more data therefore have greater opportunities to show "success".

Our proposed GPRA measures follow:

- Percent of coalitions reporting (over a two year period) improvement in past 30-day:
  o Alcohol use
  o Tobacco use
  o Marijuana use
  o Use of prescription drugs not prescribed to the respondent
- Percent of coalitions that a decrease in at least one targeted risk factor
- Percent of coalitions that report an increase in at least one targeted protective factor

Given significant problems with the Age of Onset measure, we do not believe it is appropriate to be used as a basis for performance measurement. We will use both quantitative and qualitative data to produce our final recommendations for performance measurement. The qualitative assessment will be conducted by talking to ONDCP, SAMHSA, and DFC grantees to determine whether these measures represent clear, unequivocal measures of performance. We will also conduct a quantitative analysis to determine whether other variants of the core measures (e.g., coalition has achieved improvements in both 30-day use and perception of risk) have a stronger linkage to long-term success and sustainability of coalition operations.

Analysis of Core Measures

Our primary impact analyses will be characterized by their simplicity. Given that there are inherent uncertainties in the survey sampling process (e.g., we do not know how each coalition sampled their target population for reporting the core measures, we do not know the exact number of youth served by each coalition), the most logical and transparent method of analyzing the data will be to develop simple averages of each of the core measures. Each average will be weighted by the reported number of respondents. In the case of 30-day use, for example, this will intuitively provide the overall prevalence in 30-day use for all youth surveyed in a given year. The formula for the weighted average is:

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

Where w_i is the weight (in this case, outcome sample size), and x_i is the mean of the ith observation. Simply put, each average is multiplied by the sample size on which it is based, summed, and then divided by the total number of youth sampled across all coalitions.

One key challenge in the weighting process is that some coalitions have reported means and sample sizes from surveys that are partially administered outside the catchment area (e.g., county-wide survey results are reported for a coalition that targets a smaller area within the county). Since means for 30-day use are weighted by their reported sample size, this situation would result in a much higher weight for a coalition that has less valid data (i.e., the number of youth surveyed is greater than the number of youth targeted by the coalition). To correct for this, we will cap each coalition's weight at the number of youth who live within the targeted zip codes. By merging zip codes (catchment areas) reported by coalitions with 2000 Census data, we can determine the maximum possible weight a coalition should have.[25]

To measure the effectiveness of DFC coalitions on the core measures for alcohol, tobacco, marijuana, and prescription drugs, we will conduct three related analyses:

1. *Annual Prevalence Figures*: First, we will compare data on each core measure by year and school level (i.e., middle school [grades 6–8] and high school [grades 9–12]).[26] These results provide a snapshot of DFC grantees' outcomes for each year; however, since coalitions are not required to report core measures each year, they should not be used to interpret how core measures are changing across time.

2. *Gain Scores*: Second, we will calculate the average total change in each coalition, from the first outcome report to the most recent results. By standardizing time points, we are able to measure trajectories of change on core measures across time. This provides the most accurate assessment of whether DFC coalitions are improving or not on the core measures.

3. *Benchmarking Results*: Third, where possible, results will be compared to national-level data from YRBS and Monitoring the Future. These comparisons provide basic evidence to determine what would have happened in the absence of DFC, and allow us to make inferences about the effectiveness of the DFC Program as a whole.

Together, these three analyses provide robust insight into the effectiveness of DFC from a cross-sectional (snapshot), longitudinal (over time), and inferential (comparison) perspective.

Analysis of Process Data

DFC coalitions follow the Strategic Prevention Framework (SPF), which is built upon a community-based risk and protective factors approach to prevention and a series of guiding principles that can be utilized at the Federal, State/Tribal and community levels. For the past five years, grantees have been reporting a wealth of process data corresponding to each step in the SPF (Assessment, Capacity, Planning, Implementation, and Evaluation – supported by cultural competence and sustainability at each step).

---

[25] 2010 Census data will be released starting in April 2011, with all data released by September 2013. ICF will update catchment area data as these new data come in.

[26] Coalitions were asked to report data by school level and gender; however, given that only nine coalitions have reported results exclusively by gender (out of 731 coalitions that reported on 30-day use) – and sample sizes were much larger for school-level breakouts – we do not believe that presenting data by gender will add significantly to our understanding of trends in overall prevalence figures. We will, however, present patterns in results by gender when they are notable.

Much of the data collected to date has been largely untapped, and the exploration of process data is one of the more exciting opportunities for the next phase of the evaluation. As a first step in the process of analyzing these data, work will be undertaken to assess their validity.

