Skip all front matter: Jump to Page 1

TECHNICAL
REPORT

Analysis of the Cities Readiness Initiative

Christopher Nelson, Andrew M. Parker,
Shoshana R. Shelton, Edward W. Chan,
Francesca Pillemer

Prepared for the U.S. Centers for Disease Control and Prevention

RAND HEALTH

This work was sponsored by the U.S. Centers for Disease Control and Prevention. The research was conducted in RAND Health, a unit of the RAND Corporation.

Library of Congress Cataloging-in-Publication Data

Analysis of the Cities Readiness Initiative / Christopher Nelson ... [et al.].
p. ; cm.
Includes bibliographical references.
ISBN 978-0-8330-6029-7 (pbk. : alk. paper)
I. Nelson, Christopher D. II. Rand Corporation.
[DNLM: 1. Cities Readiness Initiative. 2. Disaster Planning—organization & administration. 3. Efficiency, Organizational. 4. Program Evaluation. WA 295]

362.1—dc23

2012002978

Published 2012 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

The Cities Readiness Initiative (CRI) provides funding, program guidance, and technical assistance to improve communities' ability to rapidly provide life-saving medications in response to a large-scale bioterrorist attack, naturally occurring disease outbreak, or other public health emergency.

In 2010, the Centers for Disease Control and Prevention (CDC) asked the RAND Corporation to conduct an analysis of CRI program data collected by the CDC in order to assess (1) the current status of communities' operational capability to meet the CRI program goal of delivering medical countermeasures to entire Metropolitan Statistical Areas within 48 hours of the federal decision to deploy assets and (2) whether there is evidence that CRI has improved communities' capability to meet the 48-hour goal. The study follows up on an earlier evaluation of CRI conducted by RAND during 2007 and 2008.

The results of the study suggest that grantee's response capacities (i.e., plans, equipment, personnel, partner agreements, and protocols) are strong. Significant growth in capacities suggests that CRI has had an impact, but data on operational capabilities are not conclusive. The report concludes with recommendations for continued enhancements to CDC's ability to measure readiness and program impact.

This report should be of interest to those seeking to understand the operations of public health preparedness and homeland security programs, as well as to those interested in developing feasible approaches to evaluating these programs' effectiveness. The study was carried out between 2010 and 2011 within the RAND Public Health Systems and Preparedness initiative. RAND Health is a division of the RAND Corporation. A profile of the initiative, abstracts of its publications, and ordering information can be found at http://www.rand.org/health/centers/public-health-systems-and-preparedness.html.

Contents

Figures

Tables

Summary

The Cities Readiness Initiative (CRI) provides funding, program guidance, and technical assistance to improve communities' ability to rapidly provide life-saving medications in response to a large-scale bioterrorist attack, naturally occurring disease outbreak, or other public health emergency. Currently, the program operates in each of the 50 states and involves the participation of local "planning jurisdictions" (which have diverse structures, ranging from single public health departments to multiple municipalities working together) in 72 of the nation's largest metropolitan areas. These areas correspond roughly to the federally defined Metropolitan Statistical Areas (MSAs).

In 2010, the Centers for Disease Control and Prevention (CDC) asked the RAND Corporation to conduct an analysis of CRI program data collected by the CDC over the course of the program in order to assess

- the current status of communities' operational capability to meet the CRI program goal of delivering medical countermeasures to entire MSAs within 48 hours of the federal decision to deploy assets
- whether there is evidence that CRI has improved communities' capability to meet the 48-hour goal.

The analysis focused on both *capacities* (i.e., plans, equipment, personnel, partner agreements, and protocols) and operational *capabilities* (i.e., the ability use capacities in real-life operational contexts). At CDC's request, the study relied, where possible, on existing data to minimize the need to burden program participants with new data collection requirements. Capacities were assessed using data from a standardized written assessment tool—the Technical Assistance Review (TAR)—and capabilities were assessed from self-reported data derived from operational drill performance. These sources were supplemented by discussions with a small number of stakeholders in participating jurisdictions.

Capacity as Measured by TAR Scores Appears to Be Strong

The TAR measures the completion of a weighted composite of critical planning tasks identified by CDC. There are two closely related versions of the TAR: one for state health departments (the State Technical Assistance Review, or S-TAR) and the other for local planning jurisdictions in the participating MSAs (the Local Technical Assistance Review, or L-TAR).

State capacities. As of 2009–2010, all states' overall scores—the average of all 13 functional areas, weighted by each function's importance—were equal to or above the 79-percent

threshold deemed acceptable, with the average state scoring 94 out of 100. Although performance was strong across all functional areas, performance was somewhat lower in three particularly critical areas: coordination and guidance for dispensing,[1] security, and distribution.

MSA capacities. As of 2009–2010, planning jurisdictions in the average MSA achieved a score of 86 out of 100, with a median of 89 percent. (Note that there is no official threshold for the L-TAR.) However, there was more variability in local scores, which are aggregated into an MSA score, than among state scores. Performance was lower in the critical areas of training, exercise, and evaluation; security; and dispensing. MSAs in higher-scoring states with centralized public health systems performed better on the 2009–2010 TAR compared with MSAs in lower-scoring states and in those in less-centralized public health systems, after controlling for other factors.

Operational Testing Has Not Been Conducted at the Scale Required to Test Readiness for the 48-Hour Scenario

Planning jurisdictions conducted and reported data on 1,364 drills in 2008–2009 and on 1,422 drills 2009–2010. However, few jurisdictions have tested their capabilities at a large scale. For example, a large number of jurisdictions have tested staff call-down procedures. However, in 2009–2010, nearly 90 percent of these tests involved 100 or fewer people, thus limiting efforts to estimate the capability to contact all needed staff during a large-scale emergency. Similarly, in 2009–2010, only 32 percent of drills that tested dispensing at points of dispensing (PODs) involved 500 clients or more. POD drills with higher numbers of clients reported higher throughputs, suggesting that, if jurisdictions run more large-scale drills that place more stress on PODs, greater countermeasure dispensing capability might be revealed.

Significant Growth in Capacities Suggests That CRI Has Had an Impact, but the Data Are Not Conclusive

State TAR scores have improved consistently (the median increased from 85 in 2006–2007 to 95 in 2009–2010),[2] and the variation among states' scores has been reduced. MSA-level TAR scores showed similar patterns, with the median increasing from 52 in 2006–2007 to 89 in 2009–2010. But more variability remained among MSAs' performance than among states' performance. There was also anecdotal evidence both that CRI has improved responses to real incidents and of spillover effects in the form of states using the TAR (and similar instruments) to assess non-CRI communities.

The fact that greater "exposure" to CRI is associated with considerable increases in TAR scores is consistent with CRI having an effect on preparedness. However, the absence of data from a representative comparison group makes it difficult to rule out the possibility that other factors drove the increases. Thus, the findings must be regarded as suggestive but not conclusive.

---

1 Local planning jurisdictions are mainly responsible for operating dispensing sites.

2 The CRI program operates on a budget year that has typically been from August of one year to an end date in August of the following year. Thus, the period 2006–2007 to 2009–2010 is four years.

Implications and Recommendations

The recommendations presented in this report focus on improving systems for measuring, improving capacities and capabilities at the local and state levels, and enhancing accountability to decisionmakers and the public.

Recommendation 1: Attempt to Validate the TAR

Given heavy reliance on the TAR as a measure of CRI readiness, it is important to (1) assess the extent to which TAR scores represent actual variations in communities' preparedness and (2) confirm that the TAR scores assigned to different states and planning jurisdictions are truly comparable.

Recommendation 2: Continue Refining the Drill-Based Measures by Requiring Jurisdictions to Conduct Drills at a More Realistic Size and Scale

In keeping with Homeland Security Exercise and Evaluation Program guidelines, CDC encourages jurisdictions to conduct increasingly difficult drills that lead to full-scale exercises. Perhaps because of the burdens associated with conducting large exercises, many of the drills are conducted at a smaller scale than would be required by the CRI scenario. Requiring that at least some call-down drills call the entire POD volunteer list, and that PODs in at least some dispensing drills more closely resemble those that would be implemented in a CRI scenario (in terms of their procedures, size, staffing, and throughput), would lead to more-realistic assessments of jurisdictions' capabilities.

Recommendation 3: Improve Performance Feedback to Jurisdictions and Develop Stronger Improvement Tools

Several stakeholders perceived the need for additional performance feedback to jurisdictions in convenient, easy-to-use formats and for tools that would further assist them in closing the performance gaps revealed through the TAR and drill-based measures. CDC should consider reviewing its current tools and feedback procedures in order to better understand the extent and sources of this perceived deficiency and should, as necessary, promote or revise existing tools and develop new ones.

Recommendation 4: Seek to Leverage Assessments of Non-CRI Sites as a Comparison Group

CDC should consider efforts to encourage states to collect data on a broader range of non-CRI communities to support systematic performance comparisons between CRI and non-CRI sites.

Recommendation 5: Assess Cost-Effectiveness

In the future, it would be useful to assess the program's cost-effectiveness (i.e., costs relative to benefits found by this and other studies) in order to inform discussions about whether the program's accomplishments justify the investments made.

Acknowledgments

The study could not have been completed without the efforts of a large number of individuals. First and foremost, we thank the many officials at the state and local levels who gave generously of their time to talk with us and provided essential program documentation. We also acknowledge Deborah Loveys, Rae Benedict, Linda Neff, Elizabeth Ely, Paul Renard, Jaime Jones, and other CDC staff who provided invaluable support and assistance throughout the project. We thank Sarah Hauer and Suzanne Benedict for their expert administrative assistance and Peter Breed for his substantive input. Finally, we thank Henry Willis, Lisle Hites, Claude Setodji, and Jeanne Ringel for their in-depth reviews of earlier drafts of this manuscript.

Abbreviations

ANOVA	analysis of variance
CDC	Centers for Disease Control and Prevention
CRI	Cities Readiness Initiative
DSNS	Division of Strategic National Stockpile
DUR	drug utilization review
EOC	emergency operations center
FY	fiscal year
HHS	U.S. Department of Health and Human Services
LAIV	live attenuated influenza vaccine
L-TAR	Local Technical Assistance Review
MSA	Metropolitan Statistical Area
N	number
NA	not applicable
PHEP	public health emergency preparedness
PIC	public information and communication
POD	point of dispensing
RSS	receipt, store, and stage
SNS	Strategic National Stockpile
S-TAR	State Technical Assistance Review
std dev	standard deviation
TAR	Technical Assistance Review
WIC	women, infants, and children

CHAPTER ONE

Introduction

The Cities Readiness Initiative (CRI), arguably the nation's flagship public health preparedness program, was established by the U.S. Department of Health and Human Services (HHS) in 2004. The program is part of the Centers for Disease Control and Prevention's (CDC's) Cooperative Agreement on Public Health Emergency Preparedness (PHEP), which provides funding to 50 states, four directly funded cities, eight territories, islands, and freely associated states to improve the ability to prevent, respond to, and recover from public health emergencies. Through a combination of funding, program guidance, and technical assistance, CRI seeks to improve the ability of the nation's largest metropolitan regions to rapidly provide life-saving medications in the event of a large-scale bioterrorist attack, naturally occurring disease outbreak, or other public health emergency. The program builds on existing efforts (also funded through the PHEP Cooperative Agreement) to support the federal Strategic National Stockpile (SNS), which maintains caches of medications and other medical supplies that can be requested by states to supplement state and local supplies, when needed.

Participating metropolitan areas have been selected for the program largely on the basis of size. The program was originally rolled out to 21 cities in 2004, and it was soon expanded to include the entire Metropolitan Statistical Area (MSA) for each of the original cities, as well as to an additional 15 MSAs.[1] The first 21 MSAs are known as Cohort I, and the additional 15 MSAs as Cohort II. In 2007, the program was further expanded with the addition of 36 Cohort III MSAs (see Appendix A). Currently, the program includes 72 of the nation's largest metropolitan regions (with at least one in every state) and individual "planning jurisdictions" (which have diverse structures, ranging from single public health departments to multiple municipalities working together) within those regions. Overall, CRI covers approximately 57 percent of the U.S. population.

The program requires participating regions to prepare to deliver antibiotics or other medical countermeasures to their entire populations (often millions of people) within 48 hours of the federal decision to deploy. Between 2004 and 2010, $301.2 million was allocated to the program (Centers for Disease Control and Prevention, 2010). Appendix A provides additional information on the CRI program.

[1] Originally, the MSA boundaries were prescribed by the Office of Management and Budget. As of 2010, five MSAs participating in CRI had requested, and were allowed, to add new populations/areas to their catchment areas.

Purpose of the Study

In 2010, CDC asked the RAND Corporation to conduct an analysis of CRI program data collected by CDC in order to assess

- the current status of communities' operational capability to meet the CRI program goal of delivering medical countermeasures to entire MSAs within 48 hours of the federal decision to deploy SNS assets
- whether there is evidence that CRI has improved communities' capability to meet the 48-hour goal.

The report follows up on an earlier evaluation of CRI conducted during 2007 and 2008. Based on in-depth case studies of seven participating and two nonparticipating MSAs, and on quantitative program assessment data from 2004 and 2005–2006, the earlier study found that CRI had improved preparedness by increasing the number of staff working on countermeasure dispensing, strengthening key partnerships with other responders, supporting the development of more-detailed plans and streamlined dispensing models, and enabling the purchase of critical equipment and supplies. The study concluded that CRI's funding, focus on a clear planning scenario, and accountability requirements were important drivers but that the program's effectiveness varied according to degree of centralization, the quality of participants' relationships with state health departments, and staff turnover rates.

Organization of This Report

Chapter Two describes the data and methods used to conduct the analysis, with technical detail on statistical and qualitative analysis provided in the report appendixes. Chapter Three addresses the current status of readiness, summarizing findings on capacity (i.e., the quality of planning). Chapter Four also addresses the current status of readiness, summarizing the drill-based assessments of operational capability. Chapter Five assesses whether the CRI program has been effective in bringing about improvements in readiness. Chapter Six summarizes the report's key findings, describes their implications for policy, and provides recommendations.

CHAPTER TWO

Approach

This chapter describes the conceptual framework that guided the study and discusses primary data sources and methods used. Technical detail on data sources and methods is provided in the report appendixes.

The Conceptual Framework Highlights Capacities and Capabilities

The analysis used a conceptual framework, or logic model (Chen, 1994), to orient data collection and analysis. The framework identifies a plausible—though not validated—chain of events leading from the components of the CRI program (e.g., funding, planning scenarios, assessment/accountability, technical assistance) to the desired outcome of reducing morbidity and mortality through mass countermeasure dispensing operations.[1] Although some of the specific capabilities fostered by CRI were put to the test in the recent H1N1 influenza outbreak, there have been no incidents requiring countermeasure delivery to an entire metropolitan region within a timeline as short as 48 hours. The framework helps identify plausible proxies or predictors of readiness to respond to such a scenario.

