This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL
REPORT

Enhancing Public Health Preparedness: Exercises, Exemplary Practices, and Lessons Learned

Assessing the Adequacy of Extant Exercises for Addressing Local and State Readiness for Public Health Emergencies

Lisa R. Shugarman, Elisa Eiseman, Arvind Jain,
Nancy Nicosia, Stefanie Stern, Jeffrey Wasserman

Prepared for the
Office of the Assistant Secretary for Public Health Emergency Preparedness

HEALTH

The research described in this report was carried out under the auspices of the RAND Center for Domestic and International Health Security, a program within RAND Health, and sponsored by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Public Health Emergency Preparedness.

Published 2005 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
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

Over the past three years, the Department of Health and Human Services (DHHS) has made significant investments in state and local public health in an effort to enhance public health emergency preparedness. The RAND Corporation was contracted to work with the U.S. Department of Health and Human Services Office of the Assistant Secretary for Public Health Emergency Preparedness (OASPHEP) to develop resources and to prepare analyses to help describe and enhance key aspects of state and local public health emergency preparedness. As part of this contract, RAND was asked to compile a repository of exercises used at the local, state, regional, national, or international level that could be made available on the HHS/OASPHEP website. In addition to collecting this set of exercises, RAND was asked to develop a set of criteria by which exercise design could be evaluated, and then apply those criteria to actual exercises. These criteria are designed to serve as a template for public health officials in evaluating potential exercises to be used in their local areas. This report provides an overview of the criteria development process, methods for evaluating criteria design, methods for evaluating exercise design, and the results of our analyses. This work was carried out from October 2003 through January 2005.

This report was prepared specifically for the Office of the Assistant Secretary for Public Health Emergency Preparedness. The report is comprehensive: it covers our methodology, analyses, and results. We envision that a portion of this report (for example, much of the information in the appendices, accompanied by brief explanatory material and links to the exercise materials) would be of interest to individuals working in public health preparedness at the federal, state, and local levels. At this time, however, we have been asked by the individuals who supplied the exercises to maintain confidentiality; accordingly, all exercises have been de-identified in this report. When the confidentiality issue is resolved, DHHS will be able to make the criteria and exercise materials more widely available.

Comments or inquiries should be sent to the RAND Principal Investigators Nicole Lurie (Nicole_Lurie@rand.org) and Jeffrey Wasserman (Jeffrey_Wasserman@rand.org) or addressed to the first author of this report, Lisa Shugarman (Lisa_Shugarman@rand.org).

This work was sponsored by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Public Health Emergency Preparedness and carried out under the auspices of the RAND Center for Domestic and International Health Security, a program within RAND Health. For more information about the RAND Center for Domestic and International Health Security, please visit http://www.rand.org/health/healthsecurity/. The mailing address is RAND Corporation, 1200 South Hayes Street, Arlington VA 22202. More information about RAND is available at www.rand.org.

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

The use of emergency preparedness exercises is now widespread throughout the public health and responder communities. Exercises vary considerably in complexity and may be designed to meet one or more of a variety of objectives, including training, gap analysis, planning, and team building. Ideally, as with any quality improvement cycle, preparedness exercises are used to identify areas where improvement is needed; to inform the process of developing plans and procedures to improve performance; and finally to enable repeat testing once improvements have been made. This process is critical to achieving the long-term goal of conducting preparedness exercises in public health: to mitigate the morbidity, mortality, psychosocial stress, and social consequences of a terrorist attack or other public health emergency.

Task 4 of the project entitled *Enhancing Public Health Preparedness: Exercises, Exemplary Practices and Lessons Learned* requires that the RAND team identify and compile information regarding existing preparedness exercises and provide a preliminary critique of their design. These exercises focus on assessing and improving the readiness of local and state public health departments to respond to public health emergencies, including bioterrorism. The Task 4 results have produced tools (criteria) that can be used by DHHS and others, including state and local public health officials, to evaluate current and future exercises. DHHS requested that the criteria be broadly applicable, clearly defined, measurable, and designed to evaluate both substance and procedure. This document presents the results of our large-scale application of the final criteria to exercises suitable for evaluation (exercises that met minimum documentation requirements). From this evaluation, we provide information about the utility of the criteria for evaluating these and future exercises and the appropriateness of exercise design.

The specific questions addressed in this report are:

1. What is the feasibility of these criteria?
    a. To what extent are data available to rate the exercises?

   b. Are ratings of the criteria sufficiently variable to distinguish among exercises?
2. What is the reliability of these criteria?
   a. To what degree do evaluators agree on their rating of each criterion?
   b. Are the criteria internally consistent?
      i. Is there sufficient internal consistency to justify an overall score?
      ii. Is there sufficient internal consistency to justify scores for each criterion domain?
3. What is the validity of these criteria?
   a. Is there a relationship between criteria performance and type of exercise?
   b. Is there a relationship between criteria performance and number of documents available for evaluation?
   c. Is there a relationship between criteria performance and the type of documents available for evaluation?
4. How well designed are the exercises we reviewed?

An initial set of 20 criteria were developed and tested in an iterative process. Criteria fall into five separate domains: (1) Goals and Objectives; (2) Scenario; (3) Participation; (4) Materials; and (5) Execution and Feedback. Revisions were made to the criteria, scoring instructions, and guidance, reducing the list of criteria from 20 to 14. We obtained information describing 100 exercises, simulations and drills. Exercises ranged from tabletop to full-scale field exercises. Exercises that did not include a minimum amount of documentation (an after-action report or an exercise plan, plus at least one other document) were excluded from review, leaving 37 exercises appropriate for evaluation. Four to six evaluators rated the 37 exercises, using only the written materials available to the RAND project team. No exercise sponsors, participants, or developers were interviewed or otherwise contacted during this evaluation.

Criteria were evaluated for feasibility, reliability, and validity. Feasibility is a measure of whether there are enough data to rate the exercise and to distinguish among exercises. We found that the criteria we developed are reasonably feasible, with good dispersion across possible response categories and only a modest effect from missing information (i.e., the documentation was incomplete). Two criteria (#7 and 14) did not

demonstrate sufficient variability across response categories. However, overall, we found that there is sufficient variation across response categories, thereby enabling us to distinguish among exercises.

Additionally, the criteria are reliable and internally consistent, with an overall Cronbach's alpha (a widely used measure of internal consistency) of 0.87 (alpha > 0.80 is considered good internal consistency) and domain-specific Cronbach's alphas ranging from 0.82 to 0.85. The inter-rater reliability (i.e., the extent to which the various evaluators agreed with each other) for most criteria was reasonable; no criterion had an inter-rater reliability score greater than 0.29 (a score of zero reflects perfect agreement). However, criteria with three to four ordinal response categories were not as reliable as those with a simple Yes/No response. Feedback from reviewers indicates that there was high variability in the quality of the documentation provided by the exercise developer or state, county, or other locality that used the exercise.

Based on expert review, the criteria have good face validity. There was no relationship between criteria performance and type of exercise, or the number or type of documents available for evaluation. Similarly, the type of exercise, the number of documents available for review and the types of documents available for review are not significantly related to reviewer evaluations, strengthening our conclusion that these criteria are reasonable and valid. We cannot test the validity of the criteria against external standards, because, as of yet, there are no "gold standard" measures on which to base a comparison.

There is substantial variation in the scores for the 37 exercises. In fact, there is a threefold difference in the performance scores of the exercises. The variability across exercises within a domain was even greater. Generally, exercises that scored in the highest tertile overall also scored high across the individual domains; conversely, those that scored in the lowest tertile overall generally scored low in all the domains.

As we have discussed with DHHS, it is intended that the exercise criteria and a menu of vetted exercises be made available to state and local public health officials and others in the preparedness community. However, the individuals who supplied the exercise materials to us asked that we maintain confidentiality. Accordingly, for purposes of this report, the exercises have been de-identified.

ACKNOWLEDGMENTS

We would like to thank the reviewers who evaluated exercises for this report: Jennifer Brower, Edward Chan, Elisa Eiseman, Brian Jackson, Sarah Myers, Nancy Nicosia, Elizabeth Sloss, Amanda Spendlove, Stefanie Stern, Ana Suarez, and Robert Valdez. In addition, the authors would like to thank Dr. Kristine Gebbie and Dr. Michael Seid for their careful review of this report.

1. INTRODUCTION

TASK SUMMARY

In Task 4, RAND identified and compiled a database of extant exercises that are relevant to assessing readiness of local and state public health departments for public health emergencies, including bioterrorism. Criteria were developed through an iterative process to evaluate the design of these exercises. These criteria needed to be broadly applicable, clearly defined and measurable, and capable of evaluating both substance and procedure. The primary consumer for these criteria and RAND's evaluation of the exercises is initially DHHS, and, in the future, state and local public health officials and bioterrorism coordinators. The evaluation results can be helpful to these agencies/individuals as they consider which exercises they might select for preparedness activities in their local areas. The final product of this effort will be a set of criteria useful to evaluating exercise design and a set of vetted exercises available for use by preparedness leaders at all levels of the public/preparedness community for training, imitation, gap analysis, self-examination, and evaluation.

There were two main objectives to this study. The first main objective of this work was to develop a set of criteria for evaluating exercise design and to determine if the criteria were feasible, reliable, and valid. The second main objective was, after demonstrating these properties of the criteria, to evaluate the performance of the various exercise designs against these criteria. In this report, we answer the following questions:

1. What is the feasibility of these criteria?
    a. To what extent are data available to rate the exercises?
    b. Are ratings of the criteria sufficiently variable to distinguish among exercises?
2. What is the reliability of these criteria?
    a. To what degree do evaluators agree on their rating of each criterion?
    b. Are the criteria internally consistent?
        i. Is there sufficient internal consistency to justify scores for each criterion domain?
        ii. Is there sufficient internal consistency to justify an overall score?
3. What is the validity of these criteria?

   a. Is there a relationship between criteria performance and type of exercise?
   b. Is there a relationship between criteria performance and number of documents available for evaluation?
   c. Is there a relationship between criteria performance and the type of documents available for evaluation?
4. How well designed are the exercises we reviewed?

Questions 1a and 1b were evaluated by examining the distribution of reviewer scores across the possible score responses to assess the variability in response and any floor or ceiling effects. Question 2a was evaluated by examining the inter-rater reliability for each criterion. Questions 2bi and 2bii employed Cronbach's alpha (a widely accepted measure of internal consistency) to examine internal consistency within criteria domains and for all criteria together. These analyses were used to ascertain the reasonableness of establishing a global score for each exercise across criteria. Although there are no external "gold standards" against which we can compare our criteria, we explored the validity of our criteria (questions 3a, 3b, and 3c) by determining if the type of exercise, number of documents, or types of documents biased the global scores for each exercise. Finally, overall and domain-specific exercise performance scores were arrayed from best to worst performance to serve as a "report card" on exercise design (question 4).

Earlier deliverables for this task presented draft criteria, the results of applying those criteria to a small sample of exercises, a proposed set of revised criteria, and a subsequent application of the revised criteria to a small sample of exercises. The purpose of these evaluations was to begin to test the feasibility, reliability, and validity of the criteria and the evaluation process. This report documents additional changes to the criteria and an application of the criteria to all exercises available to RAND that met the criteria for necessary documentation. From this evaluation, this report provides information about the utility of the criteria for evaluating the design of these and future exercises.