The most common type of validity assessment will involve the linkage of free-form text responses to standardized response categories. For example, DFC grantees are asked to describe their implementation strategies and then, link that activity to one of Seven Strategies for Community Change[27]:

1. Provide information

2. Enhance skills

3. Provide support

4. Enhance access/reduce barriers (or reduce access/enhance barriers)

5. Change consequences

6. Change physical design

7. Modify/change policies

While these categories are sufficiently detailed to facilitate analyses, they may not be mutually exclusive in some cases (e.g., students caught using drugs have to attend after-school classes on substance abuse, which would both alter consequences and provide information), and strategies may not be categorized correctly in others. Moreover, some strategies may cross over multiple steps in the SPF (e.g., needs assessment strategies can be found in all five steps of the SPF).

To obtain relevant categories based on historical DFC data, we **randomly selected two hundred cases** for each SPF step. This procedure increased our confidence that we were obtaining data on both new and older grant communities, as well as coalitions along the entire range of functioning (high/low). Responses were then coded and major categories were retained. Another coder then used these dropdowns to categorize an additional 100 randomly selected responses. This quality assurance check was conducted to ensure that all major categories were included within each SPF section.   Then, reviewers edited the categories to remove redundancy among categories, to make sure those core competencies that facilitate implementation of SAMHSA's Strategic Prevention Framework (SPF) were included, to make sure identification of best processes was included,[28] and to eliminate categories that did not fit the SPF section.

Once process data are determined to be of sufficient quality, they will be used in "natural variation" studies to determine what "high performers" are doing differently than others. For example, we have found that coalitions with the strongest reductions in 30-day use had engaged in more Support Activities

---

[27] Community Anti-Drug Coalitions of America (2009). Handbook for Community Anti-drug Coalitions. Retrieved 2/16/10 from http://www.cadca.org/ and originally from the University of Kansas Work Group on Health Promotion and Community Development—a World Health Organization Collaborating Centre.

[28] Identification of best collaborative processes was identified by a University of Kansas workgroup who conducted a thorough review of the literature.

(e.g., alternative activities, mentoring, referrals, etc.). If this finding holds up with clean process data, it could form the basis of policy decisions on the DFC Program or be used as a criterion in the selection of new grantees.

Impact Analyses

Given that public use data may be widely available in particular locations, but not others (e.g., we may be able to obtain Pride Survey data for all communities in Alabama and Arkansas), we have the opportunity to conduct a number of rigorous regional/State studies using comparison group designs. These studies will have the strongest internal validity of any part of the National Evaluation, and they will help us triangulate findings and enhance the external validity (i.e., generalizability) of our results from other study components.

In complex evaluations such as DFC, this approach can be particularly valuable since it stands to reason that results from the National Evaluation may not be broadly applicable to all coalitions. As outlined in the DFC logic model, context matters, and by understanding at a more granular level what works and what does not, results can be packaged and tailored to specific types of coalitions.

Using propensity score matching to develop closely-matched comparison groups at baseline, we will study DFC communities and matched comparison (non-DFC) communities[29] from the year prior to DFC grant award to five years after implementation. This timeframe will provide sufficient opportunity for us to observe long-term outcomes and to understand the dynamics of

> **KEY QUANTITATIVE AND ANALYTIC TECHNIQUES TO BE APPLIED IN THE DFC OUTCOME EVALUATION**
>
> - **Calculating Weighted Means**: In the analyses of the core measures, the National Evaluation team will weight reported means by the size of the reported sample that took the survey. This allows us to determine the overall prevalence of 30-day use for all youth surveyed.
>
> - **Analysis of Covariance (ANCOVA)**: ANCOVA is used to determine whether the means of two groups are significantly different at post-test while controlling for pretest differences. This method will be applied in our quasi-experimental studies.
>
> - **Structural Equation Modeling: (SEM)**: SEM is a method to estimate causal relationships between a number of variables. It is often used to estimate relationships between latent variables (i.e., variables that are not measured directly, such as happiness). SEM is particularly valuable as a method to validate logic models and theories of change. We will use SEM in this study to validate the DFC logic model and understand the interrelationships between difficult-to-measure concepts that are critical to the success of coalitions.
>
> - **Logistic Regression**: Logistic regression is a multivariate method to estimate the probability of a dichotomous outcome (e.g., an outcome whose value is either 1 or 0). Logistic regression will be used in our natural variation studies, since we will be trying to determine what factors are more likely to be present in high-performing coalitions (coded as a "1") relative to other coalitions (coded as "0").
>
> - **Cluster analysis**: Cluster analysis employs statistical methods to maximize variation between clusters and minimize variation within clusters. This exploratory method is particularly valuable for developing typologies of sites by showing the dimensions on which some coalitions naturally cluster together.
>
> - **Cost Analyses**: Cost analyses will be conducted to determine the best practices for the price. We will investigate both start-up costs and maintenance costs for each best practice, and will gather in-depth cost data during our site visits. This information will be shared in Best Practices Briefs, so other coalitions will know how to replicate effective strategies.

change among multiple variables across time. For example, a given quasi-experimental study could help us understand which outcomes (e.g., perception of risk or perception of parental disapproval) precede others. Such a study could also provide valuable information about the amount of change in 30-day use

---

[29] Matched comparison communities would be identified where public use data is available from a number of surrounding communities (e.g., Pride data in Arkansas). The rule of thumb for propensity score matching is that we need a 4:1 ratio of potential comparison subjects to treatment subjects to ensure a balanced match.

that can be expected in the first year, and the pattern of change (e.g., do reductions in 30-day use follow a general downward pattern, or is there a significant drop within the first three years?). By understanding not only the amount of change--but also the dynamics of that change–we will be in a position to better understand exactly how coalitions are changing communities.