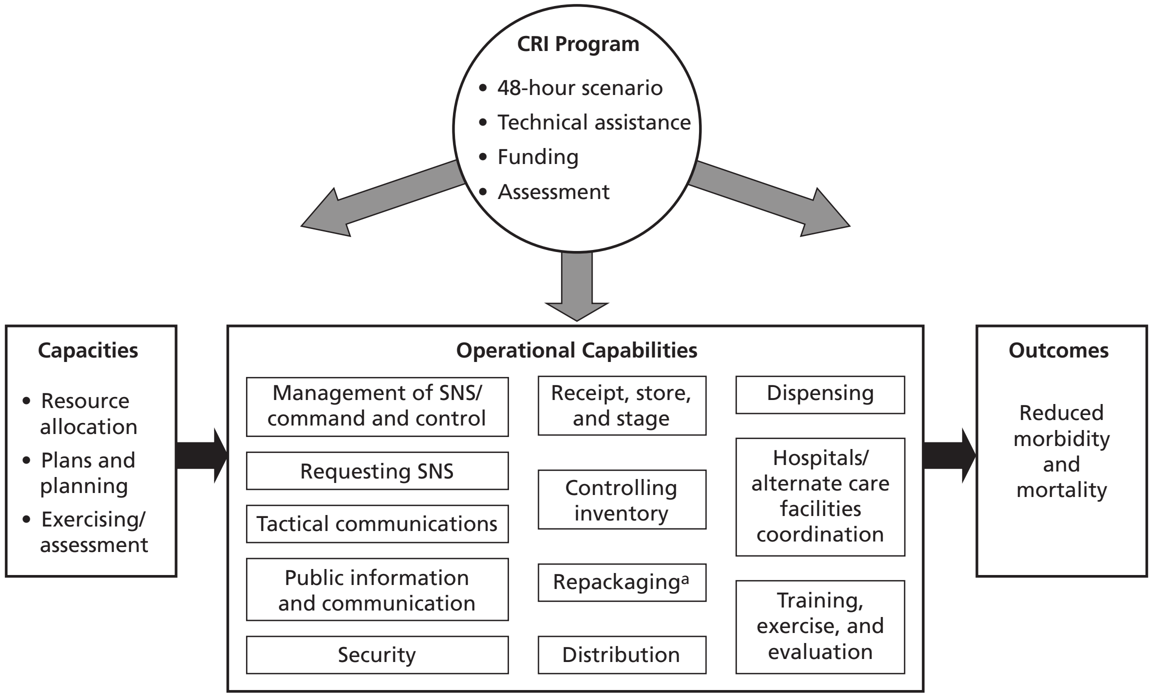
The framework, illustrated in Figure 2.1, includes *capacities* and operational *capabilities*. The term *capacities* refers to the resources and processes involved in readiness, including resource allocation, response plans, planning processes, and exercising, assessing, and improving readiness. Simply having a plan and resources is no guarantee of the ability to respond. However, resources, plans, and partnerships are generally thought to provide a necessary foundation for mounting an effective response.[2] By contrast, the term *operational capabilities* refers to the ability to put resources and plans into practice in real-life operational contexts. The capabilities addressed by CRI (and SNS) include requesting, warehousing, securing, distributing, and dispensing SNS materiel; mobilizing staff and facilities; and implementing effective public information campaigns.

---

1 For simplicity's sake, a number of intermediate outcomes (or outputs) that come between the execution of operational capabilities and ultimate health outcomes are omitted. These include (but are not limited to) throughput (patients processed per hour), dispensing accuracy (ensuring that individuals receive the correct medication), and rate of adverse reactions.

2 We are aware of no empirical evidence on the relationship between planning and execution in PHEP. However, Hallmark and Crowley (1997) provide empirical evidence that increases in the quality of planning are related to increases in levels of execution for Army tank companies conducting operational exercises at the National Training Center at Fort Irwin, California.

Figure 2.1
Conceptual Framework



[a]Applies only to states.
RAND *TR1200-2.1*

The CDC's Technical Assistance Review Provides the Best Available Representation of Capacities

The Technical Assistance Review (TAR) provides the only comprehensive assessment of capacities for medical countermeasure delivery (e.g., plans, memoranda of understanding, personnel, equipment, training, exercising). Designed by CDC, the TAR assesses capacities in 12 or 13 functional areas:[3]

1. developing an SNS plan
2. management of SNS/command and control
3. requesting SNS
4. tactical communications
5. public information and communication
6. security
7. receipt, store, and stage (RSS)
8. controlling inventory
9. repackaging
10. distribution
11. dispensing

[3] The repackaging function applies to states but not to local jurisdictions.

12. hospitals/alternate care facilities coordination
13. training, exercise, and evaluation.

There are separate, but largely similar, versions for assessing state health departments and local planning jurisdictions participating in the CRI program. CDC collected all TAR data that were reviewed during the present study.

For each area, the TAR assesses whether personnel have been assigned to perform key activities, whether community partners have been identified and assigned roles, whether plans specify how key activities will be performed, and whether protocols have been established to guide the dispensing of medications to the public. Specific planning tasks within each TAR functional area are scored by CDC program services consultants or state SNS planning officials based on scoring guidance, site visits, a review of plans, and a review of supporting activities documents.

The overall TAR score, which ranges from 0 to 100, is a weighted composite of the degree to which the assessed jurisdiction has completed critical planning tasks. Although some of the items assess whether jurisdictions have exercised components of their CRI plan, the TAR does not directly assess the operational capability to execute plans. A copy of the 2009–2010 local TAR instrument is available online.[4]

The CDC's Drill-Based Assessments Provide the Best Available Representation of Operational Capabilities

In 2008–2009, CDC rolled out a set of drill-based metrics to test the operational readiness of a small number of selected tasks related to mass countermeasure delivery.[5] The drills focus on the following tasks: staff call-down, site activation, facility setup, pick-list generation, and dispensing. The drills can be conducted as stand-alone assessments or in combination with one another as part of a larger exercise. CDC collected all data on the drill-based metrics and provided them to RAND. Appendix C provides additional technical detail on the analysis of the drill datasets.

Discussions with Stakeholders from Participating MSAs Provided Additional Information on Performance and Program Impact

Analysis of quantitative data (e.g., TAR scores, drill data) was supplemented by a small number of discussions with stakeholders in nine MSAs participating in the CRI program. These data facilitated exploration of some of the specific mechanisms through which CRI might improve preparedness, and they helped illustrate performance strengths and gaps in concrete terms.

---

[4] See Centers for Disease Control and Prevention, Division of Strategic National Stockpile, undated-a. The TAR has been adapted and refined over the years. The tool on this website has not been verified as being in alignment with the current, official SNS tool.

[5] Additional details on the drills are available at Centers for Disease Control and Prevention, Division of Strategic National Stockpile, undated-b. RAND developed prototypes of the metrics and associated data collection tools, initially under contract with the HHS Office of the Assistant Secretary for Preparedness and Response (see Nelson, Chan, et al., 2009). The measures were designed to align with the DHS Target Capabilities.

Data were collected through a combination of telephone and in-person discussions with state and local CRI officials. Sites were selected to represent variations in TAR score growth patterns.

The number of these discussions was limited in order to minimize the need to burden program participants with new data collection requirements. CDC staff were not present during the case study discussions and did not limit or review the questions asked. Appendix D provides additional technical detail on the case studies and sites chosen.

CHAPTER THREE

Evidence of Current CRI Capacities

Chapters Three and Four assess the current status of communities' readiness to meet the CRI program goal of full-community prophylaxis within 48 hours of the federal decision to deploy assets. This chapter focuses on *capacities*—the plans, equipment, personnel, partner agreements, and protocols needed to carry out this mission. The next chapter considers evidence of those communities' operational capability to put these capacities into action during an actual response.

Scores on the Technical Assistance Review (TAR) were used to assess capacities, after first taking into consideration the percentage of sites completing the TAR. The latter establishes the completeness of the data and provides insight into completion of one program reporting requirement. To supplement discussion of the TAR, case study discussions were used to explore some of the results in more detail, and statistical modeling was used to assess whether variations in TAR scores followed systematic patterns. Appendix B provides detail on data and methods.

To understand the findings in this chapter, it is important to note that, other than the directly funded cities, states are the initial recipients of the SNS and have overall responsibility for initial distribution of SNS materiel. Local planning jurisdictions may or may not engage in distribution from local or regional warehouses. (This function may be administered completely by a state.) Conversely, local planning jurisdictions are mainly responsible for operating dispensing sites, with states providing coordination and guidance. These differences are reflected in the state and local versions of the TAR.[1]

TAR Assessment Completion Rates Are High

Completion of the TAR indicates the existence of a plan and participation in the CRI program, and high reporting rates lend confidence in interpretation of TAR scores. As noted in Chapter Two, there are two closely related versions of the TAR: one for state health departments (the State Technical Assistance Review, or S-TAR) and the other for local planning jurisdictions in the participating MSAs (the Local Technical Assistance Review, or L-TAR). According to the PHEP Cooperative Agreement, all 50 states must complete the S-TAR each year. Similarly, all local planning jurisdictions that are part of participating MSAs are assessed with the L-TAR, 25 percent by CDC officials and the remaining 75 percent by the state.

[1] It is also important to note that 29 states and 130 planning jurisdictions (across 35 MSAs) with total scores of at least 90 percent in 2008–2009 were allowed to carry over those scores to 2009–2010 without reassessment.

Since 2007–2008, 100 percent of states have been assessed annually with the S-TAR. L-TAR completion rates among planning jurisdictions have consistently been at or above 97 percent within all cohorts since 2007–2008, as shown in Table 3.1.

Table 3.1
Percentage of Planning Jurisdiction TARs Completed, by Year and Cohort

Cohort	2007–2008	2008–2009	2009–2010
I	99	98	99
II	100	99	100
III	98	97	100

NOTE: The maximum number of missing TARs in any given year and cohort was four.

State TAR Scores Are Above Threshold, and They Vary Across Functional Areas

Overall state performance on the S-TAR is assessed in relation to a threshold score. Initially, the threshold score for acceptable performance on the state TAR was set at 79 out of 100; for 2010–2011, the threshold was raised to 89. This study applied the earlier threshold of 79, given that it examines TAR data collected prior to 2010–2011. A score of 79 indicates that a jurisdiction has adequately completed most of the tasks that CDC identified as necessary for meeting minimal capacity for plan readiness.[2]

Table 3.2 presents overall state scores and scores for each of the 13 functional areas described earlier. As of 2009–2010, all states' overall scores—the average of all 13 functional areas, weighted by each function's importance—were equal to or above the 79-percent threshold, with the average state scoring 94 out of 100.

There was, however, some variation among states in function-level scores, especially in the areas of public information and communication, security, repackaging, hospitals/alternate care facilities coordination, and coordination and guidance for dispensing.[3] Keeping in mind the relative importance of functions (as indicated by the importance weights provided by CDC and shown in Table 3.2), the functions that might most benefit from additional improvement efforts are dispensing,[4] which accounts for 22 percent of the overall score; security, which accounts for 10 percent; and distribution, which accounts for 10 percent.

Planning Jurisdictions in Most MSAs Scored Highly, but They Varied Considerably Across Functional Areas

Local TAR scores may be reported at the MSA level (typically, the average of all planning jurisdictions within a given MSA is reported). Yet, because there is usually no overarching

2 Given the absence of rigorous empirical data linking TAR performance to other outcomes, the threshold is based on the judgment and experience of CDC staff.

3 Local planning jurisdictions are mainly responsible for operating dispensing sites.

4 As noted earlier, dispensing is a local function, but states provide coordination and guidance.

Table 3.2
Average and Range of State-Level Overall and Functional-Area TAR Scores for 2009–2010

Functional Area	Weights (%)[a]	Mean	Std Dev	Median	Range	% Above Threshold[b]
Developing an SNS plan	3	95	10	100	67–100	NA
Management of SNS/command and control	10	96	7	100	67–100	NA
Requesting SNS	3	99	4	100	83–100	NA
Tactical communications	3	96	8	100	67–100	NA
Public information and communication	7	95	11	100	42–100	NA
Security	10	93	13	100	20–100	NA
RSS	14	97	5	98	79–100	NA
Controlling inventory	3	96	6	100	75–100	NA
Repackaging	2	88	23	100	0–100	NA
Distribution	10	94	10	100	57–100	NA
Dispensing	22	92	12	100	44–100	NA
Hospitals/alternate care facilities coordination	3	92	16	100	30–100	NA
Training, exercise, and evaluation	10	95	8	100	72–100	NA
Overall score	NA	94	6	95	79–100	100

NOTES: NA = not applicable. Std Dev = standard deviation.

[a] CDC Division of Strategic National Stockpile (DSNS) staff generated the weights.

[b] No minimum performance thresholds are defined for individual functional area scores. There were 50 observations for each row in the table.

MSA-level governance structure (although an MSA with a single jurisdiction would have unified governance), the performance thresholds for local TAR scores apply only to individual planning jurisdictions, not to MSAs as such.[5]

Table 3.3 presents a similar analysis, this time focusing on local TAR scores aggregated to the MSA level. The overall local score is the average of the 12 functional area scores, weighted by importance. As of 2009–2010, the average MSA achieved a score of 86 out of 100, with a median score of 89. However, there was more variability among local scores than among states scores, with four MSAs having an average local score below 69 and one as low as 37.

Overall, as the ranges in the last column of the table indicate, there was much more variability across functions in MSA-level TAR scores than in state TAR scores. Based on the importance weights, the functional areas that might benefit most from additional improvement are training, exercise, and evaluation, which accounts for 10 percent of the overall score; security, which accounts for 10 percent; and dispensing, which accounts for 24 percent.

Information from the case studies helps illustrate what might lie behind the somewhat lower TAR scores for security. Three sites that experienced challenges with security noted that their plans relied largely on informal arrangements and relationships with public safety rather than on the formal agreements required by the TAR. Although most of these officials indicated that relationships were improving with time and increased interaction, one of the three sites still had not created formal documentation of these relationships. The three sites also reported seeking alternative security arrangements, as necessary. Several of the case study sites reported challenges in providing documentation of activities and relationships across multiple areas of the TAR.

In sum, the overall level of performance at the state and MSA levels is quite high (as defined by the DSNS threshold), but there is considerable variation across functional areas and among MSAs. The remainder of this chapter explores patterns in that variation.

2010 TAR Scores Were Higher in MSAs in High-Performing, Centralized States

Statistical modeling was used to assess whether variations in 2010 MSA-level TAR scores followed systematic patterns and whether some types of MSAs might benefit most from additional technical assistance or other support (Appendix B provides technical detail on the analysis). Although it is not possible to draw strong causal conclusions from the available data, the results suggest that the two factors described in the remainder of this section were associated with 2010 MSA-level TAR scores.

State TAR Performance

Higher state scores on the TAR were associated with higher MSA-level TAR scores in 2010. On average, having a 1.0-point higher 2010 state score was associated with a 1.2-point increase in an MSA's 2010 score. This finding echoes an earlier study's finding (based on a small number of case studies) that strong CRI preparedness efforts in state health departments help support MSA-level efforts (reported in Nelson, Willis, et al., 2010; and Willis et al., 2009).

---

[5] The PHEP Cooperative Agreement gives money to states for preparedness, and states are to distribute a portion of that money to local planning jurisdictions within the CRI MSAs.