This report is organized as follows. The remainder of this chapter provides additional background and information on the development of the criteria. Chapter 2 describes our approach to the evaluation; Chapter 3 describes the evaluation results; and Chapter 4 presents our conclusions. The appendices present, in greater detail, our evaluation criteria, descriptive data for

each exercise, the scores for all exercises evaluated, exercise performance data, and an example of a "report card."

BACKGROUND

The use of emergency preparedness exercises is now widespread throughout the public health and responder communities. Exercises vary considerably in complexity and may be designed to meet one or more of a variety of objectives, including training, gap analysis, planning, and team building. Ideally, as with any quality improvement cycle, preparedness exercises are used to identify areas where improvement is needed; to inform the process of developing plans and procedures to improve performance; and finally to enable repeat testing once improvements have been made. This process is critical to achieving the long-term goal of conducting preparedness exercises in public health: to mitigate the morbidity, mortality, psychosocial stress, and social consequences of a terrorist attack or public health emergency.

Conducting exercises can be expensive and time-intensive. Considering the limited personnel, time, and budget constraints that state and local public health departments face, it is especially important to have a process by which preparedness exercises can be evaluated for their quality and appropriateness. The criteria presented in this document are designed to be used by state and local public health personnel and evaluating agencies to assess the quality of a potential exercise, and/or to evaluate the implementation of an exercise after the fact. Further, the criteria can be used to develop a "report card" similar to the "Consumer Reports" evaluation of consumer goods. This report card could serve as a valuable tool for state and local public health personnel as they try to identify potential replicable exercises that will meet the needs of their particular organization.

DEVELOPMENT OF THE CRITERIA

The 14 final criteria and the related scoring ranges are shown in Table 1. The criteria are indicators of the appropriateness of exercise design in the domains of: (1) Goals and Objectives; (2) Scenario; (3) Participation; (4) Materials; and (5) Execution and Feedback. In developing the criteria, the project team relied largely on the relevant literature and RAND's considerable

experience in developing and conducting exercises in both the military and public health sectors, as well as our experience in measuring quality of care.

Table 1
Criteria and Scoring Ranges

Criteria	Score Range[a]
1. The goals of the exercise are clearly stated.	1 = No, 2 = Yes
2. The objectives of the exercise are clearly stated.	1 = No, 2 = Yes
3. Exercise objectives are appropriate given the goals of the exercise.	0-3
4. The exercise addresses each of its objectives.	0-3
5. Exercise objectives are measurable within the context of the exercise.	1-3
6. The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3
7. The exercise scenario is internally consistent.	0-2
8. The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2
9. The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3
10. The exercise is designed to engage all invited participants.	0-3
11. Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1-3
12. The exercise is designed to result in action items.	1 = No; 2 = Yes
13. The exercise is designed to solicit feedback from participants.	1 = No; 2 = Yes
14. The exercise, as designed, can be completed within the scheduled timeframe.	0-2

[a] "0" = "Not enough information to make an evaluation"; for criteria with range ending in "2", 1 = "No" and 2 = "Yes"; for criteria with range ending in "3", "3" = "High".

An important step in creating these criteria was to clarify the terminology. Two key terms are "exercise" and "scenario." An "exercise" is a tool that supports decisions by explorations of a "virtual reality," such as an activity undertaken by an agency or agencies to practice skills, test readiness to respond to emergencies, or evaluate response plans or training and development programs. We classified the exercises into five basic categories: Orientation, Drill, Tabletop Exercise, Functional Exercise, and Full Scale Exercise (Gebbie, 2004). Orientation exercises can be used to familiarize staff to an agency's emergency response plan or to inform staff of changes to that plan. Drills can be used to test response time, communication, and staffing capabilities, among others. Tabletop exercises focus on training and problem solving; staff come together to discuss responses to a particular scenario and how they might respond to changing conditions within the scenario. Functional exercises test and evaluate the capabilities of the emergency response system (such capabilities might include, for example, mass vaccination and epidemiological investigation). Full scale exercises are used to test and evaluate all or most of the emergency response systems over an extended period of time. These exercises often involve more than one agency.

A "scenario" is the "story" describing the emergency event/situation around which the exercise is designed. Scenarios depict potential or hypothetical real-world events to which exercise participants respond. As the scenario unfolds, participants are asked to assess and deploy resources, make decisions, determine actions, and ultimately to apply what has been learned through the experience.

Next, we determined the characteristics of the exercise design that were the most important to measure and developed criteria for each. As mentioned earlier, exercise design criteria selected for evaluation fall into five major domains: (1) Goals and Objectives; (2) Scenario; (3) Participants; (4) Materials; and (5) Execution and Feedback. Each of these domains is defined below.

- *Goals and Objectives* – An exercise must have clear goals and objectives. A goal represents what the developers of the exercise are ultimately trying to achieve through the exercise (the "big picture"). Exercise goals usually fall into at least one of the following categories: training, problem solving, or evaluation. Objectives refer to what the developers of the exercise expect the users to specifically accomplish as a result of

having completed the exercise. Objectives can be process- or outcome-oriented, depending on the goals of the exercise. The objectives of the exercise should be appropriate given the goals of the exercise, should be addressed by the exercise, and should be measurable within the context of the exercise. Criteria # 1, 2, 3, 4 and 5 address characteristics of the exercise goals and objectives (see Table 1).

- *Scenario* – The scenario chosen must be appropriate given the goals and objectives of the exercise. It should be plausible and the threat should be realistic. Finally, the science must be realistic. For example, unrealistic disease incidence and transmission rates will leave the participants unconvinced. Criteria # 6, 7, and 8 address characteristics of the exercise scenario (see Table 1).
- *Participation* – It is important that the appropriate experts participate in the exercise. A well-designed exercise will explicitly call for professionals and representatives of various agencies appropriate to address the emergency scenario postulated. The exercise should provide information about who should participate in the exercise and which agencies should be represented. For example, depending upon the nature of the postulated event, the exercise design should suggest (or require) actual or simulated participation of appropriate agencies such as the Centers for Disease Control and Prevention (CDC), the Federal Bureau of Investigation, the Federal Emergency Management Agency (FEMA), etc.; and professionals such as local Emergency Medical Services (EMS), law enforcement, medical professionals, and community members. In addition, to be an effective training instrument, the participants must be engaged and view the exercise as being worthwhile. Criteria # 9 and 10 address characteristics of the exercise participants (see Table 1).
- *Materials* – Ideally, an exercise should be replicable by any qualified individual(s). That is, the goals and objectives would be met if someone other than the exercise developers were to implement the exercise. To achieve this, the exercise guidance and materials should be adequate to allow others to easily replicate the exercise. Instructions needed to facilitate the exercise should be provided, including how to engage participants; clear ground rules (i.e., no retribution for comments made during exercise); and guidelines for how to make sure that the exercise continues to move forward, including a list of

questions to guide participants through the discussion if needed. Criterion #11 addresses characteristics of the materials available for conducting the exercise (see Table 1).

- *Execution and Feedback* – The central purpose of conducting exercises is for the public health agency to gain and maintain proficiency. Part of this is to record lessons learned and insights gained, and to identify issues that may be resolved in subsequent exercises or through independent study. An actionable item resulting from an exercise consists of detected deficiencies in procedures, organization, equipment, etc., that can be corrected by the public health agency. Ideally, it also reports on the resources needed to make these corrections. In addition to improving overall proficiency, properly implemented training exercises can also identify specific procedures and processes that might be improved. These lessons learned (i.e., action items) are generally recorded in an after-action report. At the conclusion of an exercise, it is also useful for the facilitators to conduct a "hot wash." This is generally a plenary session in which all participants offer their opinions and insights on the exercise and suggest issues for further study. These observations, coupled with the recorded results of the exercise, constitute the after-action report. Criteria #12, 13, and 14 address whether the exercise is designed to elicit action items for the organization and feedback about the exercise from participants (see Table 1).

Development of the final criteria was an iterative process. Previous Task 4 deliverables describe this process in detail (RAND, 2004a, 2004b, and 2004c). Several draft versions of the criteria were tested by using them to evaluate actual exercises and were revised according to input received from team members during the testing. Our initial test of the draft criteria involved using the criteria to evaluate a sample of three "test" exercises. This initial test examined the effectiveness of the criteria and resulted in further refinements to the criteria (RAND, 2004a). The revised draft criteria were tested again by using them to evaluate 17 exercises (RAND, 2004b). The purpose of this second test was to begin to test the reliability and validity of the criteria and the evaluation process. Based on the results of this larger test, additional changes were made to the criteria and reported to the Department of Health and Human Services (RAND, 2004c).

The criteria underwent a number of revisions. Original drafts of the criteria addressed characteristics of both exercise design and implementation and were divided into two sets of criteria: (1) Criteria for Evaluating Potential Exercises Prior to Implementation; and (2) Criteria for Evaluating Exercises After Implementation. The original after-implementation criteria had several elements that appeared to evaluate the conduct of the exercise, rather than the design of the exercise. In addition, a lack of necessary documentation across all exercises made it difficult to evaluate the after-implementation criteria.

In subsequent revisions to the criteria, we collapsed the two sets of criteria into a single set of criteria, which focused on characteristics of the exercise design. Additionally, many of the scoring ranges in the early development of the criteria were developed on a 5-point scale. As a result of the initial tests, we determined that the 5-point scale was too fine a metric for scoring the criteria because it compromised internal consistency. Therefore, the scoring was revised to a 3- or 4-point scale for some criteria and a Yes/No scale for others. A score of zero was also included for some criteria, where "0" signifies that there is not enough information available to evaluate the criterion. Even after rescaling the criteria, we still faced the challenge of developing an overall score for each exercise (question 4). (The challenge was to develop a valid single score using the 3- or 4-point scale for some criteria, and the Yes/No scale for others.) Our solution was to "norm" all scores to a scale from zero to one. More information about this procedure is provided in the analytic methods section of this report.

Prior to the final evaluation of the exercises, we reviewed the criteria once more. Three experienced reviewers were given four exercises each to evaluate. After completion of their evaluation, the entire evaluation team met and discussed the reviewers' scores and addressed specific issues and challenges with the criteria. As a result of this effort, we dropped several criteria and created an additional criterion, modified the language of some criteria and clarified the scoring definitions and the Guidance/Examples. Among the criteria dropped, several were excluded because there was not enough documentation available across exercises to evaluate. Others were deemed to be more appropriate as part of our descriptive data collection effort and are presented as part of Appendix B. One criterion was added to mirror the criterion regarding goals (#1). These revisions resulted in a reduction in the number of criteria from 20 to 14.

In addition to input from the project team, we also shared the draft criteria with staff at DHHS, and their feedback was incorporated into the final criteria. A draft of the criteria was also

given to the Expert Advisory Panel for comment. Table 1 summarizes the final criteria and the scoring range for each. Appendix A presents the criteria, the scoring range and instructions, and guidance/examples for each criterion to instruct the reviewers in evaluating the exercise.