Analysis of these data will use methods most appropriate for the data structure and the types of research questions we are asking. Long-term impacts will be assessed using analysis of covariance (ANCOVA), and repeated measures analysis. Simple trend plots will also be developed to understand the dynamics of change across time.

Geographic Information Systems. The National Evaluation team includes staff who are highly trained in spatial analysis. By incorporating Geographic Information Systems (GIS) into our analyses, we will be able to map neighborhood assets and problems that are related to alcohol and other drug use. This information will not only provide the field with additional considerations in their identification of risk factors, it will also provide key contextual information for our analyses. Recently, there has been a number of interesting research projects that have correlated the number of liquor stores to alcohol consumption. This evaluation will provide us the opportunity to build upon those analyses and map neighborhood assets (parks, libraries, etc.) and problems (liquor stores, abandoned buildings, etc.) to augment our quantitatively derived "neighborhood typology".

Subgroup Analyses. We will also draw upon historical data and conduct time series designs to see whether "critical incidents" had an effect on outcomes. By deconstructing coalition strategies, as well as the contexts in which they operate, we can break new ground in this evaluation. For example, subgroup analyses can be conducted on any of the following dimensions (either on their own or in combination with each other):

- Maturity level
- Typology category
- Urbanicity
- Readiness for change
- Size of coalition membership
- Age of coalition
- Tenure of leadership
- Region of the country (e.g., Midwest vs. Northeast)
- School level
- Socioeconomic status of target population
- Intensity of service delivery (e.g., environmental strategies)
- Risk/protective factors present in the community
- Existing prevalence of 30-day use or perceptions of risk/disapproval
- Budget
- Neighborhood type

For each dimension listed above, we will be able to explore its correlation and contribution to positive outcomes using our natural variation methodology.

Analyzing Grantee Feedback from Technical Assistance Activities

Technical assistance for the DFC National Evaluation has been designed to accomplish two major objectives: (1) increase the reliability and validity of the data collected from coalition grantees through various technical assistance approaches; and (2) provide "give backs" (i.e., Best Practice Briefs, Policy Briefs, Evaluation Summary Results) to grantees for their use in performance improvement, advocacy work, and to support material for sustainability planning. Analysis of previous evaluation data and interviews with past evaluators revealed that grantees did not have a common frame of reference to define the evaluation data elements they were required to enter. This indicates that DFC grantees need assistance understanding the evaluation process, thus increasing their interest in the DFC National Evaluation process. Additionally, by providing grantees with "give backs" that they can use throughout the course of the evaluation, we increase their likelihood of providing meaningful, valid, and reliable data during data collection.

The Technical Assistance Team will achieve the first objective by working with the Evaluation Team to draft clear and concise definitions for all data elements to be collected. Consequently, coalition grantees will have uniform information for data elements when they are entering data into the COMET system. To further increase the quality of the data collected from grantees, an Evaluation Technical Assistance Hotline (toll-free phone number) and email address have been established. Technical Assistance Specialists provide responsive evaluation support to grantees as questions arise when they are entering the required data. Grantees' queries are logged and analyzed to develop topics for on-line technical assistance webinars.

The second objective is designed to produce materials that grantees will find useful in their everyday operations, stakeholder briefings, and when they apply for funding for future coalition operations. The Technical Assistance Team will work with the Evaluation Team to refine the format and content for "give back" materials (e.g., Best Practice Briefs, Policy Briefs, etc.) to best suit their needs and then provide grantees with these materials at various points throughout the evaluation. This was also the subject of a focus group discussion at the CADCA Mid-year Training Institute in July 2010.

Overall, these technical assistance activities help to ensure buy-in for evaluation activities, reduce response burden, improve response rates, and ultimately, improve the quality of the data along with providing grantees with evaluation data they can make use of in strengthening their prevention strategies and securing additional funding.

Analysis of the COMET System

ICF will determine and recommend how to best manage the data currently stored in COMET, along with the CCT data. Concurrently, we will also determine and recommend how best to collect and store future evaluation data. Future data collection might use an updated version of the present system, or there could be a complete re-design and a new replacement system. We see these analyses as crucial to future success. We also believe that they must include the spectrum of project skills and expertise. In other words, analysis of the existing systems needs to include the IT and Web usability perspective, but must not be totally driven by it. Nor should an evaluation from the evaluation and programmatic

perspectives ignore the IT findings. Also we need to get the actual end-user perspectives on these questions as well. Our goal will be to determine the best way forward after considering every angle.