Table 3.3
Average and Range of MSA-Level Overall and Functional-Area TAR Scores for 2009–2010

Functional Area	Weights (%)[a]	Mean	Std Dev	Median	Range
Developing an SNS plan	3	92	12	95	17–100
Management of SNS/command and control	10	88	14	92	17–100
Requesting SNS	3	96	12	100	20–100
Tactical communications	3	89	12	92	50–100
Public information and communication	7	91	9	93	64–100
Security	10	82	16	86	10–100
Regional/local distribution site	14	86	17	93	40–100
Controlling inventory	3	91	11	94	40–100
Regional/local distribution	10	84	18	88	29–100
Dispensing	24	84	11	87	27–100
Hospitals/alternate care facilities coordination	3	90	15	96	20–100
Training, exercise, and evaluation	10	86	13	88	32–100
Overall score	NA	86	10	89	37–99

NOTE: Regional/local distribution site, regional/local distribution, and hospitals/alternate care facilities coordination may each be designated as not applicable to a given planning jurisdiction. In such cases, the other weights are adjusted accordingly.

[a] CDC DSNS staff generated the weights.

Centralization

Centralization reflects the degree to which public health policy decisions are made at the state rather than local level (Parker et al., 2012). Because many SNS functions (e.g., distribution) require cooperation among state and local entities, it may be expected that states with more-centralized public health systems will have an advantage, given the higher existing level of vertical integration.[6] The assessment found that higher state scores were, in fact, associated with higher average 2010 MSA-level TAR scores. However, the size of the effect was modest. On average, the TAR scores of MSAs in states classified as highly decentralized were approximately two points lower than MSAs in states classified as highly centralized.[7]

Chapter Summary

- The TAR measures the completion of a weighted composite of critical planning tasks. All states have been assessed annually on the S-TAR, and, since 2007–2008, 97 percent or more of local planning jurisdictions have been assessed on the L-TAR. This finding highlights the completeness of the data and suggests a high degree of overall engagement in the program.
- As of 2009–2010, all states had met the CDC-defined acceptable threshold of 79 out of 100 on the S-TAR. Although scores were high across all functional areas, scores were somewhat lower for coordination and guidance for dispensing,[8] security, and distribution—three areas with relatively high weight on the S-TAR.
- As of 2009–2010, the average MSA achieved a score of 86 out of 100, with a median of 89 percent. However, there was more variability in local scores than among states, with four MSAs having an average local score below 69 and one as low as 37. There was also considerable variation among the functional areas, with lower scores for training, exercise, and evaluation; security; and dispensing.
- MSAs in higher-scoring states and those in centralized public health systems performed better on the 2009–2010 TAR compared with MSAs in lower-scoring states with less-centralized public health, controlling for other factors.

---

[6] This result may also reflect, in part, the fact that more-centralized states are less likely to have a regional/local distribution site. Scores on this functional area tended to be relatively low, on average, compared with other functional areas (see Table 3.3).

[7] In addition, the statistical analysis found that differences among the ten HHS-defined regions remain even after controlling for other factors. However, it was not possible to determine what may be driving those differences (see Appendix B).

[8] Local planning jurisdictions are mainly responsible for operating dispensing sites.

CHAPTER FOUR

Evidence of Operational Capabilities

In addition to completing the TAR, each planning jurisdiction within an MSA must conduct at least three drills per year, selected from a standard toolkit, and report a standardized set of metrics back to CDC/DSNS. Jurisdictions may choose from any combination of five drills.[1] Failure to meet this requirement can result in the withholding of funding, as mandated by the Pandemic and All-Hazards Preparedness Act.[2]

This chapter examines data from the first two years of drill data collection (2008–2009 and 2009–2010)[3] in order to assess the degree to which the assessments are actually being used and to provide baseline data from two selected drills (staff call-down and dispensing) that shed light on current capabilities and provide a basis for judging future performance. Given the relative newness of data collection, the findings should be regarded as suggestive but not yet conclusive.

Widespread Data Collection and Reporting Provide Evidence That MSAs Are Testing Operational Capabilities

Table 4.1 displays the number of drills reported by jurisdictions in the 2008–2009 and 2009–2010 program periods. The most commonly reported drill was a staff call-down (in which jurisdictions test whether they can contact key staff quickly), followed by facility setup and site activation (tasks that ensure that key sites and facilities would be available on short notice).

The prevalence of staff call-down drills is not surprising, given that item 2.4 of the TAR assesses whether jurisdictions complete quarterly call-down drills. The relative infrequency of dispensing drills might be related to the considerable time and effort required to design and conduct point of dispensing (POD) exercises. The infrequency of pick-list generation drills may be due to the fact that only a subset of local jurisdictions perform distribution and warehousing (functions primarily handled by states).

1 In 2009–2010, jurisdictions were required to report three different drill types; in 2008–2009, they could report more than one of the same drill type (e.g., three call-down drills).

2 Pub. L. 109-417.

3 DSNS aggregated drill data to the MSA level to maintain confidentiality. Jurisdictions in 45 MSAs reported more than the required three drills during this two-year period. Other jurisdictions may have completed more than three drills, but they did not report them.

Table 4.1
Type and Number of Drills Reported, by Program Period

	Number in 2008–2009 (%)	Number in 2009–2010 (%)
Staff call-down	541 (40%)	383 (27%)
Facility setup	303 (22%)	322 (23%)
Site activation	247 (18%)	222 (16%)
Dispensing	189 (14%)	208 (15%)
Pick-list generation	84 (6%)	70 (5%)
Other[a]	NA	217 (15%)
Total	1,364 (100%)	1,422 (100%)[a]

[a] Includes decisionmaking or inventory management drills and functional or full-scale exercise after-action reports.

[b] Percentages do not sum to 100 due to rounding.

A comparison of the types of drills reported in 2008–2009 and in 2009–2010 indicates that the number of call-down, site activation, and pick-list generation drills decreased and that the number of other drills—particularly facility setup and dispensing—increased. This trend corresponds to the change in the type of drills that jurisdictions were required to report over the two periods. Specifically, in 2008–2009, jurisdictions were required to submit three drills of any type, which allowed for multiple instances of the same type of drill. In 2009–2010, however, jurisdictions had to submit three different types of drills.

Jurisdictions Have Tested Call-Down Capabilities, but Not on a Large Scale

Staff call-down is the ability to contact and mobilize staff to perform emergency response functions. Many of the functions employed in countermeasure delivery (e.g., dispensing, warehousing, distribution, security, command centers) require staff to report for duty upon notification and, consequently, require that call-down be performed successfully. For staff call-down to succeed, the notifications must occur in a timely fashion (measured by calling completion time) and reach the necessary personnel (measured by acknowledgment percentage).

Calling Completion Time

To assess whether call-down processes can be conducted in a timely manner, the call-down drill includes a measure of *calling completion time*, the time elapsed during the calling process. For drills conducted during 2009–2010, the average completion time was 80 minutes, but there was considerable variation among the drills, as indicated by the standard deviation of 133 minutes. A few drills had extremely long completion times (as long as 700 minutes), but nearly 62 percent of the drills had completion times of less than 80 minutes. The median time (a statistic less susceptible than the average to extremely high or low values) was 47 minutes.[4] Table C.1 in Appendix C provides more detail related to call-down completion time results.

[4] Call-down completion times differed slightly depending on the type of call-down list tested; however, these differences were so small as to be statistically nonsignificant (Kruskal-Wallis p = 0.34). Statistical tests were also performed to explore

Acknowledgment Percentage

To assess whether call-down processes can reach all necessary personnel, the call-down drill includes a measure of *acknowledgment percentage*, the percentage of staff on the call-down list who were successfully contacted. Acknowledgment percentage is calculated by dividing the number of staff successfully contacted by the number of staff on the list.

For drills conducted during 2009–2010, the median acknowledgment percentage was 83 percent, and the average acknowledgment percentage was 78 percent. However, there was considerable variation among drills, with a standard deviation of 24 percentage points. Although 28 percent (n = 102) of the 368 call-down drills successfully contacted all of the individuals on the call-down list, approximately 13 percent of the drills reached fewer than half of the individuals on the list. Acknowledgment percentages were particularly high for public information and communication (median = 100 percent) and were particularly low for security (median = 53 percent).[5] Table C.2 in Appendix C provides a detailed tabulation of the acknowledgment percentages reported.

One possible explanation for the variation in acknowledgment percentages is that individuals on the security call-down list are often external partners, whereas individuals on the public information and communication call-down lists are more likely to be internal health department staff. The analysis also shows that call-down drills conducted using manual approaches (e.g., telephone calling trees) had higher acknowledgment rates (median = 92 percent) than those involving automatic equipment (median = 72 percent).[6] This may suggest that calls made by automated methods are easier to miss or ignore. Note that these are only hypotheses; further investigation would be needed to confirm them.

Another factor that does appear to have a major effect on acknowledgment percentages is the size of the call-down list. As shown in the "Number of Drills" column in Table 4.2, most of the call-down drills involved calling relatively few people. Nearly 60 percent of the drills

Table 4.2
Acknowledgment Percentage Compared with the Size of the Call-Down Drill

List Size	Number of Drills	Mean	Median	Std Dev
0–25	216	86	90	15
26–50	67	71	77	25
51–75	24	67	70	25
76–100	18	65	82	34
101–200	25	64	68	24
More than 200	17	46	40	28

whether degree of notice, equipment used, and list length would explain any of the variance, but none were found to be conclusive. Specifically, an ordinary least squares regression of the cube of time elapsed on list type and size, type of notice, and type of equipment used yielded an adjusted R^2 of 0.002. Thus, none of these factors explains much variation in time elapsed. For all analysis of elapsed call time, cases with negative times were omitted. Analyses using list length omitted one outlier with more than 2,000 individuals on the list.

5 A Kruskal-Wallis test found statistically significant differences in acknowledgment percentages across different types of call-down lists ($p < 0.02$).

6 Kruskal-Wallis $p < 0.001$.

involved call-down lists of 25 or fewer people, and nearly 90 percent involved 100 or fewer people. Only a small number of drills involved calling more than 200 people.

As shown in both the "Mean" and "Median" columns of Table 4.2, there is a visible relationship between the size of the call-down list and the mean acknowledgment percentage.[7] Drills that involve calling larger numbers of people exhibited lower mean acknowledgment percentages. For instance, the mean acknowledgment percentage for drills with lists of 25 or fewer people was 86 percent, compared with 71 percent for lists of 26–50 people and 46 percent for lists of more than 200 people. Median acknowledgment percentages generally followed the same pattern of decreased acknowledgment for larger list sizes, with an exception in the statistics for lists of 76–100 people.

One possible explanation for this pattern is that the smaller call-down drills may have tested emergency personnel who are accustomed to being contacted and responding on short notice, whereas larger lists may include a wider array of staff and volunteers who are not so easily reached and who do not typically respond. It should also be noted that, although the drills report the size of the call-down list that was tested, they do not report the true size of the call-down *roster*. Consequently, there is no way to know what percentage of the total call-down roster was tested in the drill. This raises the concern that call-down drills that test only small lists of core emergency staff will not be indicative of call-downs of larger lists, particularly when those larger lists include persons outside the department.

Jurisdictions Have Practiced PODs, but Tests of Throughput Are Not Yet Conclusive

The measure assessed in the POD drill—*POD throughput*, or the rate per hour at which countermeasures are dispensed to POD clients—serves as a proxy for the capability to dispense countermeasures to an entire affected population within the necessary time frame. Jurisdictions may be capable of achieving countermeasure delivery goals with many small PODs with low throughput or with a few large PODs with high throughput. POD drill throughput must therefore be interpreted in the context of a jurisdiction's countermeasure delivery plan.

Planning jurisdictions report the size of their population and the total number of PODs needed in their plans as part of the TAR. However, they report neither the expected number of clients who will be served at each POD on a POD-by-POD basis nor the throughput that each POD needs to achieve. In the absence of such detailed information, average population per POD and average throughput per POD can provide insight. Dividing the population by the number of PODs gives the number of clients who must be served by each POD. Dividing this result by the number of hours available for dispensing operations (likely 24) gives the average *required throughput* per POD.

---

7 Spearman's rho = –0.53, $p < 0.001$. The negative relationship between the two variables shown in Table 4.2 persisted—and was statistically significant—in a regression model of acknowledgment percentage as a function of list size (both the continuous and categorical versions of the variable) and list type (e.g., emergency operations center versus public information and communication).

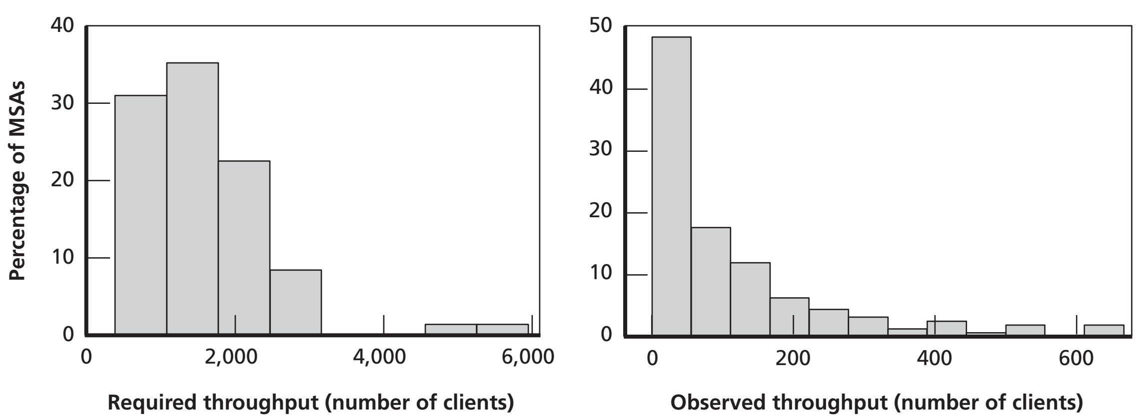
Based on the self-reported population figures and number of PODs (and excluding closed PODs),[8] average required throughput per POD for most of the MSAs ranges from approximately 400 people per hour to approximately 3,000 people per hour, with a median of 1,446 people per hour. Required throughputs are shown in the left panel of Figure 4.1. Approximately 30 percent of the MSAs have a required throughput of 1,000 or less, and another 50 percent have a required throughput of between 1,000 and 2,000.

The *observed throughput* for each drill was calculated by dividing the number of clients (or client-actors in a drill) served by the length of time that the POD was open. As the right panel of Figure 4.1 shows, only a handful of drills (6 percent) achieved throughputs of greater than 400 people per hour, which would have been the low end of the required throughput, and the highest observed throughputs involved three drills with throughputs in the range of 600–700 people per hour. Most throughputs (83 percent) were 200 people per hour or lower, and, in all but one MSA, observed throughput was lower than required throughput.

The observed throughputs are misleading, however. A common problem in POD drills is that the number of clients (or client-actors) is often quite low, especially relative to the duration of time that the POD is operating, whether in a real incident or an exercise. Only a handful of exercises involved more than 2,500 clients, and only 32 percent of the exercises had 500 clients or more. Dividing such a small number of clients by the number of hours that the PODs were open will result in low throughput numbers, not necessarily because the POD staff were slow but simply because there were not enough clients to provide a good test of the speed of POD operations.