2. FRAMEWORK FOR EVALUATION

COLLECTION OF EXERCISES

RAND started by casting the widest net in order to collect public health preparedness exercises for the Task 4 effort. Every state, three large cities (Los Angeles, Chicago, and New York City), the District of Columbia, and the U.S. territories were all recipients of grants from the Centers for Disease Control and Prevention (CDC) to develop necessary public health infrastructure to prepare for a potential bioterrorist attack. In total, there were 62 awardees. We were able to identify 55 bioterrorism coordinators in the United States and conducted phone interviews with 21 of them to learn about their involvement with the use of exercises for public health preparedness. Most of those who chose not to participate in the interview felt they did not have enough time. In each of these interviews, we asked respondents whether they, or others in their agency, had conducted public health preparedness exercises and if so, if they would be willing to share the exercise and associated materials with us for this Task. We also asked them if they knew of other experts in the field we might speak with to gather more exercises.

In addition to the bioterrorism coordinators, we contacted federal officials from several agencies, including the Department of Homeland Security, the Department of Defense, the Environmental Protection Agency, and the Department of Agriculture; private sector companies that design and conduct exercises; and national organizations of public health officials such as the National Association of County and City Health Officials (NACCHO). These efforts resulted in the development of a database that now includes 100 exercises, simulations, and drills. More than 30 states plus the District of Columbia and the territory of Puerto Rico are represented in the database.

SELECTION OF EXERCISES FOR REVIEW

The exercises included in the database varied substantially in the amount and type of documentation provided for each. In some cases, we knew that an exercise had occurred only because it was mentioned in our phone interviews; several of our respondents were not comfortable sharing any additional information or documentation to support the information we gathered through our interviews. However, for many exercises, we have at least one document

that describes some aspect of the exercise. The documents collected by RAND include the following:

- *Exercise Plan* – documents an overview of the exercise (portions of which can be distributed to participants and players), provides basic exercise development strategy, defines exercise parameters, explains the exercise, and assists participants in enhancing play (Gebbie, 2004).
- *After-Action Report* – summarizes actual exercise play, events, and participants' activities and evaluations after completion of the exercise.
- *Scenario* – describes the hazard and related events and conditions that set the stage for the exercise and provides background information for the emergency (Gebbie, 2004).
- *Facilitator Guide* – provides information for the facilitator to guide participants through the exercise and ensure that the exercise runs smoothly.
- *Participant Guide* – provides information to participants on the play of the exercise.
- *Presentation Material* – provides additional information to participants and/or facilitators (e.g., background information).
- *Participant Evaluation* – provides an opportunity for participants to relate their experiences from the exercise and offer recommendations.
- *Observer Evaluation* – reports and documents the exercise play from the point of view of third-party observers who are not actively involved in the exercise.
- *Observer Instructions* – provides instructions and guidelines for third-party observers (not actively involved in the exercise) to document or report on the exercise.
- *Other* – includes additional documents not included in the categories above, such as fact sheets, forms, and maps.

We have not yet been able to obtain a complete set of supporting material for all 100 exercises. To date, we have obtained one or more of the documents listed above for 74 exercises. For some, we have obtained only the after-action report; for others, we also have detailed exercise plans, participant guides, and presentation materials. Based on feedback from the exercise evaluators, we determined that only exercises with either an exercise plan or an after-

action report, plus one additional document, would be selected for evaluation. Based on this requirement, a total of 37 exercises, with supporting material, were evaluated for this task.

DATA COLLECTION METHODS

Ten reviewers were assigned exercises for evaluation against the final criteria. In an earlier activity, RAND reviewers were involved in observing or conducting some of the exercises we evaluated. However, since we were unable to observe all exercises, we limited our analysis to the written material only. Assignment of exercises to the reviewers was random; however, reviewers were not assigned exercises they had previously evaluated, authored, or observed. Each exercise was evaluated by a minimum of four and a maximum of six reviewers.

The reviewers scored their assigned exercises against the criteria and supplied those data to the statistician for analysis. In addition to the exercise evaluation, the reviewers provided basic descriptive information for assigned exercises. The descriptive data included the lead agency involved in executing the exercise, the type of exercise (e.g., tabletop exercise, drill, etc.), the goals/objectives of the exercise, the agent(s) used (e.g., anthrax, smallpox, etc.), source of disaster (natural vs. terrorist), the scalability of the exercise (e.g., whether the exercise can be easily tailored to fit other settings), and the resources (e.g., people, supplies) required to conduct the exercise. In addition, we include an accounting of the documents used to evaluate the exercise. Appendix B provides the complete descriptive data for each of the 37 evaluated exercises.

ANALYTIC METHODS

The purpose of our analyses is twofold: to provide information about both the utility of the criteria for evaluating these and future exercises and the appropriateness of exercise design. Below, we describe the analytic methods used to evaluate the criteria and exercise design.

As was shown in Table 1, the 14 criteria vary in the number of possible levels of response; e.g., the yes/no questions have two levels, whereas the ordinal questions have three to four levels of response. Assuming all questions are equal in their relative importance to one another and that the intervals between response choices are equal, and in order to provide a fairer comparison among criteria, all criteria scores were “normed” so that they ranged from 0 to 1,

with 1 indicating the topmost performance. For example, an ordinal criterion with four levels ("0", "1", "2", or "3") is now rescaled to "0.00", "0.33", "0.67", or "1.00".

We first present some general notation to describe the analyses:

Let X_{ijk} represent the score

for exercise i where i = 1, 2, 3,…,37

for criteria j where j = 1, 2, 3,…,14

for reviewer k where k = 1, 2,…,6 (there are 4, 5, or 6 reviewers per exercise).

Let X_{ijSD} represent the standard deviation of X_{ij1}, X_{ij2}, X_{ij3} etc.

Let X_{ijMEAN} represent the mean of X_{ij1}, X_{ij2}, X_{ij3} etc.

Using this notation, perfect agreement among reviewers would be:

$X_{ijSD} = 0$.

And the highest score (i.e., the top-most performance) would be:

$X_{ijMEAN} = 1$.

The closer the standard deviation is to zero, the more agreement there is among reviewers for a given criterion. Similarly, given the rescaled ("normed") criteria scores, the closer the mean is to one, the better that exercise performed on a given criterion. Criteria that have high inter-rater reliability and a high average score may not be very useful in discerning differences in exercise design across exercises.

We used the method of standard deviations to assess inter-rater reliability because we can apply this both (1) within each exercise (averaging each exercise's 14 scores to compute a measure of reviewer agreement) and (2) within each criterion (averaging each criterion's 37 scores to compute a measure of criterion agreement). We decided not to use either of two other common methods of computing reliability: the kappa statistic and intra-class correlations. The kappa statistic is unsuitable as this is traditionally used in a situation where there are only two reviewers reviewing all the subjects. In our case, there were four to six reviewers for each exercise, and these reviewers differed from one exercise to another. We have two dimensions of reliability: exercise (reviewer agreement) and criteria (consistency). Using the kappa statistic would not have allowed us to compute a measure of inter-rater reliability for the criteria.

Similarly, intra-class correlations (ICC) would only have allowed us to compute a reliability score for each of the exercises (reviewer agreement). It would not have allowed us to

compute a reliability score for the criteria due to the assumptions for the different ICCs. For the six different intra-class correlations as presented in Shrout and Fleiss (1979), one assumes that either (1) all subjects (in this study, the subjects are exercises) are rated by the same number of raters or (2) all subjects are rated by the same raters. Also, the use of ICCs imposes the distributional assumptions of normality on the data via the use of the method of analysis of variance (ANOVA). We did not wish to impose any distributional assumptions on our data. For these reasons, we chose our method of standard deviations to compute reliability scores (in each of the two dimensions) rather than using kappa statistics or intra-class correlations.

We also examined agreement across reviewers for individual criteria by examining the individual scores and assigning them to one of three categories based on score agreement: *perfect agreement, simple disagreement, and extreme disagreement*. *Perfect agreement* for a criterion is defined as *all* reviewers for an exercise rating the criterion with the same score. *Simple disagreement* is defined as minor differences among reviewer assessments. A maximum difference of one unit between reviewers constitutes a simple disagreement (e.g., one reviewer gives the exercise a "2" and the others give it a "3"). Simple disagreement is a function of the size of the maximum deviation among reviewer scores, but does not address the number of deviations.[1] *Extreme disagreement* is defined as major differences among reviewer assessments within an exercise. A difference in reviewer scores of two or more units constitutes an extreme disagreement.[2] There are four criteria with dichotomous scoring categories (yes/no). Any difference in scoring across reviewers for one of these four criteria on a given exercise is defined as an extreme disagreement. We calculated the distribution of agreement, using these three categories, for each criterion across exercises as well as for each exercise across criteria.

The primary questions we attempted to answer with the data related to assessing criteria feasibility, reliability, and validity, and to evaluating the exercise design for the set of exercises chosen for the study.

---

[1] For example, if there are five reviewers for an exercise, simple disagreement for Criterion #10 could indicate one minor difference (i.e., the five scores are 3,3,3,3,2) or more than one minor difference (3,3,3,2,2). In each case, however, the maximum difference between reviewers is one unit.

[2] In parallel with simple disagreement, extreme disagreement is based on the size of the difference rather than the number of differences among reviewers.

Feasibility of criteria. Simple counts of score responses for each criterion with an ordinal response scale were used to assess feasibility (the extent to which data are available and whether the criteria are sufficiently variable to distinguish good exercises from poor ones). The distribution of reviewer scores across the possible score responses was examined to assess the variability in response and any floor or ceiling effects. If we observed that scores are primarily distributed at one end of the response scale or the other, we would conclude that the response scale was not sufficiently sensitive to responses in that range.

Reliability of criteria. To determine the degree to which evaluators agreed on their rating of each criterion, we looked at the average of X_{ijSD} across i = 1, 2, 3,…, 37 for each criterion. Criteria with the lowest average standard deviations indicate consistent scoring among reviewers across exercises; criteria with the highest averages indicate lack of consistency. We also calculated the distribution of agreement across the three categories (perfect agreement, simple disagreement and extreme disagreement) for each criterion.

To assess internal consistency, we used Cronbach's alpha (a common measure of internal consistency) to examine the internal consistency of the criteria within each domain and all the criteria together. A Cronbach's alpha value greater than 0.80 indicates high internal consistency.

Validity of criteria. Validity was evaluated using linear regression analysis, and Chi-square (χ^2) statistics for categorical data. These analyses were conducted to assess bias in the evaluation of exercises, which might limit the validity of the criteria and subsequent evaluation.

Evaluation of exercise design. We examined the average (and standard deviation) of X_{ijMEAN} across j = 1, 2, 3,…, 14 for each exercise. Exercises with the highest averages indicate high performance across the 14 criteria. Exercises with the lowest averages indicate low performance. Standard deviations indicate the degree to which exercises perform well on some criteria but poorly on others.

To compare exercises, we established tertiles based on exercise performance (X_{ijMEAN}). Exercises are labeled "high," "medium," and "low," depending on their location in the performance distribution. As a result, all comparisons are made relative to other exercises in the set, rather than an external benchmark. We could have established categories of exercises based on performance using an external cutpoint (i.e., any performance scores greater than 0.75 on a scale from 0 to 1 would be considered good performance); however, any such cutpoint would be

arbitrary and is outside what we were charged to do in Task 4. An absolute standard requires having outside information about what is important with respect to public health emergency preparedness. However, this field is still in its infancy. As the field evolves, we will be able to return to this issue and start to develop external, evidence-based performance standards by which to compare exercises in the future.