Our analysis of COMET will include reviewing the effectiveness of validation procedures for submitted data files, analyzing procedures for data edits, anomalies, and any other related data problems, and evaluating how COMET can more effectively address the National Evaluation needs. We would also like to study opportunities for much broader use of automated data cleaning at the server level, although this may depend on implementation of a replacement data system. If new elements are added to DFC data collection, we will provide guidance and business rules on incorporating the elements into COMET.

As part of this analysis, we will prepare a matrix showing each data element presently collected by COMET, showing those intended to be continued going forward, those intended not to be continued, and new data elements proposed to be added.

We will evaluate the extent to which COMET is meeting DFC and Federal Government needs by conducting online focus groups and feedback sessions, and interviews with DFC grantees and other key stakeholders. If the project schedule allows, we will identify and work with a recently awarded DFC grantee to evaluate their initial experience and usage of COMET. Our Technical Advisory Group may also be engaged. Our Web team will evaluate the existing COMET interface and functionality by conducting a heuristic evaluation, a usability review, usability testing, and a Section 508 accessibility review. We also intend to design and conduct a COMET burden analysis from the grantee perspective.

Based on the results of this task, we may propose designing and building a new data collection and reporting system to replace COMET. Our determination will be based on our in-depth analysis and evaluation of the effectiveness and COMET's ability to meet the growing needs of the DFC Program. If ONDCP agrees that a new system should be built, we will build upon our existing analysis by gathering system requirements with all appropriate stakeholders. The new system will be built to accommodate the varying technical abilities of the diverse populations for which it is intended. Currently, the grantee data system is separate from the evaluation data sets and databases. This requires data to be transferred between physical contractor locations. The replacement system has the potential to alleviate the need for separate data collection and evaluation systems and databases.

The replacement system will provide for National Evaluation project data analysis, querying, and reporting. We would also like for a replacement system to provide more "give back" value to the DFC grantees in terms of reports they are able to generate on their own. We believe that availability of well-crafted data reports is a key factor in coalition sustainability. A detailed report of our findings is expected to be released to ONDCP in January 2011.

Research Synthesis

One of the largest challenges in the conduct of multi-component, mixed method evaluations is determining how to put all the pieces together to arrive at a consistent and powerful set of findings that can be used to inform both policy and practice.

At the most basic level, the DFC National Evaluation is designed to answer four overarching questions:

1. Does the DFC Program work (i.e., does the program result in better outcomes, as defined by the core measures)?

2. How and why does the DFC Program work (i.e., what are the key factors needed to ensure that a coalition is effective)?

3. In what situations does the DFC Program work better than others (i.e., are there certain settings or types of communities that are inherently more likely to achieve success)?

4. What are best practices and policies for DFC coalitions (i.e., what specific strategies, policies, and practices maximize chances of success)?

Exhibit 7 lays out all study components described in this document, and which components will contribute to answering each overarching question. Our preliminary plans to synthesize study components rely on the quality and amount of data that can be brought to bear to answer each overarching question. Assuming we have quality data, we can answer the most important overarching question (Does the DFC Program work?) by synthesizing four study components:

1. The grantee-reported outcome data on the core measures will be used to track trends and prevalence figures among all DFC grantees. Because this is the only outcome data at our disposal that covers all DFC grantees, it will be the central focus of our impact analyses.

2. Benchmarking to National surveys, such as Monitoring the Future (MTF), YRBS, and Pride Surveys, will provide a basis of comparison for DFC to national-level prevalence figures. DFC grantees cover a wide swath of the country, and we can quite easily make the argument that DFC covers a representative sample of youth in the United States. Given that, a statistically significant difference between DFC and YRBS (for example) would provide a clear indication that grantees are effective. The problem with these comparisons, however, is that we oftentimes do not know which communities were sampled; therefore, a given survey could cover only DFC communities – or it could only cover non-DFC communities. Because we cannot separate the DFC from non-DFC prevalence figures in these surveys, it is best to simply call these comparisons exploratory in nature. Still, they will be used to describe whether DFC grantees are producing results in line with National trends, or whether they are over- or under-performing relative to National averages. In addition, the National Evaluation team will conduct a comparative analysis of National surveys, such as the Behavioral Risk Factor Surveillance System (BRFSS), Youth Risk Behavior Survey (YRBS), Pride, National Survey on Drug Use and Health (NSDUH), Monitoring the Future (MTF), American Drug and Alcohol Survey (ADAS), and other widely administered surveys. This will provide a better understanding of the biases inherent in each survey and the strength of the inferences we can make by comparing DFC results to these sources.