Figure 4.2 graphs the relationship between the number of clients in the POD drill and the POD's observed throughput. The figure shows that drills with more clients demonstrated

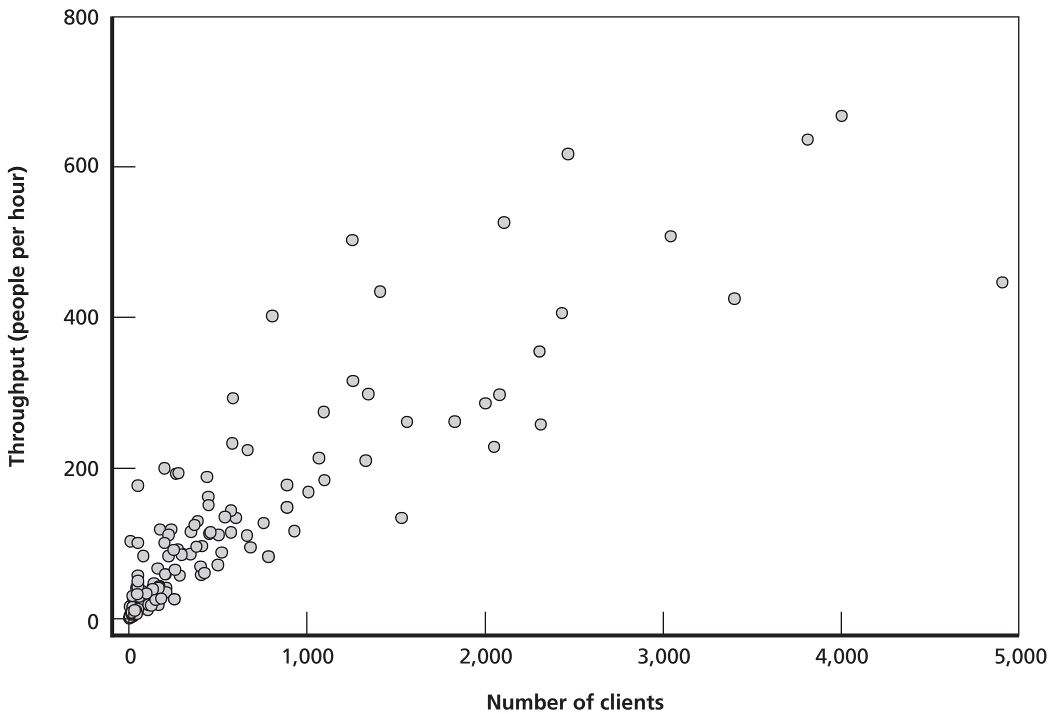
Figure 4.1
Observed and Required POD Throughputs



RAND *TR1200-4.1*

[8] Data collected by CDC on MSAs, through the TAR, includes the number of "closed PODs" in the MSA's plans. These closed PODs may be operated within health care facilities, schools, businesses, and military installations. We expect, but are not certain, that these closed PODs will typically serve fewer people than the public PODs. The data do not indicate how many people would be served by these closed PODs. Consequently, we excluded the closed PODs in our calculations of the number of people who need to be served by each (public) POD. This results in an overestimate of the required throughput per POD.

Figure 4.2
Correlation Between Throughput and Number of Clients in Drills



RAND *TR1200-4.2*

higher throughput.[9] For instance, most of the drills with 1,000 or fewer clients reported throughputs of less than 200, while drills with 3,000 or more clients reported throughputs above 400. Although further investigation is required to fully understand this pattern, there is no inherent reason why throughput should depend significantly on how many clients come to the POD.[10] Thus, it is likely that observed throughputs in many of the POD drills are low due not to POD performance but rather to a shortage in the number of clients needed to test throughput adequately.

Another reason that observed throughputs may appear much lower than the required throughputs is that many of the POD drills submitted by jurisdictions involved PODs that differed from those that would be operated in a CRI-type anthrax scenario, where pills would be dispensed. For example, of the 200 POD drills, 78 involved real-world flu vaccination PODs. Because the vaccination process is different from the oral-antibiotic dispensing process, and because these flu clinics likely had different staffing and operational infrastructure than PODs used for mass dispensing during an anthrax outbreak, it is difficult to compare the observed throughputs from flu PODs with the required throughputs for anthrax PODs.

Chapter Six presents recommendations and next steps for ensuring that the drill-based data collection system can assess communities' readiness to meet the 48-hour goal. For now,

[9] The correlation coefficient (Pearson's r) between log transformed versions of both variables (to correct for skewness) was 0.90, p < 001.

[10] It is possible that POD staff will work faster when faced with large crowds. But there will be limits on how much faster they can operate.

however, it is not yet possible to draw conclusions about readiness from the observed throughputs in the POD drills.

Time-Study Data Provide a Basis for Understanding and Improving PODs

Difficulties in obtaining good measures of throughput were anticipated during the process of designing the drills. Consequently, the data collection spreadsheets for the POD drill also ask jurisdictions to break down client processing times by POD step. If information about POD layout were available, including the proportion of clients who go to the various branches within the POD (e.g., express dispensing versus medical assessment), it would be possible to use this information, in conjunction with the processing-time data, in computer models to estimate the maximum possible throughput of the POD. However, such information is not always currently collected.

Table 4.3 shows the average processing time for nine common POD steps (as defined by those performing the drill). Table C.3 in Appendix C provides a crosswalk between the original step names and the nine common POD steps. Given the small number of drills reporting usable data for some of the steps (e.g., only four drills used an education step), it is impossible to draw strong conclusions from these data. However, they provide an indication of how data on processing times may be used to improve POD performance.

Across all drills, the average processing time for each step ranged from about 2 to 4 minutes, with education having the lowest average processing time (1.72 minutes) and medical consult having the highest (4.34 minutes). In addition, there were variations among drills in processing times for each step, suggesting that there may be wide variation in how jurisdictions have designed the steps within their PODs. Standard deviations ranged from a low of 0.54 minutes (education) to a high of 8.61 minutes (registration). Also of note is that average vaccination times (such as in a flu POD) are somewhat higher than average oral-antibiotic dis-

Table 4.3
Average Processing Time, per POD Step

Step	Average Processing Time (minutes)	Standard Deviation of Processing Time (minutes)	Number of Drills
Medical consult	4.34	4.52	8
Vaccination	4.20	3.75	58
Registration	3.76	8.61	107
Dispensing	2.86	3.94	94
Screening	2.77	2.90	37
Exit	2.26	3.58	7
Triage	2.24	2.41	27
Entry	2.17	3.06	12
Survey	2.10	1.69	7
Education	1.72	0.54	4

pensing (such as in an anthrax POD exercise) and are more variable. This shows that, although annual flu vaccination clinics provide a good opportunity for jurisdictions to practice POD procedures, the data collected in those clinics are not easily generalized to oral-antibiotic dispensing PODs.[11]

Chapter Summary

- Planning jurisdictions in the 72 MSAs participating in CRI conducted and reported data on 1,364 drills in 2008–2009 and on 1,422 drills in 2009–2010.
- A large number of jurisdictions have tested staff call-down procedures, but often not on a large scale. The limited scope of the drills limits efforts to estimate the capability to contact all needed staff during a large-scale emergency.
- Many jurisdictions have tested dispensing in PODs, but throughputs are difficult to interpret without more-stringent tests (i.e., POD drills with more clients). Generally, POD drills with higher numbers of clients report consistently higher throughputs, suggesting that, were jurisdictions to run more large-scale POD drills, higher throughputs and greater countermeasure dispensing capability might be revealed.
- The time-study data currently collected by jurisdictions provide information that can help them plan and staff their PODs. Additional insight could be gained if POD layout information were also consistently collected.

---

[11] Jurisdictions can use these processing-time statistics in two ways. First, if they have never designed a POD, these average processing times can help them understand where to allocate staff within the POD; in general, steps that take longer will require more staff than steps that take a shorter amount of time. Second, if a jurisdiction has run a POD drill and collected its own processing-time data, it can compare those results with these averages, looking for opportunities to streamline its procedures if it finds that its average processing times are longer than others'.

CHAPTER FIVE

Evidence of CRI's Effectiveness in Improving Readiness

The previous two chapters examined CRI jurisdictions' current capacity (Chapter Three) and capability (Chapter Four) to engage in medical countermeasure delivery. This chapter focuses on the extent to which the CRI program has *improved* capacities over time. Drill data are not yet sufficient to judge trends over time in capabilities.

Available data are not compatible with use of the strongest methods for assessing program effects (i.e., comparison group designs).[1] However, it is possible to assess whether, compared with less exposure, more exposure to the CRI program leads to more of the desired output (i.e., preparedness). This chapter also assesses whether there is evidence that CRI had "spill-over" effects on nonparticipating MSAs.

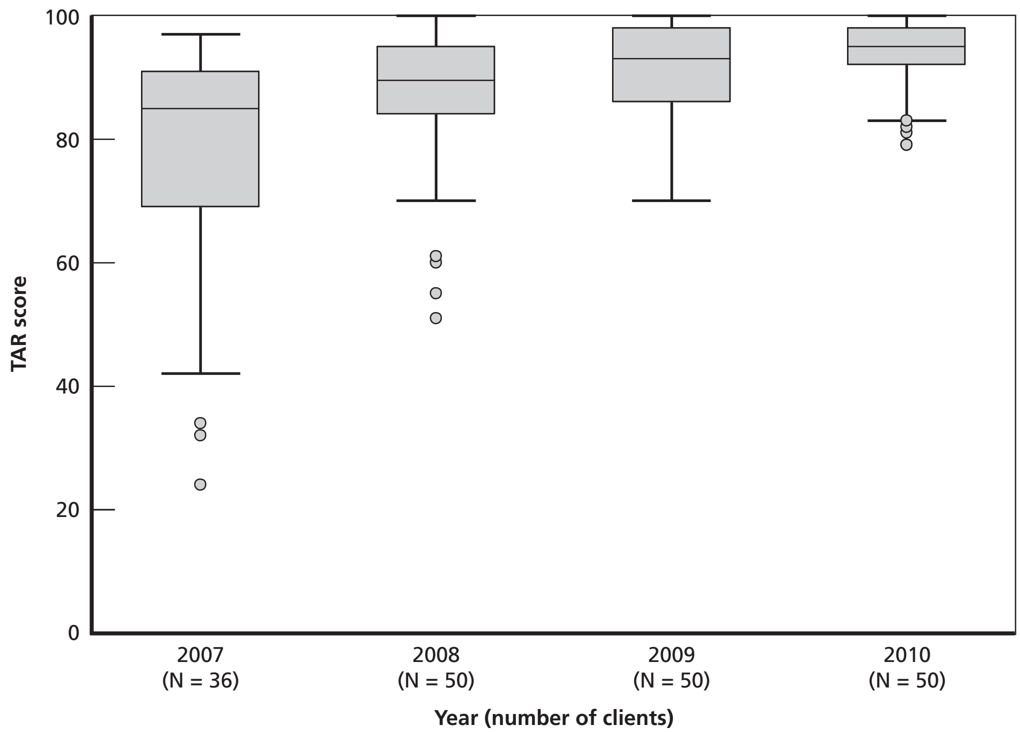
State TAR Scores Improved

Overall, state TAR scores have improved over time. The median state TAR score rose from 85 in 2006–2007 to 95 in 2009–2010. Furthermore, the range of scores has narrowed over time. These patterns are illustrated by means of boxplots in Figure 5.1. The upward trend in median scores is represented by the center lines in each box, and the decrease in spread is illustrated in the compression over time in the upper and lower edges of the box (the 25th and 75th percentiles) and in the lines above and below the box (which represent 1.5 times the distance between the 25th and 75th percentiles). Note, however, that CDC/DSNS allowed 29 states to carry over a score of 90 or better from 2008–2009 to 2009–2010 without reassessment. Hence, any change in the distribution of scores between these two years is entirely due to changes in the other sites.

Figure 5.2 plots average state TAR scores over time, by functional area. As with the overall scores shown in Figure 5.1, each average function-level score improved over time, as evidenced by the fact that all of the lines are upward-sloping. However, scores for some functions increased faster than others. The smallest increases were in requesting SNS (91 to 99) and developing an SNS plan (87 to 95), and the largest was in hospitals/alternate care facilities

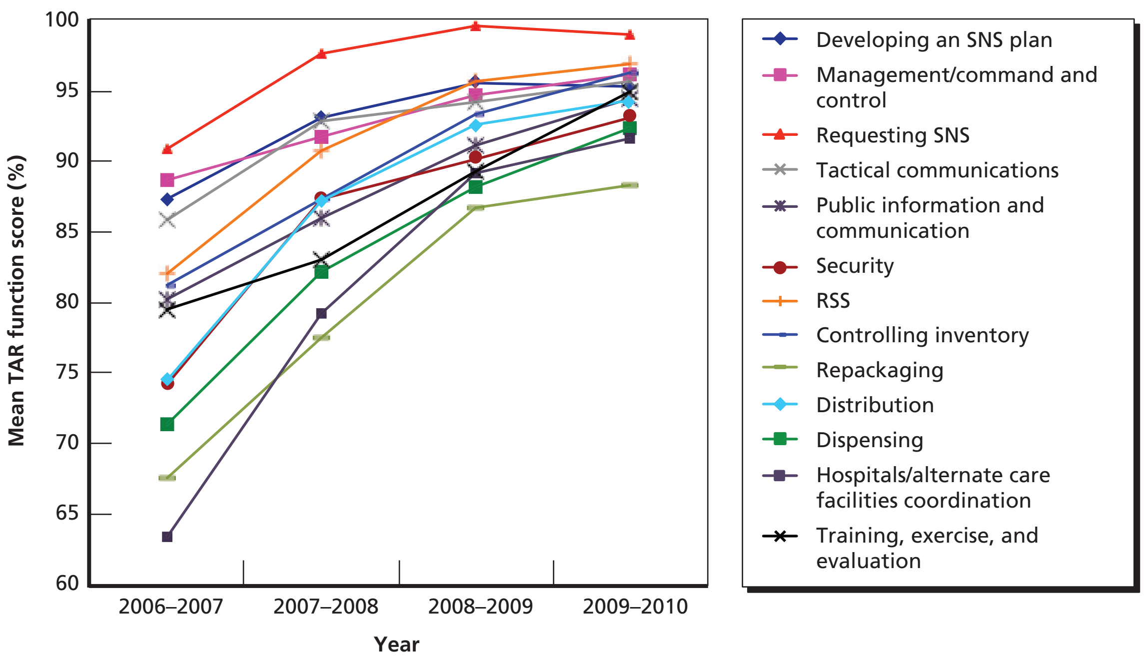
[1] Ideally, an assessment of CRI's impact would compare trends in readiness before and after entry into the program with trends in readiness over the same period at nonparticipating sites. If the participating and nonparticipating sites were similar, and if CRI sites posted *larger improvements* in readiness than non-CRI sites, this would reflect the impact of CRI. However, participating MSAs were selected largely based on size and risk; thus, MSAs that might make good comparison groups for the larger MSAs are also in the program. Moreover, the fact that performance data were not collected before the advent of the CRI program prevents before/after comparisons, and the fact that TAR has been mandated only for participating sites prevents comparisons between participating and nonparticipating sites on a common scale.

Figure 5.1
Median and Range of State-Level Overall TAR Scores, by Year



RAND *TR1200-5.1*

Figure 5.2
Average State Function-Area TAR Scores, by Year



RAND *TR1200-5.2*

coordination (63 to 93). Generally, the analysis shows greater increases in functional areas in which scores were initially lower, resulting in an overall convergence among functional areas over time.