3. RESULTS

EXERCISES EVALUATED

For this evaluation, 37 existing preparedness exercises were evaluated that focused on assessing the readiness of local and state health departments to respond to public health emergencies, including bioterrorism. The exercises represent a spectrum of exercise types, geographic regions, agencies involved, agents tested, and resources used.

The exercises were conducted in 22 states across the country and included one international exercise and a virtual (entirely online) exercise. All exercises were performed between December 1999 and December 2004. Most exercises were sponsored by the state department of health or by city or county public health agencies. The majority of exercises were designed to be implemented at the county – and in some cases multi-county – level. Others were implemented at the city (including university), state, or regional level.

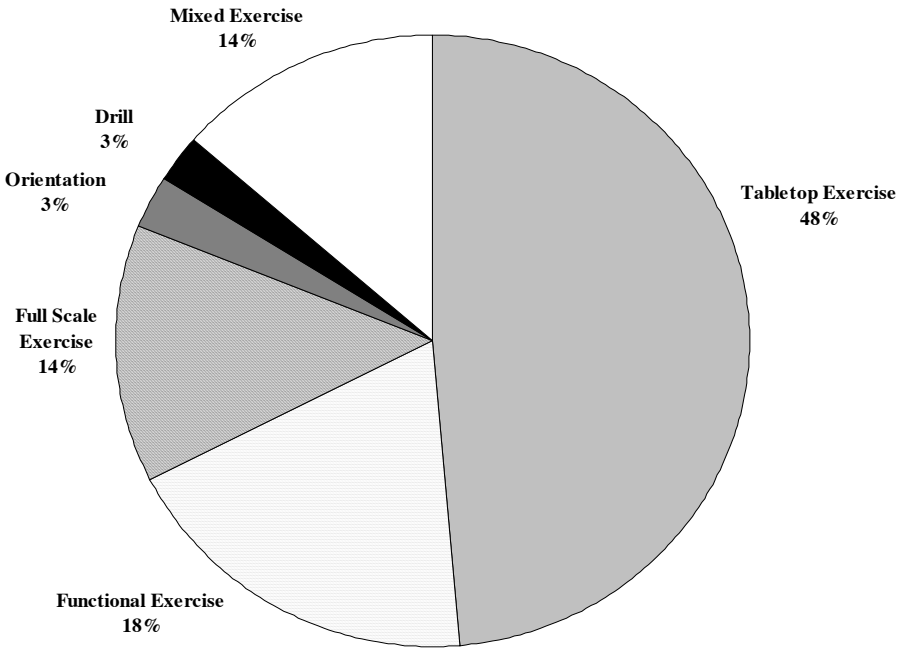

Figure 1. Distribution of Exercises by Type

Of the 37 exercises evaluated, the most frequent type was the tabletop (n=18 exercises), followed by functional (n=7), full scale (n=5), orientation (n=1), and drill (n=1) (Figure 1).

Additionally, five exercises incorporated a combination of exercise types (e.g., a tabletop and a functional exercise). Twenty-five exercises indicated that the source of the disaster was terrorist-related, while eight exercises involved a simulated natural disaster. The source of the disaster was not specified for four exercises. For purposes of evaluating the goals and objectives of the exercise, the source of the disaster is not particularly relevant. For example, to evaluate the adequacy of vaccination plans, it is not necessary to know whether the disaster is natural or terrorist in nature. Approximately one-third of the exercises presented scenarios using smallpox as the agent, by far the most common agent used in the exercises. Other exercises involved anthrax, explosive devices, food contamination and illness (including botulism, ergotism, salmonella poisoning, and shigellosis), influenza (avian and human), plague, SARS, tornadoes, and West Nile virus.

The time required to conduct the exercises ranged from a few hours to several days to a few weeks. The exercises involved a few to several thousand participants (one included thousands of volunteer "patients"). Most involved multiple state or local agencies and multiple jurisdictions.

As mentioned previously, exercises were selected for evaluation if they had either an exercise plan or an after-action report and at least one other source of documentation (i.e., scenario, facilitator guide, participant guide, presentation material, participant evaluation, observer evaluation, observer instructions, or other miscellaneous documents). Most of the exercises (n=29, 78 percent) included the scenario, an exercise plan (n=28, 76 percent), or an after-action report (n=22, 59 percent). A substantial number of exercises also had presentation materials (n=16, 43 percent), participant guides (n=14, 38 percent), facilitator guides (n=8, 22 percent), and other materials (n=15, 41 percent). The least common documents used in the evaluation were the participant evaluations (n=7, 19 percent), observer evaluations (n=5, 14 percent), and observer instructions (n=3, 8 percent).

Appendix B provides descriptive data for each exercise evaluated. However, at the request of those who supplied us with the exercise materials, all exercises have been de-identified in this report.[3]

FEASIBILITY OF CRITERIA

For the 37 exercises, a total of 161 reviews were conducted by ten reviewers. Each exercise was reviewed by four to six reviewers. Table 2 presents the results of addressing questions 1a and 1b regarding the feasibility of the criteria. (See the Introduction for a list of the specific questions addressed in this report.) In this table, we provide the count of individual rater responses for all criteria with at least three response levels. In addition, we present the percentage of all reviewer ratings that fell in each response category. We excluded criteria #1, 2, 12, and 13 from this analysis because those have only yes/no response options. To address the availability of the information, we examined the proportion of each criterion's score allocated to the "0" score ("Not enough information available to evaluate the criterion"). The proportion of all scores in the seven criteria with a "0" score option ranged from 7 percent to 32 percent. Lack of additional documentation (over and above the minimum documentation needed for evaluation) had only a modest effect on reviewers' ability to evaluate exercises.

Overall, we found that there is reasonable variation across response categories. However, there are two criteria that, in their current form, may need revision. For criterion #7 ("The exercise scenario is internally consistent") most reviewers indicated either "missing information" (score=0) or "internally consistent" (score=2). Similarly, for criterion #14 ("The exercise, as designed, can be completed within the scheduled timeframe") most reviewers indicated either "missing information" (score=0) or "the timeframe is suitable" (score=2). In both cases, the findings suggest that these criteria, in their current form, are not very useful as there is no "true" variation in responses. Additionally, score responses for criteria #4 ("The exercise addresses each of its objectives"), 6 ("The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise") and 10 ("The exercise is designed to engage all invited

---

[3] While exercises were de-identified for this report, we plan to obtain permission from the exercise developers to re-identify these exercises and make them available to the public via a website designed to communicate these findings.

Table 2
Feasibility of Criteria Across Response Categories

CRITERIA	SCORE RANGE	0	1	2	3
		N (%)	N (%)	N (%)	N (%)
1. The goals of the exercise are clearly stated.	1 = No, 2 = Yes	N/A	N/A	N/A	N/A
2. The objectives of the exercise are clearly stated.	1 = No, 2 = Yes	N/A	N/A	N/A	N/A
3. Exercise objectives are appropriate given the goals of the exercise.	0-3	27 (17%)	6 (4%)	32 (20%)	99 (61%)
4. The exercise addresses each of its objectives.	0-3	31 (19%)	2 (1%)	52 (32%)	78 (48%)
5. Exercise objectives are measurable within the context of the exercise.	1-3	N/A	27 (17%)	62 (39%)	72 (45%)
6. The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3	12 (7%)	1 (1%)	37 (23%)	111 (69%)
7. The exercise scenario is internally consistent.	0-2	21 (13%)	5 (3%)	135 (84%)	N/A
8. The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2	23 (14%)	12 (7%)	126 (78%)	N/A
9. The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3	N/A	30 (19%)	72 (45%)	59 (37%)
10. The exercise is designed to engage all invited participants.	0-3	51 (32%)	0 (0%)	27 (17%)	83 (52%)
11. Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1-3	N/A	33 (20%)	74 (46%)	44 (27%)
12. The exercise is designed to result in action items.	1 = No; 2 = Yes	N/A	N/A	N/A	N/A
13. The exercise is designed to solicit feedback from participants.	1 = No; 2 = Yes	N/A	N/A	N/A	N/A
14. The exercise, as designed, can be completed within the scheduled timeframe.	0-2	32 (20%)	7 (4%)	122 (76%)	N/A

participants") largely skip over response category 1, which for criterion #4 indicates that the exercise does not address any of the key objectives, for criterion #6 indicates that the scenario is

inappropriate, and for criterion #10 indicates that the exercise is not designed to engage any participants. Some recalibration of these response categories might be in order. It is possible that all the exercises we evaluated were designed well enough so that a low score was not reasonable; however, we can't be certain given that we do not have an external "gold standard" measure against which to compare our ratings. Although criteria #7 and 14 are not as useful and criteria #4, 6, and 10 may need to be recalibrated, the criteria produce sufficient variability in the aggregate to be able to distinguish among exercises.

RELIABILITY OF CRITERIA

The next set of questions we address are those related to the reliability of the criteria (questions 2a and 2b). Table 3 lists each of the 14 criteria along with the average of X_{ijSD} across i = 1, 2, 3,…, 37. (See column labeled Criteria Consistency.) As a group, Criteria #1, 2, 12, and 13 (with yes/no response categories) have the least variability among reviewers across exercises. This is not surprising, since we expect that as we limit the number of choices for a particular criterion, there will be less variability between the reviewers. Among the remaining ordinal criteria, the top three criteria in terms of inter-rater reliability are Criteria #6, 7, and 14. The three worst criteria in terms of inter-rater reliability are Criteria #5, 8, and 10.

Also listed in Table 3 is the average of X_{ijMEAN} across i = 1, 2, 3,…, 37. (See column labeled Scoring of Criterion.) The quantities closer to one indicate that the reviewers scored the exercises at the topmost possible score for this criterion. If a criterion has high inter-rater reliability (a desirable property) and also a very high average of X_{ijMEAN} across i, then this criterion may not necessarily be useful in discerning differences between exercises. Criteria #1 and 2 demonstrate both the highest inter-rater reliability and highest mean scores, suggesting that these two criteria may not be as useful as others in discerning differences between exercises.

Figure 2 presents the distribution of agreement categories across exercises for each criterion. As mentioned above, the criteria with yes/no responses (Criteria #1, 2, 12, and 13) have the highest levels of perfect agreement, ranging from a high of 81 percent of all exercises for Criterion #1 to 65 percent for Criterion #13. Not surprisingly, intra-exercise variation was greater for the remaining ten criteria (those with three- to four-point scales) than for the

dichotomous criteria. Nevertheless, there was still a substantial amount of agreement across reviewers.

Table 3
Scoring and Consistency for Reliability of Criteria Across Exercises

Criteria	Score Range	Scoring of Criterion[a]	Criteria Consistency[b]
1. The goals of the exercise are clearly stated.	1 = No, 2 = Yes	0.89	0.10
2. The objectives of the exercise are clearly stated.	1 = No, 2 = Yes	0.85	0.13
3. Exercise objectives are appropriate given the goals of the exercise.	0-3	0.75	0.22
4. The exercise addresses each of its objectives.	0-3	0.69	0.27
5. Exercise objectives are measurable within the context of the exercise.	1-3	0.63	0.29
6. The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3	0.82	0.20
7. The exercise scenario is internally consistent.	0-2	0.82	0.19
8. The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2	0.78	0.26
9. The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3	0.59	0.23
10. The exercise is designed to engage all invited participants.	0-3	0.63	0.28
11. Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1-3	0.54	0.20
12. The exercise is designed to result in action items.	1 = No; 2 = Yes	0.80	0.16
13. The exercise is designed to solicit feedback from participants.	1 = No; 2 = Yes	0.73	0.15
14. The exercise, as designed, can be completed within the scheduled timeframe.	0-2	0.78	0.17

[a] These values reflect the average of $X_{ij\text{MEAN}}$ across *i*. Higher values indicate better scoring.