3. A series of state- or regional-level quasi-experimental studies will provide the most rigorous test of impacts for DFC grantees. By developing closely matched comparison groups at baseline and then tracking results across time, we will be in the position to make stronger inferences about the effectiveness of DFC on the core measures.

4. The GPRA analyses involve wrapping up outcome data (e.g., percent of coalitions reporting improvement in past 30-day alcohol, tobacco, marijuana, and prescription drug use). These

frameworks for reporting will be used in quasi-experimental studies where possible to both validate the measures and to provide a stronger basis of evidence for the measures.

The synthesis of these impact analyses will not only help us determine whether DFC grantees are making a difference at a given point in time, but also whether trends on the core measures are moving more strongly than the nation as a whole. These findings will be strengthened by quasi-experimental studies, which will allow us to make stronger statements about the effectiveness of the DFC Program.

The next two overarching questions (How does DFC work? and In what situations does DFC work the best?) will be anchored by our Natural Variation Study, which describes what "high performing " coalitions are doing differently from others. Across all coalitions, we will have extensive process data, and a typology of coalitions that will provide important covariates or subgroups for our analyses. By linking these processes and typologies to outcomes, we can exploit the natural variation in coalition operations to determine what works best in given situations.[30] The use of quasi-experimental studies will corroborate these findings in a number of settings, which will add to the generalizability of results. Site visits will provide a strong mixed-method component to the evaluation that will greatly enhance inductive learning from the experience of select, accomplished coalitions; help identify robust best practices with strong external validity (e.g., they work across diverse environments); and provide grounded interpretation of results. Site visit data collection will support (a) developing comparable site-level variables to support meta-analysis of the relation between measured site characteristics and measures of effect, (b) social network analysis to explain the interpersonal and organizational dynamics of coalitions, and (c) case studies will also help us both corroborate our findings and describe specific settings in which some strategies work better than others. These studies will be done at a more granular level, but what we lose in generalizability will be made up in terms of the specificity of our findings. This level of detail produced by the evaluation team will be highly valuable for practitioners looking to implement modifications to their prevention strategies, either from a service or policy context. Critical incidents analyses will allow us to understand the impact of key attenuating circumstances (e.g., change in leadership) on outcomes – and also whether the combination of circumstances (e.g., change in leadership combined with the loss of a key partner) has multiplier effects.

The final overarching question (What are best practices/policies for DFC coalitions?) will be developed based on the results of our intensive case studies (i.e., site visits to coalitions). These case studies will allow us the opportunity to determine how best practices/policies can be replicated, and also the opportunity to collect cost and sustainability data to determine what the best practices are *for the price*.[31] Process data reported through COMET will allow us to determine which coalitions are engaged in such best practices, which will allow us to more carefully observe outcome trajectories of these coalitions to ensure that results are holding up across time.

---

[30] This can be an incredibly complex inquiry since context will differ in each coalition. By using multivariate analyses to control for a set of potential moderating variables, we hope to isolate the conditions under which certain strategies work best. Answering complex questions such as this also require extensive qualitative data.

[31] We will have to consider both the investment of DFC grant funds into the community and the coalition's ability to leverage other funding.

| EXHIBIT 7. STUDY COMPONENTS DESIGNED TO ANSWER OVERARCHING EVALUATION QUESTIONS				
Study Component	**Does DFC Work?**	**How and Why Does DFC Work?**	**In What Situations Does DFC Work Better?**	**What Are Best Practices and Policies for DFC Coalitions?**
Grantee Reported Outcome Data (biannual reports)	☐			
Benchmarking to National Surveys	☐			
Natural Variation Study		☐	☐	
Grantee Reported Process Data (COMET)		☐		☐
Coalition Classification Tool/ Typology of Coalitions		☐	☐	
Quasi-Experimental Comparison Group Studies	☐	☐	☐	
GPRA Analyses	☐			
Critical Incidents Analysis		☐	☐	
Case Studies		☐	☐	☐
Geographic Information Systems (GIS) Analysis			☐	
Social Network Analysis		☐		☐
Sustainability Study				☐
Cost Study				☐

Ultimately, the exact strategies needed to synthesize results from each study component will depend upon our results and the quality of the data that we can obtain. Our goal is to "tell a story" about how DFC coalitions are working and to synthesize findings in such a manner as to be useful and actionable for both policymakers and practitioners.