Note, however, that it is possible that the functions with high scores in 2007 might have showed greater growth if a more demanding assessment tool had been used (i.e., scores may be experiencing a "ceiling effect"). Many of the stakeholders consulted for the case studies believed that the TAR provides a better description of the early stages of preparedness than of more-advanced levels of preparedness, which would be consistent with the notion of a ceiling effect.[2]

MSA TAR Scores Also Improved

Figure 5.3 provides boxplots of MSA-level TAR scores over time. The median overall MSA-level TAR scores rose over time (from 52 in 2006–2007 to 88 in 2009–2010), with corresponding decreases in the range of scores. Over the four years, the percentage of MSAs scoring 69 percent or higher increased from 28 to 94.

However, as Figure 5.4 shows, some functions' average scores remained considerably lower than others, and convergence over time occurred in some, but not all, functions (perhaps due to a less pronounced ceiling effect). For instance, scores for requesting SNS were consistently high in all four years, with a four-year average score of 88, whereas the functional areas of regional/local distribution (most commonly the responsibility of states, not local jurisdictions) and dispensing had the lowest average scores (each with a four-year average of 70). Hospitals/alternate care facilities coordination showed marked improvement between 2006–2007 and 2009–2010.

Qualitative analysis highlighted a number of specific improvement efforts represented in the TAR data. Five of nine sites focused on improving drills and exercises, and four of nine focused on expanding their volunteer base. Three of nine sites focused on PODs, and one switched from a highly decentralized system with many smaller PODs to a more centralized system with fewer PODs in larger venues near mass transportation, thus reducing the number of staff required.

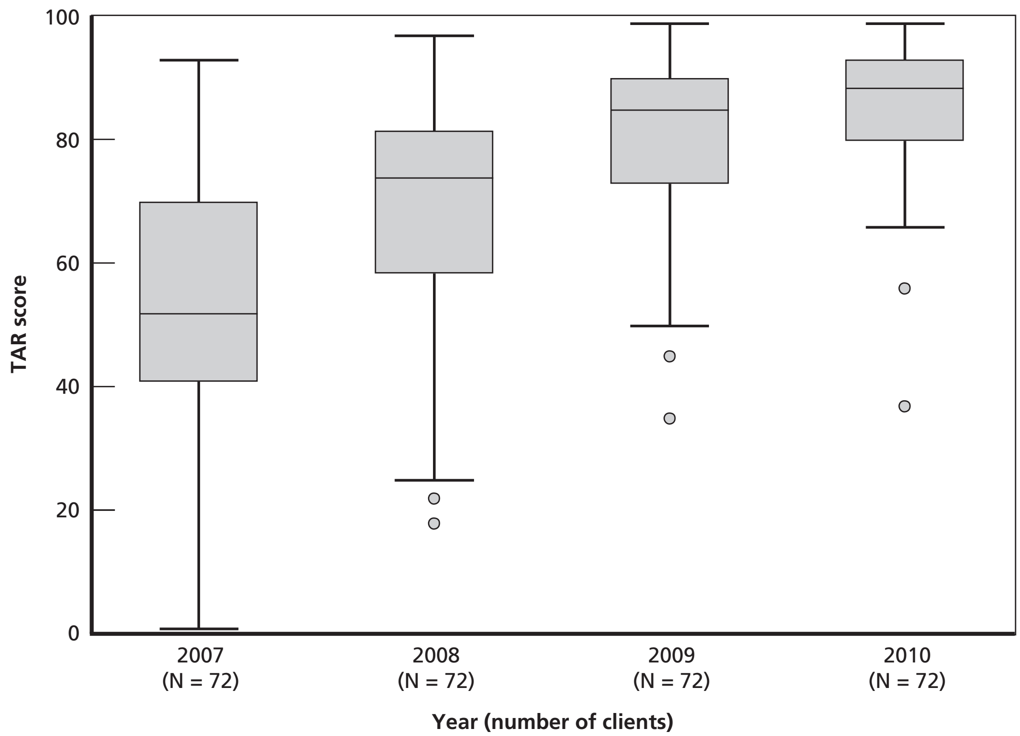
Other sites developed in-state resources. One site acquired its own warehouse and stockpile with CRI funding, and another purchased a warehouse management system. Finally, one site developed a special needs task force that hosted a seminar series on cultural competency for staff and volunteers, led training for closed PODs, and tested communications for special populations.[3]

Another way to explore the potential effects of increased exposure to the CRI intervention is to compare TAR scores across the three cohorts of CRI MSAs that entered the program in 2004, 2005, and 2006. The orange lines in Figure 5.5 represent the average of all overall TAR scores within each cohort, by year. The black lines represent the trends in individual MSA average scores. There is a great deal of variability in trajectories across MSAs. Yet, average overall TAR scores for the three cohorts became more similar over time (statistically significant cohort

[2] An earlier study (Willis et al., 2009) produced the same finding.

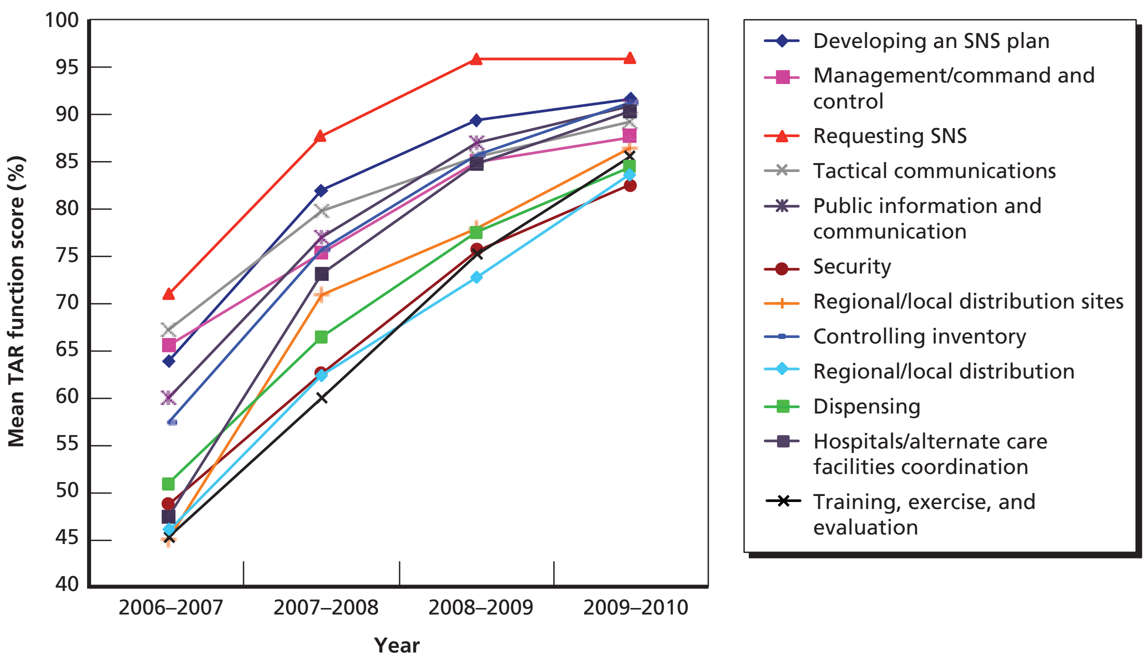
[3] Development of vulnerable populations plans for dispensing only accounts for up to 2 TAR points, making them difficult to detect in TAR data analysis.

Figure 5.3
Median and Range of MSA-Level Overall TAR Scores, by Year



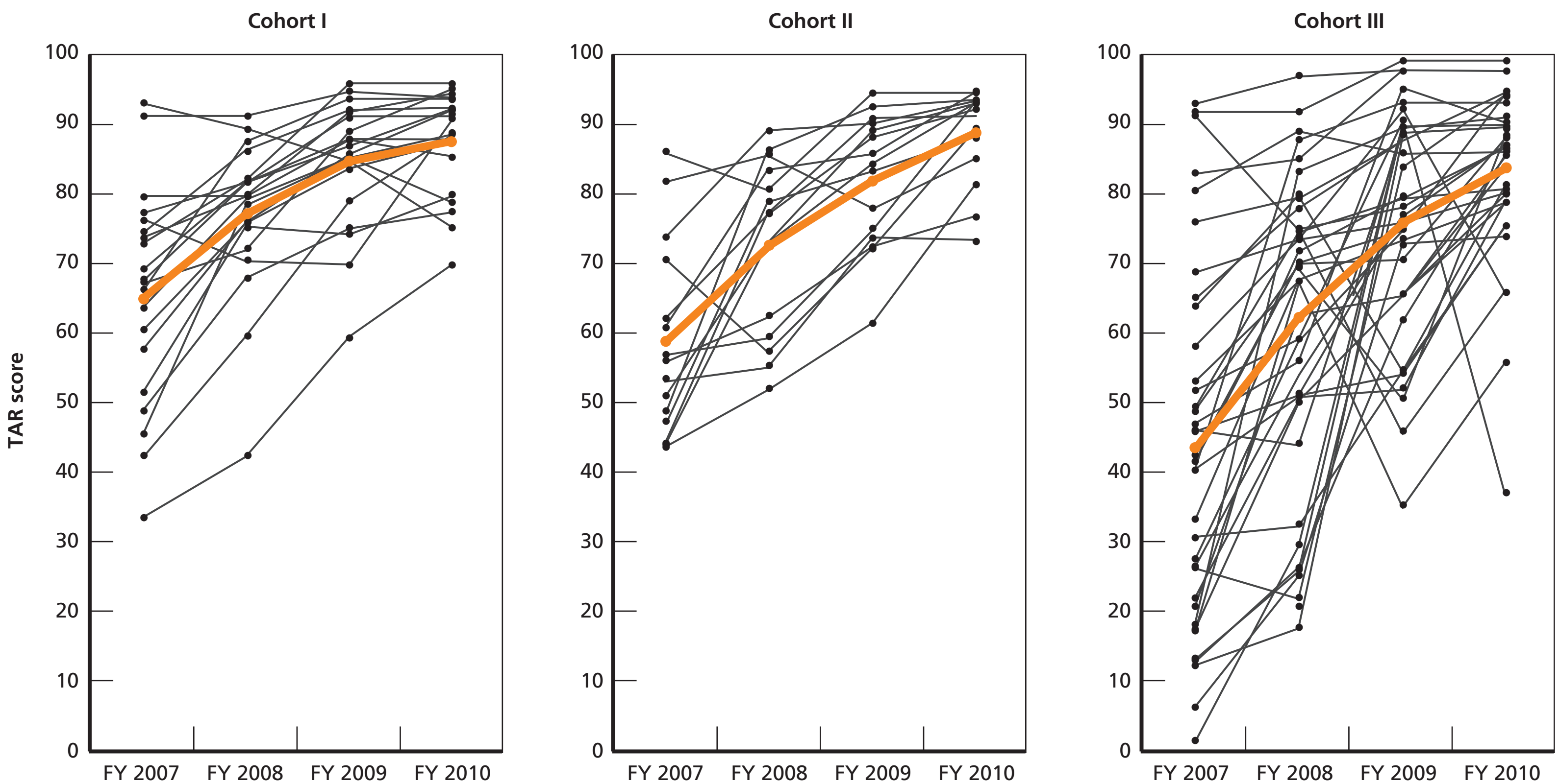
RAND *TR1200-5.3*

Figure 5.4
Average CRI MSA-Level Function-Area TAR Scores, by Year



RAND *TR1200-5.4*

Figure 5.5
Individual MSA-Level and Average Overall TAR Scores, by Year



NOTE: Orange lines represent the average of all overall TAR scores within each cohort. Black lines represent the trends of individual MSA average scores.
FY = fiscal year.

RAND *TR1200-5.5*

mean differences in 2006–2007 disappear by 2009–2010). Furthermore, both between- and within-cohort variances (a statistical measure of the range in scores) appear to be shrinking.[4] It is important to bear in mind, however, that the cohorts differ in important ways (e.g., earlier cohorts include larger MSAs, on average). Thus, differences among them may not provide evidence of a causal impact of CRI. When paired with findings from an earlier CRI evaluation (Willis et al., 2009), these findings show mostly continuous growth in TAR scores. The earlier report found that the average Cohort II MSA received a TAR score of 49 in 2006 and the average Cohort III MSA a score of 43 in 2006–2007. Thus, the growth in these cohorts' scores continues growth trends that began earlier. The long-term trend in Cohort I is less straightforward. The earlier study found that the cohort's average TAR score grew from 50 to 72 between 2004 and 2005–2006, a statistically significant increase. Taken together with the findings presented in Figure 5.5, this suggests a brief interruption in the upward trend between 2005–2006 and 2006–2007. However, changes in the TAR instrument and scoring practices over time make it difficult to interpret this pattern.

In sum, there is suggestive, but not conclusive, evidence of improvement as sites are exposed to more of the CRI intervention. There is interesting variation across sites and functional areas. A closer examination of those variations might help determine the degree to which these improvements were caused by CRI. These analyses are described in the next sections.

MSAs in High-Performing, Centralized States Improved More Quickly

Exploring whether there are systematic patterns in MSA's TAR trends (see Figure 5.5) might provide additional insight about how CRI works and which types of MSAs might benefit from additional technical assistance or other forms of support. Using statistical models to examine the relationship between MSA-level TAR score increases and a variety of demographic and governance variables shows that the following factors were associated with 2010 MSA-level TAR growth:

- *State TAR performance.* Larger increases in state TAR scores were associated with higher MSA-level TAR scores, after taking into account where those MSAs started. The state effect was apparent both in 2008 state score *levels* (a 1.0-point state increase was associated with a 1.1-point MSA increase) and in state score *growth* from 2007–2008 to 2009–2010 (a 1.0-point state increase during that period was associated with a 1.2-point increase in MSA scores during the same period).
- *Centralization.* Centralization was associated with stronger TAR scores, but the size of the effect was modest. On average, 2009–2010 TAR scores were 2.4 points lower in highly decentralized states than in highly centralized states, even after accounting for 2007–2008 TAR levels.
- *Regional differences.* Some differences among HHS regions remained even after controlling for other factors.

---

[4] A detailed graphical and statistical analysis shows that the variance within cohorts shrinks over time, as does the variance among cohort means. The result is that the cohort distributions of overall cohort TAR scores, which were significantly different on average in 2006–2007 (ANOVA $F(2, 69) = 7.56$, $p < 0.01$), become nonsignificant by 2009–2010 ($F(2, 69) = 1.89$, $p > 0.10$).

However, this analysis should be viewed as preliminary, since there might be other factors (e.g., state and local revenue streams, staff turnover) driving TAR scores that we were unable to measure. Appendix B provides a detailed description of this analysis.

Stakeholders Report CRI Impact on Real-World Responses

Discussions with stakeholders and a review of documentation provided by CDC about specific ways in which CRI capabilities were used in responses to real incidents provided additional anecdotal support of the notion that CRI has improved readiness.[5]

H1N1

Three sites noted that efforts to prepare for rapid dispensing helped their staff streamline the setup of vaccination clinics during the H1N1 response. Others reported using CRI plans to allow the public to download and complete registration forms via the Internet prior to arrival at the site. Another site noted that staff planning for CRI helped improve the site's ability to correctly estimate staffing needs for vaccination sites.