[b] These values reflect the average of $X_{ij\text{SD}}$ across *i*. Lower values indicate higher consistency.

Perfect agreement among those with ordinal response scales ranged from a high of 65 percent (Criterion #14) to a low of 14 percent (Criterion #5). Simple disagreement ranged from 11 percent (Criteria #7 and 14) to 68 percent (Criterion #9). Extreme disagreement ranged from 11 percent (Criteria #9 and 11) to 49 percent (Criterion #10).

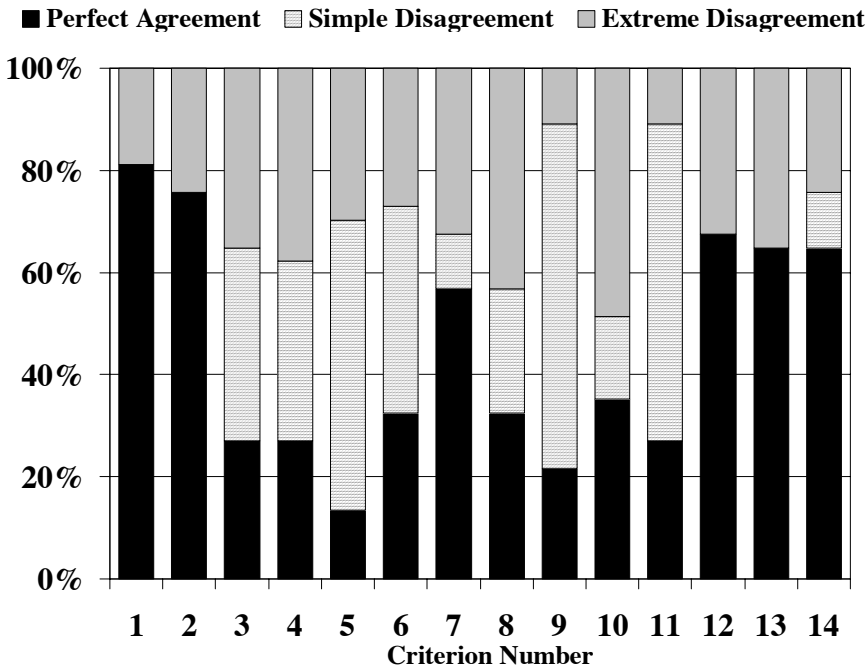

Figure 2. Reviewer Agreement on Criteria

Disagreement is associated with rating ranges and options. For Criteria #3, 4, 6, 7, 8, 10 and 14, reviewers had the option of reporting that there was not enough information to make a determination (i.e., "0"). These criteria had particularly high levels of extreme disagreement. The prevalence of zero scores in responses associated with extreme disagreement indicates that some reviewers found it difficult to identify the appropriate information in the documentation for these criteria. This result raises a concern about the quality of the documentation used to assess the exercises.

Table 4 summarizes our analyses of the internal consistency of the criteria (questions 2bi and 2bii). One of the assumptions we made was that all 14 criteria were equal in importance relative to one another. To justify creating a "global" score for each exercise, we first had to

determine that the criteria were internally consistent, or that they generally assess the same characteristics, skills, or qualities. We tested the internal consistency of the five domains of criteria as well as the overall internal consistency of the criteria. As Table 4 suggests, the internal consistency of the five criteria domains is very good, with all of the Cronbach's alpha values above 0.80. The overall consistency is also good, with a Cronbach's alpha of 0.87.

In addition to the results presented above, we performed a factor analysis on the 14 criteria for the 37 exercises to determine if the criteria "hang together" as a group (using X_{ijMEAN} – data not shown). The highest factor (with the largest eigenvalue) has all 14 criteria loading on to that factor with loadings for the ten non-yes/no criteria ranging from 0.39 to 0.85. The four Yes/No criteria have loadings ranging from 0.24 to 0.69. These findings confirm that the criteria are internally consistent overall and across the separate domains. Therefore, it is appropriate to construct an overall performance score for the exercises as a simple average across the 14 criteria.

Table 4
Internal Consistency of Criteria

Domain	**Cronbach's Alpha**[a]
Goals and Objectives (Criteria #1, 2, 3, 4, and 5)	0.85
Scenario (Criteria #6, 7, and 8)	0.83
Participants (Criteria #9 and 10)	0.84
Materials (Criterion #11)	N/A[b]
Execution and Feedback (Criteria #12, 13, and 14)	0.82
Overall	0.87

[a] A value greater than 0.80 is indicative of strong internal consistency.

[b] N/A: Not applicable. To estimate internal consistency, there must be at least two criteria in the domain.

The scoring of Criteria #3, 4, and 5 are dependent on how Criteria #1 and 2 are scored. To the extent that there is disagreement in those first two criteria, the disagreements will cascade into the three subsequent criteria. Those three criteria also have three to four response categories, thereby increasing the potential for disagreement. Thus, we see poorer consistency on these criteria relative to the first two.

Criterion #5 rated exercise objectives as to how measurable they were within the context of the exercise. Two issues affected scoring and agreement. First, this criterion is dependent on Criterion #2, which evaluates whether or not objectives are clearly stated. Consequently, Criterion #5 scores are linked and dependent on reviewers' scores for Criterion #2. Criterion #5 cannot score well if Criterion #2 scored poorly. Second, this criterion is dependent on the interpretation of "measurable" and how objectives were stated in the exercises. Some objectives were clearly quantifiable (e.g., vaccinate 3,000 people in 5 hours) while others were more subjective (e.g., convene stakeholders). The nature of these objectives could affect scoring of this criterion. To clarify any potential confusion, the guidance suggested that an objective could be considered measurable if, on a yes/no scale, it could be determined whether or not the objective was satisfied. These issues potentially affected scoring and consistency.

Criterion #8 was one of the least consistent criteria. The criterion asked the reviewer to determine if "the exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction." Exercises need to test the capabilities and resources of the jurisdiction rather than be calibrated to them. Feedback from reviewers suggests that exercise documentation does not consistently supply sufficient information about the capabilities of the participating jurisdictions, requiring greater reliance on reviewer judgment, which makes this a more difficult criterion to score reliably.

Criteria #9 and 10 were not as reliably scored as some of the other criteria. These criteria ask the reviewer to evaluate the appropriateness of the exercise design as it pertains to who should participate in the exercise and whether or not the exercise actively engages the participants. For Criterion #9, the source of most of the disagreement was what we deemed "simple" disagreement, meaning reviewers disagreed by a maximum of one unit on the scoring range. In most cases, the reviewers disagreed as to whether only "some" or "clear" guidance was provided in the documentation. Criterion #10 had more extreme disagreement, meaning that reviewers disagreed by two or more units on the scoring range. In this case, reviewers disagreed

as to whether there was sufficient information in the documentation to determine if some or all participants were actively engaged in the exercise. Reviewer feedback suggests that inclusion of the word "designed" in the criterion ("The exercise is *designed* to engage all invited participants") may have made the criterion difficult to interpret. Rephrasing the criterion or providing better guidance/examples might improve the consistency of this criterion.

Although Criterion #14 was relatively consistent, reviewer feedback suggests that we may need to revisit the construction of this criterion. The criterion asks the reviewer to assess the reasonableness of the scheduled timeframe of the exercise. Response categories were: "0"="No timeframe provided for completing the exercise *or* unable to ascertain if timeframe is appropriate from available exercise documentation"; "1"="Timeframe not appropriate"; "2"="Timeframe appropriate." Reviewers scored this criterion at the extremes of the scoring range. They either thought the timeframe was appropriate or felt there wasn't enough information available or no timeframe was provided. It is unclear, in its current state, how informative this criterion is.

VALIDITY OF CRITERIA

Given that there are no external standards, we must rely on face validity (i.e., the reasonableness of the criteria) to evaluate the criteria. In part, validation of the criteria was accomplished by expert review of the criteria as they were being developed to ensure they appeared reasonable and addressed the appropriate domains in evaluating the exercise design. We have also conducted a set of analyses to explore potential sources of bias coming from the type of exercise being evaluated and the number and types of documents available to conduct the evaluation. If any of these characteristics or resources associated with the exercise influenced the way they were scored, we would have to conclude that the criteria are not valid in their current form.

To assess if there was an association between the type of exercise conducted and the performance scores (question 3a), we compared the exercise agreement and performance tertiles by the different exercise categories (i.e., orientation, drill, tabletop exercise, functional exercise, full scale exercise, and mixed). Given our sample size, the cell sizes were too small to report the data, but there did appear to be a trend in the data toward better performance for exercises including more than one type of exercise (mixed).

The number and types of documents available for each exercise evaluation varied significantly (questions 3b and 3c). Inclusion in the evaluation required that each exercise have at least two documents: an after-action report or exercise plan and one other document. Some exercises had only this minimum documentation, while others provided as many as eight documents. It is possible that the documentation affected criteria performance. Nevertheless, bivariate and multivariate analysis of the relationship between the criteria performance within each exercise and the number and type of documents did not reveal any significant relationships (data not shown). However, such analyses cannot account for the quality of the documentation. The existence of a document, such as an after-action report, does not tell us anything about the quality of the document. It is likely that the quality of these documents is what determines the performance of criteria within an exercise.

EVALUATION OF EXERCISES

Table 5 presents all 37 exercises by level of inter-rater reliability (i.e., consistency), listed in ascending order by the average of X_{ijSD} across j = 1, 2, 3, …, 14. As explained in the table footnotes, we identified each exercise by a two-letter code, indicating the type of exercise (i.e., drill, orientation, tabletop exercise, etc.). The distribution is broken into tertiles: High Agreement, Medium Agreement, and Low Agreement.[4] Five exercises had scores less than 0.10, indicating very high reviewer agreement.

Regarding the performance of each exercise, an overall score was computed and is presented (along with its standard deviation) in Table 6. This overall performance score is the average of X_{ijMEAN} across j = 1, 2, 3, …, 14. This distribution is broken into tertiles, labeled "High," "Medium," and "Low" Performance.[5] For two exercises with similar performance scores, comparisons of their standard deviations help indicate whether a particular exercise has

[4] It is important to note that these exercises are labeled "High," "Medium," and "Low" *relative to each other*, as opposed to an external benchmark.

[5] Again, it is important to note that these exercises are labeled "High," "Medium," and "Low" *relative to each other*, as opposed to an external benchmark.