Structural equation models (SEM), which can be used to validate the DFC logic model, will provide the most comprehensive synthesis of study components. SEM allows us to improve the documentation of the steps that occur from DFC strategies through the distal outcomes of reduced substance use, improved health and behavioral outcomes, and coalition sustainability. The main advantage of SEM over multivariate regression and other analytic procedures is the fact that SEM partials out measurement error, allowing for increased power to detect significant relationships between independent variables and dependent variables. This analytic procedure allows researchers to model which variables predict other variables in a chain like fashion (much like a logic model) that ultimately impact output and outcome variables. ***Thus, this procedure is well-suited to deconstruct the chain of events required to produce an ultimate impact as graphically displayed in logic models***. Exhibit 8 displays a sample of an SEM model that demonstrates how we can assess the various linkages between community and collaborative level variables, as well as how they impact outcomes. In the sample model, collaborative perceived effectiveness is the dependent variable, but there are innumerable other outputs and

outcomes. We can assess how collaborative variables and dynamics differentially impact certain outcomes and produce meaningful results for practitioners and policy makers. Other possible dependent variables include: (1) number of community strategies; (2) depth of community strategies; (3) coalition sustainability; and (4) member satisfaction with the coalition.

We will use Mplus software to conduct our SEM analyses since this statistical software utilizes full-information maximum likelihood (FIML) estimation to obtain values for missing data. Based on our past experience surveying large numbers of coalitions, there is potential for missing data and we need to anticipate and be able to correct for this. FIML does not impute data, but rather uses all the available raw data to estimate any given parameter (Arbuckle, 1996). In many applications of this approach, correct maximum likelihood estimation with missing data can be obtained under mildly restrictive assumptions concerning the missing data mechanism (Rubin, 1976). This approach will increase the power to detect significant differences above and beyond other missing data imputation procedures. Moreover, Mplus allows for the modeling of dichotomous process and outcome variables, which may include many of our "checklist" items.

EXHIBIT 8. SAMPLE STRUCTURAL EQUATION MODEL
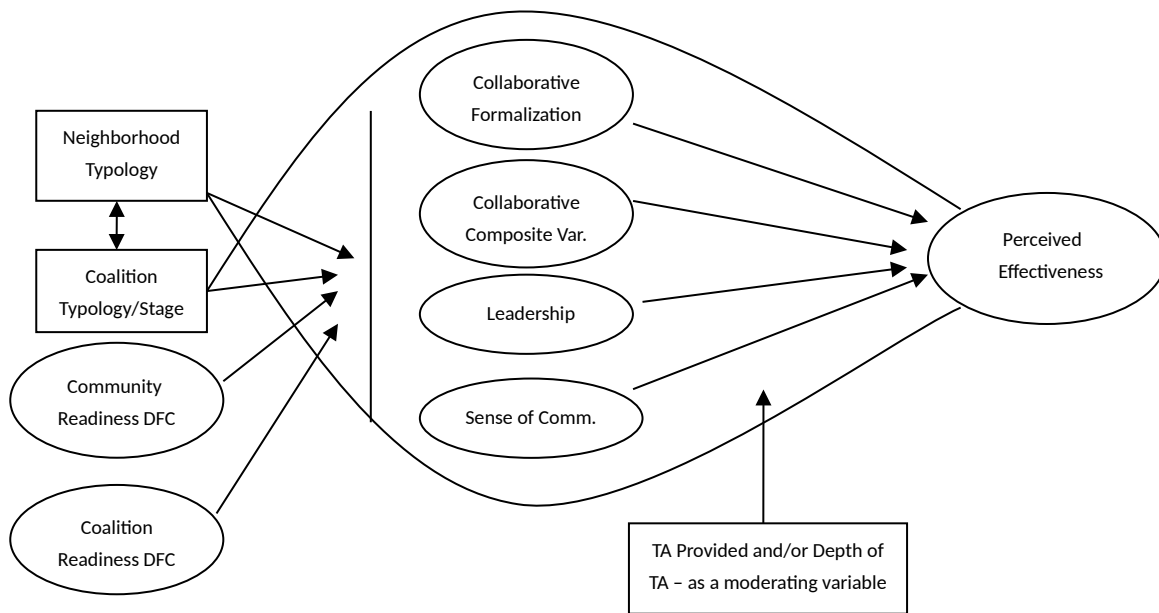
Challenges and Limitations

A number of challenges and limitations exist due to the structure of the grant requirements, the nature of the evaluation, and the availability of data. Each challenge is described below, along with a brief description of how each given challenge can be overcome:

- **We are not confident about whether the core measures are reported for a representative sample within each coalition**. DFC grantees are asked to report data on the core measures

every two years; however, very little guidance has been provided on sampling plans. We are not certain whether each coalition is providing a representative sample or whether they are "creaming" the results. In the next round of the National Evaluation, we will provide additional guidance on sampling and provide grantees with a target sample size and sampling procedures for their youth surveys. Although we will not be able to guarantee the delivery of data that are representative of the coalition at large, we will still provide guidance to grantees to make sure samples are as representative as possible.