Infectious Diseases

Similarly, sites reported using capabilities developed through CRI program planning initiatives when responding to other infectious-disease threats. As just noted, several sites reported using elements of their CRI POD models for seasonal influenza vaccination. One MSA used PODs to dispense treatments for both tuberculosis and mumps outbreaks. In other instances, PODs were used to collect and "dispense" (i.e., disseminate) information. One MSA reported leveraging its CRI call-down lists and POD activation protocols to set up a POD in response to a tuberculosis outbreak. Potentially affected community members were called to the POD to provide and receive information. Another MSA used CRI planning efforts to set up a POD to share information with the public as part of a response to a mycoplasma infection. These examples suggest that CRI protocols can be used to dispense information, as well as medical countermeasures.

Natural Disasters

There are also examples of MSAs using capabilities developed for the CRI program to respond to natural disasters. For instance, one MSA reported leveraging the tactical communications plans, public information messaging, incident management team protocols and procedures, and distribution networks developed for the CRI program to support an extensive response (e.g., evacuation, mass care, sheltering) to a major flood. Another MSA reported using CRI distribution capabilities to deliver supplies to a reservation during an emergency. Finally, another MSA reported that mutual aid agreements, volunteer coordination mechanisms, and tactical communications systems developed with CRI resources helped in responding to a major tornado.

---

[5] Given the limited scope of the qualitative data collection, it was not possible to validate the connections to CRI investments made in the documentation or in the stakeholder discussions.

Sporting and Political Events

Three MSAs reported using CRI-derived capabilities to prepare for sporting and political events. These MSAs reported benefiting from the metropolitan-wide coordination facilitated by the CRI program and from specific capabilities supported by CRI (e.g., command and control, public information). Several sites also noted that planned events can provide opportunities to practice and test CRI capabilities (e.g., rapid facility setup).

However, MSAs that were included as part of the qualitative analysis reported that the impact of CRI might be stronger if they received clearer performance feedback from CDC. Several stakeholders noted that they spent considerable time collecting and sending data to meet program requirements but felt that the data were not synthesized in a way that allowed them to fully understand their performance or identify needed improvements. Further, MSAs thought that feedback would help validate their efforts. In addition to a general desire for feedback, the case study sites suggested specific types of feedback they would find useful. In particular, two sites noted that it would be especially useful to receive CDC feedback in person for drills, exercises, or training. Given the small number of sites involved in the qualitative data collection, these findings should be regarded as preliminary, and they warrant further examination.

Administration of TAR in Non-CRI Sites Provides Evidence of Program Spillover

Programs often have spillover effects on nonparticipating sites—effects that should be taken into account when weighing a program's overall merit. An earlier study of CRI suggested that one specific spillover was that state health departments had applied the TAR and other program doctrine to non-CRI jurisdictions (Willis et al., 2009). To help explore this more broadly, CDC/DSNS developed and administered a survey on the extent to which states apply the TAR, the drill-based assessments, or other assessment tools in non-CRI jurisdictions.[6]

Forty-four of 50 states (88 percent) reported collecting SNS performance data from non-CRI jurisdictions. Of these, 41 states (82 percent) reported using the DSNS L-TAR. Fifteen states were willing to share their assessment tools, providing a total of 334 TAR-like assessments. Of those, 231 used formats either identical or virtually identical to DSNS's TAR, and 103 used completely different assessment tools. The CDC/DSNS survey did not assess whether efforts to measure preparedness in non-CRI sites is driving *improvements* in preparedness. However, some case study sites reported that lessons from drills, exercises, and real incidents were sometimes transferred from CRI to non-CRI MSAs and, in the case of multistate MSAs, across state borders.[7] One discussant noted that participating MSAs might influence broader public expectations as citizens in smaller, non-CRI communities learn about preparedness and response activities in CRI MSAs. However, another discussant thought the smaller communities could also influence the larger ones, offering an example of a situation in which H1N1-

---

6 The survey was distributed to CDC program services consultants (the CDC officials responsible for direct technical assistance to CRI participants), who administered the survey to state CRI coordinators in March 2011. All 50 states responded, and CDC shared the data with RAND for analysis and interpretation.

7 Though not relevant to spillover to non-CRI sites, one case study example revealed that exercise lessons were transferred between two CRI MSAs in different (though bordering) states.

related school closures in a non-CRI community in a neighboring state led to public demand for school closures in a CRI MSA. One MSA specifically reported using CRI funding to assist in preparedness efforts for neighboring counties.

The number of non-CRI communities assessed and the qualitative data received provide some evidence that the assessment component of CRI is having a broader impact. The existence of the quantitative assessment data suggests that, with more-systematic sampling of non-CRI sites, future studies might be able to compare participants and nonparticipants using a common instrument.[8]

Chapter Summary

- State TAR scores have improved consistently since 2007, and the variation among state performance has decreased. The median state score increased from 85 in 2006–2007 to 95 in 2009–2010, and score increases were found in all functional areas.
- MSA-level TAR scores showed similar patterns, with the median increasing from 52 in 2006–2007 to 88 in 2009–2010. However, more variability remained among MSAs' performance than among states'. The TAR scores of MSAs in high-performing states and in those with centralized public health systems were higher than those of other MSAs, after accounting for earlier levels.
- There was anecdotal evidence that CRI has improved responses to real incidents. For example, several sites reported using elements of their CRI POD models for seasonal influenza vaccination, and others used capabilities developed for the CRI program to respond to natural disasters.
- Widespread state use of the TAR to assess non-CRI communities represents a spillover effect of CRI.
- Although TAR score increases are consistent with the claim that the CRI program caused improvements, the data are not sufficient to rule out the possibility that other factors drove the increases. Thus, the findings must be regarded as suggestive but not conclusive.

---

[8] Spillover to non-CRI sites is a sign of program impact, but it also limits those non-CRI sites' utility as a comparison group for estimating that impact. To be useful for evaluation, future data collection from non-CRI sites would need to include assessments of the degree and type of technical, financial, or other support non-CRI sites receive.

CHAPTER SIX

Policy Implications and Recommendations

The Cities Readiness Initiative uses a combination of funding, program guidance, and technical assistance to improve the ability of the nation's largest metropolitan regions to rapidly provide life-saving medications in the event of a large-scale bioterrorist attack, naturally occurring disease outbreak, or other public health emergency. The purpose of this report was to analyze program data collected by CDC in order to assess

- the current status of communities' operational capability to meet the CRI program goal of delivering medical countermeasures to entire MSAs within 48 hours of the federal decision to deploy assets
- whether there is evidence that CRI has improved communities' capability to meet the 48-hour goal.

This concluding chapter highlights key findings from the study and provides recommendations for moving forward.

Capacity as Measured by the TAR Appears to Be Strong

Overall, TAR scores suggest a high degree of capacity, with states achieving a median score of 95 out of 100 and planning jurisdictions in MSAs achieving a median score of 88 on the TAR. However, there was considerable variation in performance across MSAs. Scores on the state TARs were somewhat lower in the functional areas of security, dispensing (for which states provide coordination and guidance), and distribution. Scores on the local TARs were lower in the areas of security; dispensing (local jurisdictions conduct actual dispensing); and training, exercise, and evaluation. These functional areas might benefit from additional planning attention.

Testing of Operational Capabilities Has Not Been Conducted at the Scale Required to Test Readiness for the 48-Hour Scenario

Planning jurisdictions conducted and reported data on 1,364 drills in 2008–2009 and on 1,422 drills 2009–2010. A large number of jurisdictions have tested staff call-down procedures and POD dispensing. However, most of these drills have not been conducted on a large scale, thus limiting efforts to estimate the capability to contact all needed staff during a large-scale emergency and the ability to compare drill performance against the performance required by jurisdictions' countermeasure delivery plans. Efforts to estimate POD throughputs have been similarly hampered by the fact that most reported POD drills have used a relatively small number of clients.

Significant Growth in Capacities Suggests That CRI Has Had an Impact, but the Data Are Not Conclusive

Overall, states and MSAs participating in CRI have exhibited significant increases in TAR scores since 2006–2007. The median state TAR score rose from 85 in 2006–2007 to 95 in 2009–2010, and the average MSA score rose from 52 in 2006–2007 to 88 in 2009–2010. There was also anecdotal evidence that CRI has improved responses to real incidents and anecdotal evidence of spillover effects in the form of states using the TAR (and similar instruments) to assess non-CRI communities. The fact that greater "exposure" to CRI is associated with considerable increases in TAR scores is consistent with CRI having an effect on preparedness. It is also consistent with findings from an earlier evaluation that found growth in Cohort I TAR scores between 2004 and 2005–2006 and evidence (based on in-depth case studies) that CRI had led to increases in staff, strengthening of key partners, and other factors associated with readiness to respond to a large-scale public health emergency (Willis et al., 2009).

However, the absence of data from a representative comparison group makes it difficult to rule out the possibility that other factors drove the increases. Thus, the findings about CRI's impact must be regarded as suggestive but not conclusive. Furthermore, available data did not allow a determination of the program's impact on nonpreparedness aspects of public health, and the present study did not seek to determine whether any positive impacts justify the program's cost.

Implications and Recommendations

Although the analysis suggests that additional effort may be needed to address a number of lower-scoring functional areas, the study's reliance on secondary data and broad patterns of performance limits the ability to make detailed technical recommendations. Moreover, given earlier research highlighting the community-specific nature of CRI readiness (see, e.g., Nelson, Chan, et al., 2010; Nelson, Willis, et al., 2010; Willis et al., 2009), it is not clear that there is a one-size-fits-all approach to improving CRI capacities and capabilities. Thus, this report's recommendations focus on improving systems for (1) measuring and improving capacities and capabilities and (2) enhancing accountability, which can apply to all aspects of countermeasure delivery.

It is important to note that these recommendations must be evaluated in light of recent reductions in funding for preparedness. Thus, improvements in the ability to measure and evaluate CRI preparedness must be balanced against the costs and burdens that enhanced measurement efforts inevitably bring.

Recommendation 1: Attempt to Validate the TAR

Given decisionmakers' (and the present study's) heavy reliance on the TAR as a measure of CRI readiness, efforts to systematically validate the TAR seem warranted. First, it is particularly important to assess the extent to which TAR scores represent actual variations in communities' preparedness (i.e., genuinely prepared communities score higher, less-prepared ones lower). The relative infrequency of large-scale responses makes this difficult, but it may be possible to compare TAR scores with performance on exercises (including, but not limited to, the drill-based measures) and with the quality of responses to more-frequent incidents, such as outbreaks of food- or water-borne diseases. Second, it is important to systematically confirm that the TAR scores assigned to different states and planning jurisdictions are truly comparable (i.e., that equally prepared communities get similar scores). This might involve, but should not be limited to, assessing whether different evaluators assign the same scores to the same jurisdiction (i.e., assessing inter-rater reliability).

Recommendation 2: Continue Refining the Drill-Based Measures by Requiring Jurisdictions to Conduct Drills at a More Realistic Size and Scale

Given that many jurisdictions' TAR scores are approaching the maximum value of 100, it will be important to continue developing the drills as a way to gauge further progress. In keeping with Homeland Security Exercise and Evaluation Program guidelines, CDC encourages jurisdictions to conduct increasingly difficult drills that lead to full-scale exercises. Perhaps because of the burdens associated with conducting large exercises, many of the drills are conducted at a smaller scale than would be required by the CRI scenario. Although the number of drill-based measures reported to CDC represents an important accomplishment, the following additional refinements could significantly improve the drills' utility as measures of the ability to deliver medications at the throughput required by the CRI scenario:

- *Call-down drills*. CDC should consider additional efforts to encourage states and planning jurisdictions to conduct and report on at least some call-down drills that call the entire POD volunteer list. Smaller-scale tests have provided jurisdictions with the opportunity to test their procedures, but they do not test how these procedures would operate when scaled up, and they do not provide a good measure of the availability of the much larger number of people who would need to be mobilized in a CRI scenario. Testing the entire POD list would provide confirmation of the currently self-reported number of POD staff recruited. Further, the finding that acknowledgment percentages were lower for longer call-down lists (see Chapter Four) suggests that jurisdictions may find that the availability of volunteers will be reduced compared with the availability of core emergency staff. Implementing this recommendation would not require changes to the staff call-down drill template. However, it would impose additional burdens on states and planning jurisdictions.
- *Dispensing drills*. The analysis in Chapter Four suggests that few dispensing drills involve enough clients to provide a valid measure of the jurisdiction's maximum dispensing capability. Recruitment of sufficient numbers of clients for a dispensing drill is vital for determining the maximum throughput of the PODs. CDC should also consider additional efforts to ensure that PODs in at least some exercises more closely resemble those that would be implemented in a CRI scenario—in terms of their procedures, size, staffing, and throughput—so that drill performance can be compared against the performance needed in an actual response. If this is not practical, then "scale-model" PODs could be used. That is, jurisdictions might exercise a half-scale POD, with half the targeted

throughput and using half the usual client care staff.[1] Similarly, data on individual step processing times are important for quality improvement. Encouraging or requiring more jurisdictions to collect such data, along with information on POD layouts, would enable the use of computer modeling to help improve POD operations, thereby providing a powerful tool for benchmarking and improvement.

Recommendation 3: Improve Performance Feedback to Jurisdictions and Develop Stronger Improvement Tools

Several stakeholders perceived the need for additional performance feedback to jurisdictions in convenient, easy-to-use formats and for tools that would further assist them in closing the performance gaps revealed through the TAR, drills, and exercise submissions. The small sample size of stakeholders makes this finding tentative, and DSNS provides considerable technical assistance to program participants (see Appendix A). However, further attempts to understand the sources of these perceptions seem warranted. Specifically, CDC should consider reviewing its current approach to providing performance feedback in order to better understand the extent and sources of this perceived deficiency. Depending on the results of the review, CDC might consider additional efforts to raise awareness of, or revise, existing tools and consultant feedback processes. If gaps remain, new tools and processes can be developed, tested, and rolled out.

Recommendation 4: Seek to Leverage Assessments of Non-CRI Sites as a Comparison Group

The fact that many states are assessing non-CRI jurisdictions using the TAR (or similar instruments) was presented as evidence of CRI's spillover effects on nonparticipating jurisdictions (see Chapter Five). If similar data from a more representative range of jurisdictions were available to CDC, an opportunity to draw systematic performance comparisons between CRI and non-CRI sites might arise. Thus, CDC should consider efforts to encourage states to collect data on a broader range of non-CRI communities. As noted earlier, the fact that CRI sites were selected largely by size and risk means that the comparison group would not be fully equivalent. But this challenge could be at least partially addressed by comparing changes over time in the two groups and by collecting information on the extent to which non-CRI sites have been exposed to CRI.

Recommendation 5: Assess Cost-Effectiveness

This report has documented what the CRI program has accomplished, but the study did not seek to determine whether those accomplishments were worth the concomitant investment of public resources. Attention to cost-effectiveness in future studies might be valuable to decisionmakers seeking to intelligently allocate scarce resources.

---

[1] Information on the actual scale tested would have to be recorded and reported so that the resulting data could be properly interpreted.

APPENDIX A

Additional Information on the Cities Readiness Initiative

The CRI program consists of four key components: dedicated funding, a planning scenario, assessment at the state and local levels, and technical assistance. This appendix describes each of these components in greater detail.