Table 5
Exercise Consistency

Exercise Code[a]	Number of Reviewers	Reviewer Agreement Tertile[b]	Exercise Consistency[c]
FU5	4	HIGH	0.05
TE17	4	HIGH	0.06
TE7	4	HIGH	0.07
TE11	5	HIGH	0.07
TE2	4	HIGH	0.09
TE15	4	HIGH	0.10
TE5	4	HIGH	0.12
FS4	4	HIGH	0.13
TE1	4	HIGH	0.14
FS1	4	HIGH	0.14
FU3	4	HIGH	0.15
MU3	4	HIGH	0.15
TE16	5	MEDIUM	0.17
TE4	4	MEDIUM	0.18
TE8	4	MEDIUM	0.18
TE13	4	MEDIUM	0.18
TE6	4	MEDIUM	0.20
FS5	4	MEDIUM	0.20
MU5	5	MEDIUM	0.20
DR1	5	MEDIUM	0.21
TE3	5	MEDIUM	0.22
TE10	5	MEDIUM	0.22
FS3	5	MEDIUM	0.22
MU2	4	MEDIUM	0.22
FU2	5	MEDIUM	0.24
MU1	4	LOW	0.25
FU6	4	LOW	0.26
TE14	5	LOW	0.26
FU4	4	LOW	0.27
FU1	4	LOW	0.27
TE18	5	LOW	0.27
OR1	5	LOW	0.28
TE9	4	LOW	0.29
FU7	6	LOW	0.29
MU4	4	LOW	0.33
FS2	4	LOW	0.34
TE12	4	LOW	0.34

[a] Exercise code abbreviations - OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises

[b] These rankings are relative to each other, rather than to some external benchmark.

[c] These values reflect the average of X_{ijSD} across *j*. Lower values indicate higher consistency.

Table 6
Exercise Performance

Exercise Code[a]	Number of Reviewers	Overall Performance Tertiles[b]	Scoring of Exercise[c]	SD[d]
TE11	5	HIGH	0.97	0.06
TE17	4	HIGH	0.96	0.07
TE7	4	HIGH	0.95	0.08
FU5	4	HIGH	0.95	0.12
TE15	4	HIGH	0.93	0.09
FS1	4	HIGH	0.92	0.10
TE5	4	HIGH	0.90	0.12
TE16	5	HIGH	0.88	0.16
MU2	4	HIGH	0.85	0.14
MU5	5	HIGH	0.85	0.17
MU3	4	HIGH	0.85	0.20
MU1	4	HIGH	0.83	0.15
TE8	4	HIGH	0.83	0.16
TE2	4	MEDIUM	0.83	0.32
FS5	4	MEDIUM	0.82	0.21
FU3	4	MEDIUM	0.82	0.20
TE13	4	MEDIUM	0.79	0.21
TE3	5	MEDIUM	0.78	0.21
FU6	4	MEDIUM	0.78	0.20
FS4	4	MEDIUM	0.77	0.24
TE9	4	MEDIUM	0.77	0.23
FU2	5	MEDIUM	0.76	0.27
DR1	5	MEDIUM	0.74	0.29
FS3	5	MEDIUM	0.72	0.27
TE6	4	MEDIUM	0.72	0.29
TE1	4	LOW	0.72	0.38
OR1	5	LOW	0.70	0.29
TE18	5	LOW	0.70	0.31
FU4	4	LOW	0.67	0.34
MU4	4	LOW	0.64	0.25
TE10	5	LOW	0.64	0.35
TE12	4	LOW	0.51	0.30
FU1	4	LOW	0.49	0.37
TE14	5	LOW	0.43	0.37
TE4	4	LOW	0.39	0.37
FS2	4	LOW	0.38	0.31
FU7	6	LOW	0.33	0.32

[a] Exercise code abbreviations - OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises

[b] These rankings are relative to each other, rather than to some external benchmark.

[c] These values reflect the average of X_{ijMEAN} across *j*. Higher values indicate better performance.

[d] SD = Standard Deviation of X_{ijMEAN}.

scores that fluctuate more across criteria. Although tertile development was somewhat arbitrary, tertiles easily show the distinction among exercises with the same or similar performance scores. For example, exercises "TE8" and "TE2" both have the same mean performance score, yet the former is labeled "High" and the latter is labeled "Medium". This determination was made because TE2's standard deviation was double that of TE8.

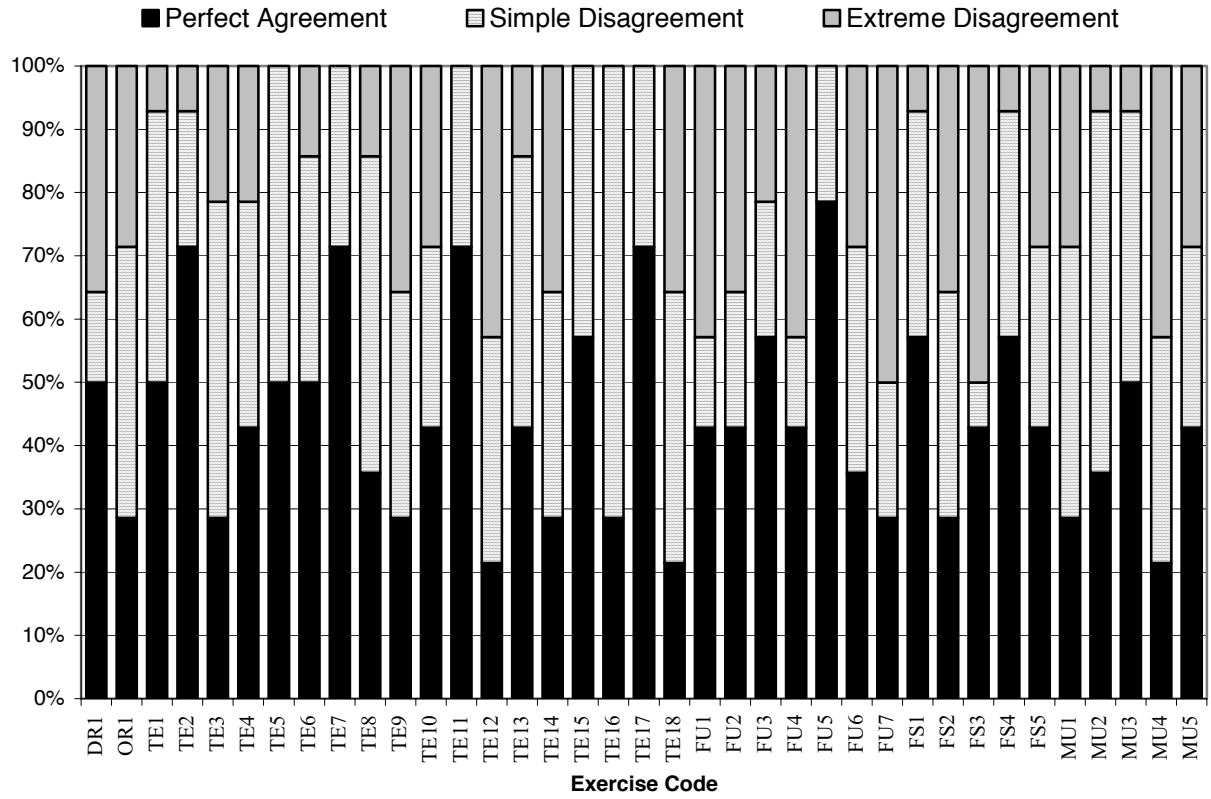

Figure 3. Reviewer Agreement on Exercises[6]

Figure 3 summarizes one aspect of exercise performance by showing, for each exercise, the number of criteria for which there is perfect agreement, simple disagreement or extreme disagreement. In general, exercise performance was good on this metric: 14 out of the 37 exercises had perfect agreement on 7 or more criteria. Six out of the 37 exercises had extreme disagreement on 6 or more criteria; however, seven exercises had no extreme disagreements. The

[6] Exercise code abbreviations: OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises.

highest performing exercise ("TE11"), based on X_{ijMEAN}, had over 70 percent perfect agreement on criteria across reviewers and no extreme disagreement. Exercise "FU6", selected because it was at the midpoint of the performance distribution based on the tertiles, had almost equal numbers of criteria for which there was perfect agreement, simple disagreement, and extreme disagreement. Seven out of fourteen criteria were classified as extreme disagreements across reviewers for the lowest performing exercise ("FU7"), while less than 30 percent of criteria had perfect agreement. Exercises that had a high number of extreme disagreements had very few simple disagreements, suggesting more difficulty in evaluating certain criteria given the documentation available.

The final question in this task was to evaluate the overall performance of the exercises and how they compare to each other (question 4). The performance scores in Table 6 present the performance scores for each exercise, ordered from highest to lowest performance. There is a threefold difference in the performance scores across exercises (in other words, the score for the highest-scoring exercise is three times that of the lowest-scoring exercise). The "high" and "middle" tertile exercises are highly clustered. Conversely, the "low" tertile is highly dispersed. Tables 7 and 8 present the prototypes of a "report card" for the exercises; Table 7 presents the numeric overall scores and the scores by each domain, while Table 8 organizes the data into "high," "medium," and "low" tertiles. The information is summarized at the domain level, which may be more useful than the detailed criteria level to a public health official or bioterrorism coordinator reviewing possible exercise options. Appendix D presents the exercise scores for reviewer agreement, overall performance, and performance on individual criteria for each exercise. Appendix E presents this information summarized by tertiles.

Table 7
Overall and Domain-Specific Performance Scores

		Domains				
Exercise Code[a]	**Overall Performance[b]**	**Goals and Objectives[b]**	**Scenario[b]**	**Participation[b]**	**Materials[b]**	**Execution and Feedback[b]**
TE11	0.97	0.95	1.00	1.00	1.00	0.93
TE17	0.96	0.93	0.97	1.00	1.00	0.96
TE7	0.95	0.94	0.96	1.00	0.75	1.00
FU5	0.95	0.88	1.00	1.00	0.88	1.00
TE15	0.93	0.93	0.93	0.94	0.75	1.00
FS1	0.92	1.00	0.96	0.90	0.75	0.83
TE5	0.90	0.88	0.93	0.88	0.63	1.00
TE16	0.88	0.89	0.94	0.67	0.90	0.93
MU5	0.85	0.92	0.87	0.78	0.40	0.93
MU2	0.85	0.75	1.00	0.79	0.75	0.92
MU3	0.85	0.68	0.92	0.85	1.00	1.00
MU1	0.83	0.88	0.81	0.69	0.50	1.00
TE8	0.83	0.86	0.90	0.65	0.63	0.92
TE2	0.83	0.98	0.96	0.81	1.00	0.42
FS5	0.82	0.84	0.88	0.54	0.63	1.00
FU3	0.82	0.83	0.76	0.81	0.50	1.00
TE13	0.79	0.83	0.90	0.48	0.50	0.92
TE3	0.78	0.86	0.79	0.53	0.80	0.80
FU6	0.78	0.86	0.97	0.71	0.50	0.58
FS4	0.77	0.82	0.94	0.38	0.50	0.88
TE9	0.77	0.74	0.92	0.46	0.88	0.83
FU2	0.76	0.85	0.56	0.75	0.20	1.00
DR1	0.74	0.66	1.00	0.42	0.20	1.00
FS3	0.72	0.80	0.72	0.38	0.20	1.00
TE6	0.72	0.57	1.00	0.75	0.63	0.71
TE1	0.72	0.94	1.00	0.83	0.50	0.08
OR1	0.70	0.75	0.88	0.45	0.60	0.67
TE18	0.70	0.78	0.84	0.85	1.00	0.23
FU4	0.67	0.93	0.33	0.50	0.00	0.92
MU4	0.64	0.58	0.65	0.48	0.38	0.92
TE10	0.64	0.88	0.90	0.32	0.60	0.20
TE12	0.51	0.53	0.61	0.06	0.25	0.75
FU1	0.49	0.54	0.64	0.00	0.25	0.67
TE14	0.43	0.57	0.70	0.30	0.20	0.07
TE4	0.39	0.12	0.86	0.69	0.63	0.08
FS2	0.38	0.20	0.58	0.19	0.25	0.67
FU7	0.33	0.31	0.60	0.04	0.08	0.36

[a] Exercise code abbreviations - OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises

[b] These values reflect the average of X_{ijMEAN} across *j*. Higher values indicate better performance.