- **Core measures are reported every two years, which makes interpretation of year-to-year change difficult**. The National Evaluation team will conduct cohort studies to understand whether the group of coalitions reporting in even years is substantively different from the coalitions reporting in odd years. Other grantees report data for every year, which adds to the complexity. This contextual information will allow us to understand whether year-to-year fluctuations represent positive or negative movements in results.

- **There are no public use data files reported at the community level that can be used to develop a comparison group on the core measures.** Because we cannot develop a comparison group for every coalition using national-level data, we will have to exploit pockets of similar data (e.g., Arkansas Pride data) that can be used to develop smaller, yet rigorous, impact analyses. The triangulation of these smaller studies will provide a wealth of information for practitioners and policymakers – and answer practical questions, such as "In what settings do DFC coalitions work better than others?"

- **Response burden needs to be kept to a minimum**. Our data collection plan calls for a net reduction in reporting burden. Although we have limited "evaluation capital"[32] at our disposal, we believe that reducing reporting burden will actually add to the quality of the evaluation data and overall, we will have more findings to share with confidence. It may seem paradoxical that less data collection will result in more findings, but in our experience, that pattern has held across many of our studies.

- **Difficulty in linking coalition strategies to community-level changes.** Attribution is a significant challenge in this evaluation since DFC grants focus on developing an infrastructure to reduce substance use in the community; direct service provision is not intended to be the primary focus of DFC grantees. It is certainly difficult to attribute lower rates of substance use to the presence of better lighting in a public park; however, because we are conducting a number of separate studies, the triangulation and replication process inherent in our study design will increase our ability to attribute processes to outcomes. We will also develop structural models to link processes to outcomes.

---

[32] Evaluation capital refers to the amount of burden we can impose on grantees before those burdens ultimately result in lower quality data – or less cooperation with evaluation staff.

5. Evaluation Products

With such a large number of stakeholders in this evaluation, the National Evaluation team will need to develop a number of evaluation products. Our plan includes a dissemination strategy that will ensure that coalitions get both a "give back" for their data collection efforts and practical guidance for implementing best practices. Anticipated products include:

A. **Best Practices Briefs** will summarize best practices and will provide information on (a) the extent of evidence underlying the practice, (b) qualitative evidence from staff who have implemented the practice, (c) key considerations in the replication of best practices [i.e., helpful hints gathered from coalition staff], and (d) a summary of the costs involved in replicating the practice. Cost-effectiveness results will go beyond answering not only what works best, but rather what works best for the price. This will provide much more practical guidance for the field when decisions are made about adopting best practices. The National Evaluation team will collect detailed cost information on identified best practices during site visits using a structured protocol. This protocol will be vetted to key decision makers prior to its use to ensure that all appropriate cost centers and considerations are captured. Please see the text box on this page for a sample outline of a Best Practices Brief.

> **SAMPLE OUTLINE OF A BEST PRACTICES BRIEF**
>
> 1. **Introduction to Best Practices Briefs**
>   a. Layout of the document
>   b. How to use best practices briefs
> 2. **Overview of the Best Practice**
>   a. Data supporting best practice (impacts found)
>   b. Overview of our level of confidence in the data
>   c. Detailed description of the best practice
>   i. Overview of the practice
>   ii. How practice is implemented in coalition
>   iii. Number of students/parents/staff engaged in practice
>   d. Theories/other research supporting best practice
> 3. **Cost**
>   a. Estimated implementation costs
>   b. Estimated maintenance costs
>   c. Comparison of costs to other strategies
> 4. **Communication**
>   a. Tips for how to communicate the need for this best practice to policymakers
>   b. Tips for the types of questions that policymakers will ask practitioners regarding the practice.
> 5. **Contacts/Resources**
>   a. Contact information of grantees who can provide advice on implementing the best practice
>   b. Further reading/resources on best practices
> 6. **Optional: One page fact sheet that can be used in discussions with policymakers/funders**

B. **Policy Briefs**, which will be similar in scope to best practice briefs, but they will be tailored to policymakers. Policy context will be included in lieu of helpful hints for replication.

C. **Interim and Final Reports** are the core products of our evaluation. Our reports are typically structured to distill complex evaluation methods and results into easily accessible and practical findings for practitioners.

D. A **Sustainability Study** that will share critical information with DFC grantees about preparing for sustainability of coalition initiatives and outcomes. Shortly prior to the end of each grantee's DFC grant, the National Evaluation team will administer an online survey that will ask coalitions

to identify (a) whether they are sustaining operations, (b) what funding, if any, they have received, and (c) best practices for sustainability. The results of this survey (which will be administered by phone if we do not receive a response online) will be shared with current grantees and will ensure that the seed money provided by ONDCP is spent wisely.

E. *Web Content* will provide grantees with additional findings and information to improve practice.

Prior to the development of any products (especially the practice briefs and policy briefs), the National Evaluation team will meet with ONDCP and its partners (e.g., SAMHSA, CADCA, etc.) to ensure there is no duplication in our efforts to provide information to grantees. We will also vet products with our Technical Advisory Group, which is comprised of grantees, researchers, and experts with on-the-ground experience, to ensure that they meet the highest standards of quality and provide the most practical results possible.