Dedicated Funding

As with other monies awarded through the PHEP Cooperative Agreement, CRI funds are provided to state health departments, which in turn use those funds to support local and regional entities. Currently, the PHEP Cooperative Agreement mandates that states provide 75 percent or more of the funds to a fiscal agent (usually a local health department) in the CRI-funded metropolitan area;[1] however, the requirement was 51 percent during the period covered in this report. Four sites (New York City, Los Angeles, Chicago, and Washington, D.C.) receive CRI funds directly and function, for award purposes, very much like states.

As noted in Chapter One, CRI participants entered the program in three cohorts, listed below. In 2005, 15 Cohort II MSAs were added; 36 Cohort III MSAs were added in 2007. Currently, a total of 72 MSAs (listed at the end of this appendix) participate in CRI, covering an estimated 57 percent of the U.S. population.

The Planning Scenario

As noted earlier, the CRI program goal is to improve cities' readiness to dispense countermeasures to all individuals in a metropolitan area within 48 hours of the federal decision to deploy assets. This goal was originally derived from a planning scenario involving a large-scale, outdoor aerosolized release of *bacillus anthracis*, the agent that causes inhalational anthrax. Although anthrax is not contagious, it poses a challenge in that, by the time individuals are symptomatic, the probability of death approaches 100 percent (see, e.g., Pile et al., 1998). However, administration of antibiotics before the appearance of symptoms (which usually manifest within 48–72 hours after exposure) may prevent occurrence of the disease. In consultation with experts in the field, HHS determined early in the program that reaching the 48-hour goal would likely prevent at least 95 percent of cases of inhalational anthrax. In its 2011 program guidance, CDC expanded the scenario, stating that, based on new lessons learned from

[1] Fiscal agents can be part of an emergency management agency or can consist of multiple municipalities working together.

the H1N1 response and elsewhere, CRI supports medical countermeasure delivery "for *all-hazards events*, which includes the ability of jurisdictions to develop capabilities for U.S. cities to respond to a large-scale biologic attack, with weaponized anthrax as the primary threat consideration" (Centers for Disease Control and Prevention, 2011).

The delivery and use of medical countermeasures from the SNS include several steps. First, states (or local health departments working through their states) request SNS materiel from CDC. CDC then delivers materiel to a state-designated medical warehouse or RSS facility. In accordance with the medical materiel distribution plan of the state or directly funded city, the materiel is then placed onto trucks for distribution to the PODs, where medications are dispensed to individuals. (Note that CDC distinguishes between the *distribution* of medications to PODs and the *dispensing* of medications to individuals.) POD facilities are often located within high school gymnasiums, armories, or other facilities with large rooms.

Assessment at the State and Local Levels

The third key component of the CRI program is its requirements for assessment and accountability. The TAR is a tool developed by CDC to assess the extent to which a jurisdiction has completed planning in core functional areas (13 in the case of states and directly funded cities, 12 in the case of local planning jurisdictions). Details are provided in Chapter Two.

Technical Assistance

The final component of the CRI program is the technical assistance provided by CDC program services consultants. CRI is notable among CDC programs in that it provides technical assistance directly to local and regional entities, in addition to state health departments and directly funded cities. The consultants assist in identifying resources and opportunities for training. They also conduct the TAR in approximately 25 percent of the planning jurisdictions in each CRI MSA (with the state conducting the remaining 75 percent). Scoring the TAR requires the application of considerable judgment in determining whether awardees have met specific program requirements. CDC DSNS utilizes a number of venues, including educational webcasts, guidebooks, training classes, and other forms of technical assistance to provide information and share lessons learned from federal, state, and local perspectives. Awardees are also provided with written guidance materials that include specific details on the core functional areas and the related capabilities required to receive, distribute, and dispense SNS materiel.

MSAs Within CRI Cohorts

Cohort I (2004)	Cohort II (2005)	Cohort III (2006)
Atlanta, GA	Baltimore, MD	Albany, NY
Boston, MA	Cincinnati, OH	Albuquerque, NM
Chicago, IL	Columbus, OH	Anchorage, AK
Cleveland, OH	Indianapolis, IN	Baton Rouge, LA
Dallas, TX	Kansas City, MO	Billings, MT
Denver, CO	Milwaukee, WI	Birmingham, AL
Detroit, MI	Orlando, FL	Boise, ID
Houston, TX	Portland, OR	Buffalo, NY
Las Vegas, NV	Providence, RI	Burlington, VT
Los Angeles, CA	Riverside, CA	Charleston, WV
Miami, FL	Sacramento, CA	Charlotte, NC
Minneapolis, MN	San Antonio, TX	Cheyenne, WY
National Capital Region (DC)	San Jose, CA	Columbia, SC
New York City, NY	Tampa, FL	Des Moines, IA
Philadelphia, PA	Virginia Beach, VA	Dover, DE
Phoenix, AZ		Fargo, ND
Pittsburgh, PA		Fresno, CA
San Diego, CA		Hartford, CT
San Francisco, CA		Honolulu, HI
Seattle, WA		Jackson, MS
St. Louis, MO		Little Rock, AR
		Louisville, KY
		Manchester, NH
		Memphis, TN
		Nashville, TN
		New Haven, CT
		New Orleans, LA
		Oklahoma City, OK
		Omaha, NE
		Peoria, IL
		Portland, ME
		Richmond, VA
		Salt Lake City, UT
		Sioux Falls, SD
		Trenton, NJ
		Wichita, KS

Technical Detail on the Study's TAR Analysis

This appendix provides technical detail on the study's analysis of variations in MSAs' 2010 TAR scores and of growth in MSAs' TAR scores.

Explaining Variation in 2010 MSA-Level TAR Scores

To provide insight into the mechanisms through which CRI works, and to develop guidance about where technical assistance and other forms of support should be targeted, multivariate regression analysis was used to explain variations in 2010 MSA-level TAR scores. Variables included in the analysis were selected based on a review of the literature on public health systems and services (e.g., Mays, Halverson, et al., 2004; Mays, Scutchfield, et al., 2010) and on earlier research on the CRI program (e.g., Nelson, Willis, et al., 2010; Willis et al., 2009).

State TAR scores were included in the analysis because an earlier qualitative study of CRI (reported in Nelson, Chan, et al., 2010; and Willis et al., 2009) found that MSAs benefitted from strong planning and clear leadership from state health departments. *Governance structure* was included because the same study found that MSAs working in decentralized governance structures appeared to experience more difficulties in meeting program objectives. A variable representing the three CRI *cohorts* was included to account for the possibility that length of exposure to the program affects performance. Finally, variables representing *population demographics* and region were included because researchers have found that these factors explain some of the variation in the quality of other public health services (e.g., Mays, Halverson, et al. 2004; Mays, Scutchfield, et al., 2010).

Inspection of the data revealed high multicollinearity of cohort and demographics with the remaining variables, so the cohort and demographics variables were dropped from the analysis. The cross-sectional model estimated is represented by

$$TAR_i^3 = \beta_0 + \beta_2 S_i + \beta_3 G_i + \beta_4 R_i + \varepsilon_i,$$

where $TAR_{i,2010}^3$ is the overall 2010 TAR score for a given MSA i. The variable was cubed to address skewness. S is a set of variables representing the performance of states on the state version of the TAR, G is a set of variables representing governance structures in the states involved in each MSA, R is a set of variables representing geographic region, and ε_i is residual variation.

The variables were operationalized as follows. State performance is the overall TAR score for the state containing the MSA.[1] The model also included a variable representing change in state scores. Governance was measured using a scale ranging from –4 (most decentralized) to +4 (most centralized). The scale was derived by subtracting the number of functions (of four) for which any local agency has authority from the number of functions (of four) for which the state has authority. (Information on the functions was taken from the National Association of County and City Health Officials' 2008 profile of local health departments; see Parker et al., 2012).[2] Region was measured with a set of dichotomous variables coded 1 if a given MSA is in each of the ten HHS regions and 0 if not (the excluded category is HHS Region 1).[3]

Results are provided in the "Model 1" column of Table B.1.[4] See Chapter Three for an interpretation of the findings. Note that the coefficients are interpreted as the marginal effect of a one-unit increase in the independent variable on the cube of TAR scores.

Assessing Variations in MSA-Level TAR Score Growth

Another multivariate regression analysis sought to explain variations in TAR scores, after accounting for earlier year TAR scores. Detailed discussions with DSNS staff about TAR data collection revealed two concerns. First, a number of MSAs with high scores in 2008–2009 were allowed to "carry over" their scores into 2009–2010 (see Chapter Three for details). This makes scores between those two years difficult to compare. Second, there were concerns that data collection in 2006–2007 was not standardized enough to allow accurate conclusions to be drawn about that year. Thus, the analysis focuses on the difference between 2007–2008 and 2009–2010 TAR scores. The estimated model is represented in

$$TAR_{it}^3 = \beta_0 + \beta_1 TAR_{i,previous} + \beta_2 G_i + \beta_3 S_i + \beta_4 R_i + \varepsilon_i ,$$

where TAR_i^3 is the cubed overall 2010 TAR score for a given MSA i. $TAR_{i,previous}$ is the overall TAR score from a previous year (e.g., 2007–2008), G is a variable representing governance structures in the states involved in each MSA, S is a variable representing S-TAR scores of states included in the MSA, R is a set of variables representing geographic region, and ε_i is residual variation. The coefficient β_1 provides the extent to which TAR scores in 2010 depend on where they were in 2008, after accounting for other variables in the model, and the coefficients β_2, β_3, and β_4 show systematic variation in mean MSA TAR scores associated with variations in

[1] For the 17 MSAs that involve more than one state (e.g., the Kansas City MSA includes portions of both Missouri and Kansas), the score was derived by averaging overall TAR scores for each state, weighted by the number of planning jurisdictions in each.

[2] An alternate model included a variable representing whether MSAs were in states with substate regional structures (Wasserman et al., 2006). Including the variable did not add to the proportion of explained variance (as determined by an F-test) and did not appreciably change any of the remaining coefficients. Thus, it was dropped in the interest of parsimony.

[3] For a map of the HHS regions, see Department of Health and Human Services, undated. Table B.1 compares each HHS region with HHS Region 1, determining whether they are significantly different. Additional models were estimated using each HHS region as the baseline (with similar results).

[4] Regressions of the independent variables against all other variables did not reveal problems with high multicollinearity. Inspection of the residuals did not reveal any serious deviations from the assumption of homoskedastic errors, and models using Huber-White robust standard errors produced similar findings.

Table B.1
Modeling Variations in TAR Score Levels and Changes

Explanatory Variable	Model 1: 2010 Scores Only, Cubed	Model 2: Change from 2008 to 2010	Model 3: Change from 2009 to 2010	Model 4: Change from 2008 to 2009
Intercept	–0.73**	–0.95***	–1.02***	0.02
2008 MSA score		0.44***	0.65***	0.56***
2010 state score	1.53***			
2008 state score		1.38***	1.36***	0.12
2008–2010 change in state score		1.95***	1.2**	0.31
Centralization	0.02**	0.03***	0.02*	0.03***
HHS Region 2	0.16*	0.16*	–0.01	0.37***
HHS Region 3	–0.07	–0.8	–0.13*	0.11
HHS Region 4	–0.05	–0.03	–0.12*	0.16**
HHS Region 5	0.04	0.08	–0.04	0.21***
HHS Region 6	–0.15*	–0.05	–0.23***	0.27***
HHS Region 7	0.10	0.19**	–0.02	0.36***
HHS Region 8	–0.15*	–0.15**	–0.13*	–0.03
HHS Region 9	0.05	0.10	–0.07	0.30***
HHS Region 10	–0.18*	–0.18**	–0.20**	0.09
N	71	70	70	70
Adjusted R^2	0.48	0.54	0.57	0.59

NOTE: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

governance, state TAR scores, and region.[5] The variables were operationalized as described above.

Results are provided in the "Model 2" column of Table B.1. See Chapter Five for an interpretation of the findings. The results from the analysis of years 2007–2008 to 2008–2009 and years 2008–2009 to 2009–2010 (see the "Model 3" and "Model 4" columns in Table B.1, respectively) are also presented. The results across the three models were remarkably consistent, and the models explained a considerable amount of variance (adjusted R^2 was between 0.54 and 0.59). The analysis began with assessment of how much of the explained variance might be associated with the MSA's previous year's TAR score. Comparison of results from a trimmed model (which included only the previous year's TAR score) with results from the full model found that past state TAR performance accounted for less than half the variance accounted for by the full models. An interaction analysis between state TAR performance and degree of centralization was difficult to interpret because of outliers that affected the results.

[5] Direct modeling of change scores (i.e., 2009–2010 scores minus 2007–2008 scores) was considered but ultimately rejected. Literature on the subject (see Allison, 1990; Cohen and Cohen, 1983) suggests that analysis of change scores tends to underestimate effects when (1) there is any degree of unreliability in the component scores, (2) the component scores are correlated with each other, (3) the component scores have differential variance, and (4) change is correlated with the baseline score—all of which likely apply here. Thus, 2009–2010 scores were estimated as a function of 2008–2009 scores, 2008–2009 scores as a function of 2007–2008 scores, and 2009–2010 scores as a function of 2007–2008 scores (with the results provided in Table B.1). As an additional robustness check, a growth curve model (see, e.g., Raudenbush and Bryk, 2002) of 2007–2010 growth was estimated. In this analysis, the intercept (starting value) and slope (growth rate) parameters from a linear regression model that included the initial starting value (intercept) and growth rate (slope) were modeled as a function of the covariates shown in Table C.1. Results from these models (not shown here, but available upon request) were similar.

APPENDIX C

Technical Detail on the Analysis of Drill Data

This appendix contains data that supplement the information provided in Chapter Four regarding the metrics collected in drills used to assess operational capabilities.

Table C.1 provides descriptive statistics on call-down completion time, broken out by the type of list tested, notice given, and type of call-down equipment used. The categorizations for list type are not perfect: 14 percent of the drills reported testing more than one list simultaneously.

Table C.2 gives descriptive statistics on acknowledgment percentages, broken out by type of list tested, notice given, and type of call-down equipment used.

In the raw data provided by awardees, many different names were used to describe essentially similar steps. To compute statistics of processing times for the steps, we recategorized these steps into a smaller number of grouped steps. Table C.3 shows the original name for the step, the name of the new "grouped" step, and the number of drills that include the step.

Table C.1
Descriptive Statistics on Call-Down Completion Time

	N	Mean (minutes)	Median (minutes)	Std Dev (minutes)	5th Percentile (minutes)	95th Percentile (minutes)
All drills	351[a]	80	47	113	3	301
List type						
EOC	43	55	26	66	2	150
PIC	14	42	26	51	1	180
RSS	8	89	45	139	1	420
Security	6	182	78	219	9	540
Distribution	18	113	60	171	1	720
POD	75	69	46	70	6	180
Other	121	96	60	143	3	390
Tested more than one list	53	69	45	83	3	240
Notice						
Full	32	72	60	65	0	198
Partial	64	80	46	101	3	260
None	256	80	44	121	2	330
Equipment						
Automatic	159	88	60	130	1	360
Manual	188	70	35	95	4	180
Other	11	96	44	125	1	427

NOTES: EOC = emergency operations center. PIC = public information and communication.

[a] A total of 368 drills were reported. However, sample sizes in this table account for drills that were missing data on one or more drill attributes. The table also excludes seven cases with reported call-down times of less than zero.