Table 8
Overall and Domain-Specific Performance Tertiles

		Domains				
Exercise Code[a]	**Overall Performance[b]**	**Goals and Objectives[b]**	**Scenario[b]**	**Participation[b]**	**Materials[b]**	**Execution and Feedback[b]**
TE17	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
TE7	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
FU5	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
FS1	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM
TE11	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM
TE16	HIGH	HIGH	HIGH	MEDIUM	HIGH	MEDIUM
TE5	HIGH	HIGH	MEDIUM	HIGH	MEDIUM	HIGH
MU5	HIGH	HIGH	MEDIUM	MEDIUM	LOW	MEDIUM
TE15	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH
MU1	HIGH	HIGH	LOW	MEDIUM	MEDIUM	HIGH
MU2	HIGH	MEDIUM	HIGH	MEDIUM	HIGH	MEDIUM
TE8	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM
MU3	HIGH	LOW	MEDIUM	HIGH	HIGH	HIGH
TE2	MEDIUM	HIGH	HIGH	HIGH	HIGH	LOW
FS4	MEDIUM	MEDIUM	HIGH	LOW	MEDIUM	MEDIUM
FU6	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM	LOW
TE13	MEDIUM	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM
FS5	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH
FU3	MEDIUM	MEDIUM	LOW	HIGH	MEDIUM	HIGH
TE3	MEDIUM	MEDIUM	LOW	MEDIUM	HIGH	MEDIUM
FU2	MEDIUM	MEDIUM	LOW	MEDIUM	LOW	HIGH
FS3	MEDIUM	MEDIUM	LOW	LOW	LOW	HIGH
TE6	MEDIUM	LOW	HIGH	MEDIUM	MEDIUM	LOW
DR1	MEDIUM	LOW	HIGH	LOW	LOW	HIGH
TE9	MEDIUM	LOW	MEDIUM	LOW	HIGH	MEDIUM
TE1	LOW	HIGH	HIGH	HIGH	MEDIUM	LOW
TE10	LOW	HIGH	MEDIUM	LOW	MEDIUM	LOW
FU4	LOW	HIGH	LOW	MEDIUM	LOW	MEDIUM
OR1	LOW	MEDIUM	MEDIUM	LOW	MEDIUM	LOW
TE18	LOW	MEDIUM	MEDIUM	HIGH	HIGH	LOW
TE4	LOW	LOW	MEDIUM	MEDIUM	MEDIUM	LOW
TE12	LOW	LOW	LOW	LOW	LOW	MEDIUM
MU4	LOW	LOW	LOW	LOW	LOW	MEDIUM
FS2	LOW	LOW	LOW	LOW	LOW	LOW
FU1	LOW	LOW	LOW	LOW	LOW	LOW
FU7	LOW	LOW	LOW	LOW	LOW	LOW
TE14	LOW	LOW	LOW	LOW	LOW	LOW

[a] Exercise code abbreviations - OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises

[b] These rankings are relative to each other, rather than to some external benchmark.

LIMITATIONS OF ANALYSES

One caveat relates to the ordinal nature of the criteria. For example, for Criterion #7, the three options for the reviewer are "0"= "There isn't enough information to determine," "1"= "No," and "2"= "Yes." There are seven other criteria (Criteria #3, 4, 6, 8, 10, 11, and 14) for which the lowest scoring option is similar to "There isn't enough information to determine this." Implicit in the above analysis is the assumption that all criteria scoring options are ordinal. Hence for these eight criteria, we assume that "not having the information" is in fact worse than "having the information." However it is possible that the latent information would suggest that the exercise performs very well on these criteria. We do not know if the missing information is not available because written documentation does not exist or because the needed information was simply not sent to RAND. Many of the individuals contacted during the initial stages of this project were hesitant to share the documentation associated with the exercise, for proprietary reasons or because it contained sensitive information. However, we also know through feedback from the reviewers and through our analyses of the data that the quality of the documentation received by RAND for some of the exercises was also poor. Poor documentation would probably lower an exercise's performance score, but it is not clear at this point if failing to send documentation should be considered worse than poor performance.

4. CONCLUSIONS

In this section, we review the findings of our analyses as they relate to the four questions posed in Task 4. The headings of each of the sections that follow are the questions answered by this report.

1. What is the feasibility of the criteria?

The criteria are reasonably feasible. Lack of documentation above and beyond the minimum required for evaluation has only a modest effect on exercise ratings. This finding is consistent with analyses to answer question 3b, in which we found that the number of documents did not influence the scoring of the exercises. Two criteria (#7 and 14) did not demonstrate sufficient variation across response categories. However, overall, we find that there is sufficient variation across response categories, indicating that the criteria are sufficient to distinguish good exercises from poor ones.

2. What is the reliability of the criteria?

The criteria are reasonably reliable and demonstrate high internal consistency. No criterion had an inter-rater reliability score greater than 0.29 (a score of zero reflects perfect agreement). Those criteria with two response categories were the most reliable. Although we reduced the number of response categories for the criteria with multiple response options from earlier draft versions, there are still three or four response categories for many of them, increasing the potential for disagreement.

As discussed previously, several criteria were modified to include a "0" response category, reflecting the reviewers' findings that the exercise documentation was not sufficient to score the exercise well. The major source of disagreement between reviewers was due to differences in their interpretation of the availability of necessary documentation.

RAND had previously been involved in observing or conducting some of the exercises that we evaluated. We concluded from that activity that the best circumstances for evaluating exercises were both to review all available written material and to observe the exercise. One

observer noted that she was able to conclude that only some of the key stakeholders participated in the exercise because she had directly observed the exercise; shortcomings in the participant list were not detailed in the after-action report.

3. What is the validity of these criteria?

Based on expert review, the criteria have good face validity. Additionally, the type of exercise, the number of documents available for review and the types of documents available for review do not significantly bias reviewer evaluations, strengthening our conclusion that these criteria are reasonable and valid. We cannot test the validity of the criteria in the conventional way (against agreed-upon standards) because there are no gold standard measures against which to compare our criteria.

4. How well designed are the exercises?

There is substantial variation in the global exercise scores. In fact, there is a threefold difference between the highest and lowest performance scores. The variability across exercises within domains was even greater. The Goals and Objectives domain scores ranged from a perfect score of 1.0 to 0.12. The Scenario scores ranged from 1.0 to 0.33. Participation and Materials scores ranged from 1.0 to 0.0, and the Execution and Feedback scores ranged from 1.0 to 0.08. Generally, exercises that scored in the highest tertile overall also scored high across the individual domains, although only three exercises scored in the highest tertile across the board. Conversely, those that scored in the lowest tertile overall generally scored low in all the domains; four exercises scored in the lowest tertile across the board. With an important caveat (see "Document Quality" below), these results indicate that the criteria are useful in determining whether an exercise is well designed in relation to the other exercises evaluated.

DOCUMENT QUALITY

While we do find substantial variation across exercises, it is unclear how much of that variation is due to real differences in the quality of exercise design and how much is due to the quality of the exercise documentation. This concern is supported in part by the strong correlation between exercise consistency and performance; the more consistently scored exercises (those

with high levels of agreement) were also the better performing exercises. The more poorly performing exercises were also sources of relatively high levels of disagreement on scoring.

To evaluate the exercises well, the reviewers rely on good-quality documentation. Having more documentation is not necessarily better, as we demonstrated previously. Differences in performance were also not explained by having certain types of documentation (i.e., having an after-action report or an exercise plan). In the future, we may develop an instrument to evaluate the quality of available documentation and make that information available to users so that they can take document quality into account when selecting exercises for their own constituencies.

REFERENCES

Gebbie, K. (2004). Defining emergency exercises: A working guide to the terminology used in practicing emergency responses in communities and public health agencies. http://www.cumc.columbia.edu/dept/nursing/institute-centers/chphsr/def-exer.pdf (Accessed November 6, 2004).

Shrout, P.E. and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2): 420-428.

The RAND Corporation (2004a). Enhancing public health preparedness: Exercises, exemplary practices and lessons learned: Criteria for evaluating preparedness exercises (Task 4 deliverable 1). RAND: Santa Monica, CA.

The RAND Corporation (2004b). Enhancing public health preparedness: Exercises, exemplary practices and lessons learned: Applying criteria to a sample of exercises (Task 4 deliverable 2). RAND: Santa Monica, CA.

The RAND Corporation (2004c). Enhancing public health preparedness: Exercises, exemplary practices and lessons learned: Validating the evaluation process (Task 4 deliverable 3). RAND: Santa Monica, CA.

APPENDIX A:
CRITERIA FOR EVALUATING PUBLIC HEALTH EXERCISES

CRITERIA	SCORING	GUIDANCE/EXAMPLES
1. The goals of the exercise are clearly stated	1=No	A goal represents what the developers of the exercise are ultimately trying to achieve through the exercise (the "big picture"). Exercise goals usually fall into at least one of the following categories: training, problem solving or evaluation.
	2=Yes	Identifying the goals of the exercise is critical since the exercise's goals will impact the relative importance of a particular criterion in assessing the exercise and how a particular criterion should be applied to the exercise.
2. The objectives of the exercise are clearly stated	1=No	Objectives refer to what the developers of the exercise expect the users to specifically accomplish as a result of having completed the exercise. Objectives can be process or outcome-oriented, depending on goals of exercise.
	2=Yes	
3. Exercise objectives are appropriate given the goals of the exercise	0 = Not enough information available to evaluate the criterion	Objectives refer to what the developers of the exercise expect the users to specifically accomplish as a result of having completed the exercise. Objectives can be process or outcome-oriented, depending on goals of exercise.
	1 = None of the stated objectives are appropriate or none are stated	If you can identify the goal (even if it is not clearly stated), use that information to respond to this criterion.
	2 = Some but not all of the stated objectives are appropriate	Rater needs to determine whether the objectives correspond to the goals of the exercise.
	3 = All of the stated objectives are appropriate	For example, if the goal of the exercise is to evaluate, objectives might include: (1) To assess the level of X capability; (2) To evaluate ability to receive and distribute Y.
4. The exercise addresses each of its objectives	0 = Not enough information available to evaluate the criterion	Rater needs to determine whether the components of the exercise are designed to test the stated objectives (i.e., do the exercise components map back to the exercise objectives?).