The upshot of these evaluation activities will be a stronger evidence base, along with more practical information for coalitions. Ultimately, we feel that this approach dovetails well with the needs of the grantees, as well as the mission of ONDCP and SAMHSA.

REFERENCES

Allen, N. (2005). A multilevel analysis of community coordinating councils. *American Journal of Community Psychology*, *35* (1/2), 49-63.

Beyers, J.M., Bates, J.E., Pettit, G.S., & Dodge, K.A. (2003). Neighborhood structure, parenting processes, and the development of youths' externalizing behaviors: A multilevel analysis. *American Journal of Community Psychology*, *31*(1/2), 35-53.

Brooks-Gunn, J., Duncan, G.J., Klebanov, P.K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, *99*(2), 353-395.

Brounstein, P. & Zweig, J. (1999). *Understanding Substance Abuse Prevention Toward the 21st Century: A Primer on Effective Programs*. Washington, DC: U.S. Department of Health and Human Services.

Colder, C.R., Mott, J., Levy, S., & Flay, B. (2000). The relation of perceived neighborhood danges to childhool aggression: A test of mediating mechanisms. *American journal of Community Psychology*, *28*(1), 83-94.

Cross, J.E., Dickman, E., Newman-Gonchar, R., Fagan, J.M. (2009). Using mixed-method design and network analysis to measure development of interagency collaboration. *American journal of Evaluation (30)*3, 310-329.

Crowel, R. C., Cantillon, D., & Bossard, N. (April, 2009). *Building and sustaining systems change in child welfare: Lessons learned from the field*. Symposium presentation at the 17th annual National child Care Conference on Child Abuse and Neglect, Atlanta, GA.

Donnermeyer, J.F., Plested, B.A., Edwardes, R.W., Oetting, G., & Littlethunder, L. (1997). Community readiness and prevention programs. *Journal of the Community Development Society*, *28* (1), 65-83.

Drug Free Communities Support Program Policy Report (2003). Caliber Associates.

Elliott, D.S., Wilson, W.J., Huzinga, D., Sampson, R.J., Elliot, A., & Rankin, B. (1996). The effects of neighborhood disadvantage on adolescent development. *Journal of Research in Crime and Delinquency*, *33*(4), 389-426.

Engstrom, M., Jason, L.A., Townsend, S.M., Pokorny, S.B., & Curie, C.J. (2002). Community readiness for prevention: Applying stage theory to multi-community interventions. *Journal of Prevention and Intervention in the Community*, *24*(1), 29-46.

Florin, P., Mitchell, R., & Stevenson, J. (1993). Identifying training and technical assistance needs in community coalitions: A developmental approach. *Health Education Research*, *8*, 417-432.

Florin, P., Mitchell, R., & Stevenson, J. (2000). Predicting intermediate outcomes for prevention organizations: a developmental perspective. *Evaluation & Program Planning*, *23*, 341-346.

Foster-Fishman, P.G., Cantillon, D., Pierce, S.J., & Van Egeren, L.A. (2007). Building an active citizenry: The role of neighborhood problems, readiness, and capacity for change. *American Journal of Community Psychology*, *39*, 91-106.

Gruenewald, P.J. (1997). Analysis Approaches to Community Evaluation. *Evaluation Review*, *21*(2), 209-230.

Granner, M.L., & Sharpe, P.A. (2004). Evaluating community coalition characteristics and functioning: A summary of measurement tools. *Health Education Research*, *19*(5), 514-532.

Kakocs, R.C., & Edwards, E.M. (2006). What explains community coalition effectiveness?: A review of the literature. *American Journal of Preventive Medicine*, *30*(4), 351-361.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). McGraw Hill, New York.

Porowski, A., Landy, L., & Robinson, K. (2004). *Key outcomes and methodological strategies employed in the Drug-Free Communities Support Program*. Presented at the Annual Meeting of the Society for Prevention Research, Quebec City, Canada.

Roussos, S.T., & Fawcett, S.B. (2000). A review of collaborative partnerships as a strategy for improving community health. *Annual Review of Public Health*, *21*, 369-402.

Sampson, R.J., Morenoff, J.D., & & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions for research. *Annual Review of Sociology*, *28*, 443-478.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Influence*. Boston: Houghton Mifflin company.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

Watson-Thompson, J., Fawcett, S.B., & Schulz, J. A. (2008). Differential effects of strategic planning on community coalitions in two urban neighborhoods. American Journal of Community Psychology 42, 25-38.

Wolf, T. (2001). Community Coalition Building – Contemporary Practice and Research: Introduction. *American Journal of Community Psychology*, *29*(2), 165-172.