Table C.2
Descriptive Statistics on Acknowledgment Percentages

	N	Mean (%)	Median (%)	Std Dev (%)	5th Percentile (%)	95th Percentile (%)
All drills	367[a]	78	83	24	25	100
List type						
EOC	47	86	91	17	45	100
PIC	14	92	100	12	69	100
RSS	8	66	79	34	20	100
Security	8	61	53	30	29	100
Distribution	19	83	89	23	12	100
POD	81	75	75	23	33	100
Other	125	78	83	24	25	100
Tested more than one list	53	73	79	26	19	100
Notice						
Full	36	72	78	25	14	100
Partial	67	78	81	24	29	100
None	259	79	86	23	25	100
Equipment						
Automatic	164	69	72	24	20	100
Manual	193	86	92	20	38	100
Other	11	74	27	27	10	100

[a] A total of 368 drills were reported. However, sample sizes in this table account for drills that were missing data on one or more drill attributes. The table also excludes seven cases with reported call-down times of less than zero.

Table C.3
Crosswalk Between Original POD Step Names and New Grouped Tasks

Original Step Name	Grouped Step (if applicable)	Number of Drills Testing This Step
Approval		1
Assessment		1
Check-in	Registration	3
Checkout	Exit	1
Collection		3
Demographics	Survey	5
Discharge	Exit	1
Dispensing	Dispensing	68
Dispensing/vaccination	Vaccination	1
Dispensing–express	Dispensing	9
Dispensing–family special needs	Dispensing	1
Dispensing–general	Dispensing	1
Dispensing–injection	Vaccination	3
Dispensing–mist	Dispensing	3
Dispensing–mist/injection	Vaccination	1
Dispensing–pills	Dispensing	1
Dispensing–special needs	Dispensing	2
Drive-through vaccinations	Vaccination	1
DUR		1
Education	Education	4
Entrance	Entry	1
Entry	Triage	1
Exit	Exit	3
Exit/education	Exit	1
Express dispensing	Dispensing	1
Family		1
Flow		1
Flow control		1
Form collection	Registration	1
Form completion	Registration	4
Form completion/medical screening	Registration	1

Table C.3—Continued

Original Step Name	Grouped Step (if applicable)	Number of Drills Testing This Step
Form completion/registration	Registration	1
Forms	Registration	33
Forms distribution	Registration	1
Forms review	Registration	2
Forms/screening	Registration	5
Forms/triage	Registration	1
Forms–checking	Registration	1
Forms–collection	Registration	1
Forms–completion	Registration	2
Forms–distribution	Registration	1
Forms–review	Registration	4
Greeting/screening	Screening	1
Greeting	Entry	10
Greeting/forms	Registration	4
Greeting/forms review	Registration	1
Greeting/registration	Registration	1
Greeting/screening	Screening	7
Greeting/triage	Triage	4
Greetings/forms	Registration	1
Greetings/triage	Triage	1
Griage	Triage	3
Handicapped		1
Help		1
Insurance		2
Intake	Entry	1
LAIV express	Vaccination	1
Med eval	Medical consult	3
Med screening	Medical consult	2
Medical consult	Medical consult	1
Medical evaluation	Medical consult	1
Medical screening	Medical consult	1
Observation		2

Table C.3—Continued

Original Step Name	Grouped Step (if applicable)	Number of Drills Testing This Step
Oversize vehicle vaccinations	Vaccination	1
Client tracking	Registration	1
Payment		3
Pediatric vaccinations	Vaccination	1
Reception	Registration	1
Reception/triage	Triage	3
Recovery		1
Registration	Registration	26
Registration, screening	Registration	2
Registration/forms	Registration	3
Registration/greeting	Registration	1
Registration/screening	Registration	1
Registration/screening/triage	Registration	1
Registration/triage	Registration	1
Review/screening	Screening	1
Screening	Screening	23
Screening/dispensing/observation/checkout	Dispensing	1
Screening/forms	Screening	3
Screening/forms/vaccine	Vaccination	1
Screening/triage/dispensing	Dispensing	1
Screening1	Screening	1
Special needs	Dispensing	4
Screening/vaccination	Vaccination	1
Survey	Survey	1
Survey/exit	Exit	1
Tally	Survey	1
Tickets		1
Triage	Triage	15
Triage/dispensing	Dispensing	2
Triage/forms	Registration	2
Vaccination	Vaccination	46
Vaccination/families	Vaccination	1

Table C.3—Continued

Original Step Name	Grouped Step (if applicable)	Number of Drills Testing This Step
Vaccine triage	Screening	1
Waiting		3
Waiting area		1
WIC		1

NOTE: DUR = drug utilization review. LAIV = live attenuated influenza vaccine. WIC = women, infants, and children.

APPENDIX D

Technical Detail on Stakeholder Discussions

Nine CRI sites were selected for stakeholder discussions based on patterns of TAR score growth. TAR score patterns were divided into four categories:

1. low growth over time, which often corresponded to higher scores at the initial assessment
2. average growth over time
3. high growth over time, which often corresponded to lower scores at the initial assessment
4. unusual pattern over time, which was characterized by reversals in trends over time.

After RAND selected sites, DSNS contacted potential participants via an introductory email describing the study and requesting participation. Members of the RAND team then contacted the sites that agreed to participate, and all subsequent contact went directly through RAND. Participating sites are listed here:

- Albany, New York
- Baton Rouge, Louisiana
- Dover, Delaware
- Miami, Florida
- New Haven, Connecticut
- New York, New York
- Pittsburgh, Pennsylvania
- Portland, Maine
- Salt Lake City, Utah

Data were generated through discussions with sites' CRI coordinators and related staff and through analysis of site-level documents. Where possible, discussions were conducted in person at the SNS Summit in Atlanta, Georgia, in July 2011; other discussions were conducted via telephone between July and September 2011. Additional data came from feedback provided by approximately 50 self-selected state and local CRI officials at an interactive meeting hosted by RAND at the SNS Summit. CDC officials were present to introduce the discussion session, but neither CDC officials nor any other federal partners were present for the discussion. Discussion guides for the interviews and interactive session are reproduced in this appendix.

Written notes were taken during the nine case study discussions and during the facilitated group discussion that occurred at the summit. The notes were analyzed using qualitative data analysis software (ATLAS.ti). The notes were coded according to interview/discussion guide structure and by other key themes that emerged during analysis.

Approval for the qualitative data collection was granted by RAND's Human Subjects Protection Committee. Although CDC was aware of the names of case study participants, and although the sites are listed in this report, no specific findings are associated with a specific site or respondent. Since the purpose of the study was to assess the CRI program and not the level of preparedness in any particular planning jurisdiction, this lack of association does not greatly limit the conclusions that can be drawn from the analysis. In some instances, however, the protections slightly limited our ability to fully document the evidence behind our findings.

Guide for Small-Group Discussions

Thank you for agreeing to talk with us about public health and emergency preparedness. Before we begin, let me assure you that your responses to these questions will be held in strict confidence, except as required by law. Summary information from these interviews, together with material taken from public documents, will be presented to the Centers for Disease Control and Prevention (CDC), as part of a project evaluating the Cities Readiness Initiative. The names of the sites being contacted will be included in the reports; however, no observations or statements will be attributed to specific individuals, organizations, cities, counties or states. Study results will be made available to each site upon request. The notes from our discussions and reports and plans we collect from you may be used in other RAND studies of public health emergency preparedness. However, if data is retained for this purpose, then individuals' names and organizations will be removed from the notes. Your participation in this discussion is completely voluntary. We would like to have your responses to all of the questions; however, if you are uncomfortable with any question we can skip it. We estimate that the meeting will take no more than sixty minutes.

Do you have any questions about our confidentiality procedures before we begin? (If yes, respond to all questions. If no, proceed with discussion.)

Personal history

- Discuss interviewee's background, training, position, tenure, and experiences with PHEP

Recent History

- What have your CRI efforts looked like in the past few years?
  - What have you been focusing on developing/improving? Is it call-down lists? POD protocols? Supply chain? Other?

CRI Impact

- H1N1 and other real world experience (seek specific examples)
  - Aside from the H1N1 pandemic, has your jurisdiction responded to any real events that leveraged CRI capabilities? [If there are no H1N1 events, follow-up with: Did the H1N1 pandemic response require your jurisdiction to leverage its CRI capabilities?]
  - What were some key success factors or challenges?

Guide for Small-Group Discussions—Continued

[Listen for references to CRI program elements (technical assistance, TAR, funding, scenario focus) and follow up accordingly]

- Have any of these success factors of challenges been influenced by CRI? How has the CRI program influenced them (either positively or negatively)?

Measurement

- TAR
  - How well do you think your jurisdiction's TAR scores reflect the level of preparedness in your jurisdiction?
  - Are there sections of the TAR that are better or worse at capturing your jurisdiction's level of preparedness?
  - Over time, TAR scores in your jurisdiction have [been steady, increased, variable, started off high/low]. What contextual information might help explain [these changes or this pattern]?
- Drills

[If respondent is not in a position to address the issue of drills, ask if we can be directed to the correct individual]

- Do you regularly conduct drills? Which ones?
- How well do you think your jurisdiction's performance on the drills reflects your preparedness capabilities?
- Have there been any key lessons learned from the drills?
- Have drills led to any changes in SNS-related plans?

Future challenges

- What are your top 3 current challenges in meeting CRI program goals?
- How are you trying to address those challenges?
- What barriers/support are impacting your efforts to address the challenges?

Guide for Group Discussion

1. Questions/comments on RAND CRI study
2. Aside from H1N1 response, what types of real events have leveraged CRI capabilities?
3. If you could design the next generation program (CRI 2?) for mass prophylaxis/treatment, how would it look different than the current CRI program?

If there is time:

4. Have CRI capabilities led to greater all-hazards preparedness? Are there specific examples you could share (we might need to offer to follow up individually to save time)
5. Has CRI experience led to greater preparedness in non-CRI jurisdictions? Are there specific examples you could share (we might need to offer to follow up individually to save time)

Back-up question:

6. Have there been any ways in which you have been able to learn from other CRI participants' experiences?

References

Allison, Paul D., "Change Scores as Dependent Variables in Regression Analysis," *Sociological Methodology*, Vol. 20, 1990, pp. 3–114.

Centers for Disease Control and Prevention, "Public Health Emergency Preparedness Cooperative Agreement Budget Period 10 Extension (FY 2010) Funding," revised, August 2010. As of January 6, 2012: http://www.cdc.gov/phpr/documents/Revised_PHEP_BP10_Extension_Funding_Table_Aug2010.pdf

Centers for Disease Control and Prevention, "RFA/PA Number: CDC-RFA-TP11-1101cont11," final version, June 13, 2011. As of January 6, 2012: http://www.cdc.gov/phpr/documents/PHEP_FY_2011.pdf

Centers for Disease Control and Prevention, Division of Strategic National Stockpile, "Local Technical Assistance Review," form, undated-a. As of January 6, 2012: http://www.docstoc.com/docs/77847714/Local_Technical_Assistance_Review_Tool

Centers for Disease Control and Prevention, Division of Strategic National Stockpile, "SNS-Related Public Health Emergency Preparedness Drills," factsheet, undated-b. As of January 6, 2012: http://www.bt.cdc.gov/cotper/coopagreement/08/pdf/SNSDrillsFactSheet.pdf

Chen, Huey Tsyh, "Theory-Driven Evaluations: Need, Difficulties, and Options," *American Journal of Evaluation*, Vol. 15, 1994, pp. 79–82.

Cohen, J., and P. Cohen, *Applied Multiple Regression/Correlation for the Behavioral Sciences*, Hillsdale, N.J.: Lawrence Erlbaum, 1983.

Department of Health and Human Services, "HHS Region Map," web page, undated. As of January 6, 2012: http://www.hhs.gov/about/regionmap.html

Hallmark, Bryan W., and James C. Crowley, *Company Performance at the National Training Center*, Santa Monica, Calif.: RAND Corporation, MR-846-A, 1997. As of January 16, 2012: http://www.rand.org/pubs/monograph_reports/MR846.html

Mays, Glenn P., Paul K. Halverson, Edward L. Baker, Rachel Stevens, and Julie J. Vann, "Availability and Perceived Effectiveness of Public Health Activities in the Nation's Most Populous Communities," *American Journal of Public Health*, Vol. 94, No. 6, 2004, pp. 1019–1026.

Mays, Glenn P., F. Douglas Scutchfield, Michelyn W. Bhandari, and Sharla A. Smith, "Understanding the Organization of Public Health Delivery Systems: An Empirical Typology," *Milbank Quarterly*, Vol. 88, No. 1, 2010, pp. 81–111.

Nelson, Christopher, Edward W. Chan, Anita Chandra, Paul Sorensen, Henry H. Willis, Stephanie Dulin, and Kristin Leuschner, "Developing National Standards for Public Health Emergency Preparedness with a Limited Evidence Base," *Disaster Medicine and Public Health Preparedness*, Vol. 4, No. 4, December 2010, pp. 285–290.

Nelson, Christopher, Edward W. Chan, Carol E. Fan, Debra Lotstein, Leah B. Caldarone, Shoshana R. Shelton, Amy L. Maletic, Andrew M. Parker, Alexandria Felton, Amanda Pomeroy, and Elizabeth M. Sloss, *New Tools for Assessing State and Local Capabilities for Countermeasure Delivery*, Santa Monica, Calif.: RAND Corporation, TR-665-DHHS, 2009. As of January 6, 2012: http://www.rand.org/pubs/technical_reports/TR665.html

Nelson, Christopher, Henry H. Willis, Edward W. Chan, Shoshana R. Shelton, and Andrew M. Parker, "Federal Initiative Increases Community Preparedness for Public Health Emergencies," *Health Affairs,* Vol. 29, No. 12, 2010, pp. 2286–2293.

Parker, Andrew M., Shoshana R. Shelton, Kristy Gonzalez Morganti, and Christopher Nelson, "Assessing Relationships Between State and Local Public Health Organizations: Evidence From the NACCHO 2008 Profile of Local Health Departments," *Journal of Public Health Management and Practice*, Vol. 18, No. 2, 2012, pp. 156–159.

Pile, James C., John D. Malone, Edward M. Eitzen, and Arthur M. Friedlander, "Anthrax as a Potential Biological Warfare Agent," *Archives of Internal Medicine*, Vol. 159, 1998, pp. 429–434.

Public Law 109-417, Pandemic and All-Hazards Preparedness Act (PAHPA), December 19, 2007.

Raudenbush, Stephen W., and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis*, 2nd ed., Thousand Oaks, Calif.: Sage Publications, 2002.

Wasserman, Jeffrey, Peter Jacobson, Nicole Lurie, Christopher Nelson, Karen A. Ricci, Molly Shea, James Zazzali, and Martha I. Nelson, *Organizing State and Local Health Departments for Public Health Preparedness*, Santa Monica, Calif.: RAND Corporation, TR-318-DHHS, 2006. As of January 16, 2012: http://www.rand.org/pubs/technical_reports/TR318.html

Willis, Henry H., Christopher Nelson, Shoshana R. Shelton, Andrew M. Parker, John A. Zambrano, Edward W. Chan, Jeffrey Wasserman, and Brian A. Jackson, *Initial Evaluation of the Cities Readiness Initiative*, Santa Monica, Calif.: RAND Corporation, TR-640-CDC, 2009. As of September 29, 2011: http://www.rand.org/pubs/technical_reports/TR640.html