CRITERIA	SCORING	GUIDANCE/EXAMPLES
	1 = Exercise does not address any of the key objectives	If Criterion #2=1 and Criterion #3=0 then score Criterion #4=0; If Criterion #2=1 and Criterion #3=1 then score Criterion #4=1.
	2 = Exercise addresses some but not all of the key objectives	For example: If an exercise objective is to assess gaps in communications between public health officials and hospital emergency room providers, the exercise should be designed to specifically test the communication processes and procedures by, for example, making the scale or details of the event such that this communication is necessary.
	3 = Exercise addresses <u>all</u> of the key objectives	Do not use concluding statements made in the after-action report as evidence that this criterion was met; the rater must come to his/her own conclusion based on the description of the exercise components themselves.
5. Exercise objectives are measurable within the context of the exercise	1 = None of the objectives are measurable	Rater should first determine what capacities/processes need to be tested in order to assess whether or not an objective is met, and then decide whether the exercise will generate, and incorporates ways to collect or document, the information regarding these capacities/processes.
	2 = Some but not all of the objectives are measurable	Note that some objectives can be appropriately measured in a yes/no scale if they focus on whether or not something happened while other objectives which focus on to what extent something happened are more appropriately assessed with quantitative measures.
	3 = All of the objectives are measurable	If Criteria #1,2, 3 and/or 4 are scored poorly (0's or 1's), be more cautious about scoring this one.
6. The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise	0 = Not enough information available to evaluate the criterion	"Scenario" refers to the story describing the emergency event/situation at different stages. Rater needs to determine if all of the different dimensions of the emergency event (e.g. how the emergency situation evolves throughout the exercise, is it consistent with known epidemiology) are at all possible given the scenario timeframe and what is likely in real life. The scenario only needs to be appropriate in terms of what the exercise is trying to accomplish (i.e. its goals and objectives).
	1 = None of the scenario is appropriate	**Consider:** The method of introduction of the agent, official and public response, etc. Also, consider whether the timeframe covered by the scenario is appropriate given the exercise objectives.
	2 = Some but not all of the scenario is appropriate	For more information about specific agents see:

CRITERIA	SCORING	GUIDANCE/EXAMPLES
	3 = All of the scenario is appropriate	Anthrax http://www.bt.cdc.gov/agent/anthrax/anthrax-hcp-factsheet.asp Smallpox http://www.bt.cdc.gov/agent/smallpox/overview/disease-facts.asp Plague http://www.cdc.gov/ncidod/dvbid/plague/index.htm http://www.bt.cdc.gov/agent/plague/index.asp Others http://www.bt.cdc.gov/agent/agentlist.asp
7. The exercise scenario is internally consistent	0 = There isn't enough information to determine 1 = No 2 = Yes	Rater needs to consider if the characteristics/parameters of the emergency event are possible given other facts of the scenario and that the event remains plausible. For example, the public health resources the exercise describes as being needed should be consistent with the size and type of the emergency event described in the scenario. Do not consider whether public health resources are realistic given a participating health jurisdiction.
8. The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0 = There isn't enough information to determine. 1 = No 2 = Yes	In determining if the exercise presents a realistic depiction of the capabilities and resources available to participants, the rater should consider the following: personnel, equipment, IT infrastructure, lab capacity, partner organizations, procedures. This is especially relevant for exercises in which the main goal is to test proficiency, since the capabilities and resources depicted in the exercise should reflect capabilities and resources that could be made available.

CRITERIA	SCORING	GUIDANCE/EXAMPLES
9. The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1 = No guidance is provided.	Rater must consider whether the exercise materials identify the players who should be invited to participate in the exercise. An exercise might provide a list of fields or subject area expertise to be represented at the exercise, and this would be appropriate.
	2 = Some guidance is provided.	Note that the list of key individuals/organizations invited to participate does not necessarily correspond to the list of key individuals/organizations mentioned in the exercise scenario. If a list of agencies/staff who did participate is provided in the documentation, that would suffice as "some guidance."
	3 = Clear guidance is provided.	It is entirely appropriate if some of the participants/organizations/ activities are simulated in the scenario.
10. The exercise is designed to engage all invited participants.	0 = Not enough information available to evaluate the criterion.	Rater needs to look at the different components of the exercise and determine if all the participants have been given an appropriate role during the exercise and enough of a role within the exercise to merit their presence during the exercise.
	1 = Exercise is designed to engage none of the participants.	If Criterion #9=0 then Criterion #10=0.
	2 = Exercise is design to engage some but not all of the participants.	Consider only the participants identified by the exercise developers; that is, do not consider key participants that were excluded by the exercise developers.

CRITERIA	SCORING	GUIDANCE/EXAMPLES
	3 = Exercise is designed to engage <u>all</u> of the participants.	Do not use concluding statements that say "everyone was engaged" made in the After Action Report as evidence that this criterion was met; the rater must come to his/her own conclusion based on the description of the exercise components themselves. **Example:** if mental health professionals are included but there is nothing in the exercise regarding the worried well, etc. then that would be a poor use of their time. **Example:** Having police involved in an epidemiologic investigation (also not an appropriate use of their time).
11. Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1 = <u>No</u> written materials are provided or instructions are unavailable for evaluation.	The rater needs to assess whether, given the materials developed for the exercise AND similar resources, others aside from the exercise developers could conduct the exercise.
	2 = Materials are provided but are not clear and/or detailed enough.	Rater should consider whether the instructions needed to facilitate the exercise are provided. This would include a complete list of resource materials given the exercise scenario and objectives. **Consider:** Explicit instructions for engaging participants; clear ground rules (i.e., no retribution for comments made during exercise); guidelines for how to make sure that the exercise continues to move forward, including a list of questions to pose to participants to guide them through the discussion if needed.
	3 = Clear materials are provided with satisfactory detail.	Rater needs to determine whether the information needed by participants is provided. These materials include instructions/guidelines to participants on how the exercise should proceed. This would include a complete list of resource materials given the exercise scenario and objectives. **Examples:** If an exercise involves response to a smallpox outbreak, resource materials might include case definitions and CDC vaccination protocols. If an exercise involves testing the logistics of distributing the Strategic National Stockpile, materials that simulate Stockpile contents should be available.
12. The exercise is designed to result in action items.	1 = No, or not described in available exercise documentation	"Action items" refer to next steps to be taken by the exercise users and/or participants to address preparedness issues based on the results of the exercise; action items are also referred to as recommendations for improvement, an improvement plan or a list of next steps.
	2 = Yes	Given that what we are looking for is a structural element of the exercise (planned a priori), if there is no mention in either the exercise plan or specified in the objectives that action items (or "next steps") are to be developed, this criterion should be scored as "1".

CRITERIA	SCORING	GUIDANCE/EXAMPLES
13. The exercise is designed to solicit feedback from participants.	1 = No, or not described in available exercise documentation	Rater needs to determine whether getting feedback from participants is actually built into the exercise process. This could include a participant evaluation form that asks participants to rate various exercise components or an opportunity at the end of the exercise for participants to share their views about the exercise (often called a "hot wash").
	2 = Yes	We are looking for a structural element of the exercise (planned a priori). The feedback should be provided after the exercise is completed, reflecting how well the exercise went or where it went wrong. The feedback could be specifically related to the structure of the exercise (e.g., not enough time provided to complete all activities specified in the exercise; key personnel were not included in the exercise, etc.) or what was learned from the exercise (e.g., we learned that we need a better process to vaccinate the population; we are not prepared to respond to an anthrax scare, etc.). Do not use the fact that feedback is documented in the after-action report as evidence that this criterion was met; the rater must come to his/her own conclusion based on the description of the exercise components themselves. ***However,*** if the after-action report does indicate that a feedback process was in place (i.e., feedback forms are described and summarized, etc.) and provides detail as to what feedback was solicited and when, that can sufficiently reflect that the exercise was designed to provide feedback.
14. The exercise, as designed, can be completed within the scheduled timeframe.	0 = No timeframe provided for completing the exercise OR unable to ascertain if timeframe is appropriate from available exercise documentation	The rater needs to decide whether, given the amount of time allocated to completing the exercise, there is enough time to complete all of the exercise components (i.e., if the exercise is meant to be completed in one day, can it in fact be completed in one day given its design/components).
	1 = Timeframe not appropriate	If there is a master event list, describing step by step the activities, that is a suitable substitution for a listed timeframe.
	2 = Timeframe appropriate	

APPENDIX B:
DESCRIPTIVE DATA BY EXERCISE

EXERCISE CODE	OR1	DR1
Lead Agency for Exercise	Not documented	State Department of Health
Geographic Region	County	County
Type of Exercise	Orientation	Drill
Goals and Objectives of Exercise	Exchange information regarding West Nile Virus surveillance and response activities and become familiar with emergency management procedure. Determine response based on hypothetical scenarios. Identify gaps.	The exercise was designed to test a mass vaccination plan of the region's population (300,000 residents) over ten days. The drill tested one of six regional clinics (serving 41,000 residents) in hospital response, transportation, volunteer call-down, emergency operations center/media and vaccination clinic. The drill tested shortfalls, strengths, key players, and system flexibility/effectiveness.
Agent Used in Exercise	West Nile Virus	Smallpox
Source of Disaster (Terrorist or Natural)	Natural	Not specified
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	7 hours	8 hours
	Approximately 40 participants	180 participants, including 80 volunteers
	Not documented	Multiple agency activity
	Single jurisdiction	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE1	TE2
Lead Agency for Exercise	State Department of Health	County Emergency Operations Center, American Red Cross
Geographic Region	State	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	This exercise was designed to detect an outbreak and follow steps set up for an outbreak investigation. Objectives included illustrating local, state, and federal roles and responsibilities in the event of an outbreak.	The goal was to implement local and regional decision-making. The objectives were to: review quarantine, public information, and medical monitoring procedures; examine interface between local, state, federal agencies and private sector; and discuss ways to integrate medical and criminal investigation.
Agent Used in Exercise	Botulism	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	Not documented	Not documented
	Not documented	17 agencies participated, approximately 25 participants
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE3	TE4
Lead Agency for Exercise	Center for Infectious Disease Preparedness, University School of Public Health	County Emergency Management
Geographic Region	City (University)	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	A tabletop exercise to gain skills in preparing for and responding to emerging infectious disease outbreaks (e.g., SARS) and bioterrorism threats. The exercise is designed to address policy, organizational, planning and human/material resource gaps.	This exercise was designed to initiate discussions on what responses should be given events presented in exercise, specifically an outbreak of avian influenza in poultry that were shown at the County Fair. The goal was to focus on response and recovery.
Agent Used in Exercise	SARS (Severe Acute Respiratory Syndrome)	Avian Influenza (H5N1 AI)
Source of Disaster (Terrorist or Natural)	Natural	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	3 hours	Not documented
	Minimum of 11 participants	102 participants
	Multiple agency activity	Multiple agency activity
	Single jurisdiction	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	☐ Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials **X** After-action report **X** Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE5	TE6
Lead Agency for Exercise	State Department of Public Health and Environment	County Public Health Department
Geographic Region	State	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goal was to familiarize agencies with response protocols. The objectives were to demonstrate the ability to coordinate relations with other stakeholders; notify and work with public health organizations; prioritize and use public health resources; and obtain assistance and work cooperatively with other stakeholders.	The goal was to discuss how the county's Public Health and Mosquito and Vector Control Agencies monitor health threats. The objective was to convene relevant stakeholders and identify action needed for prevention and response. A West Nile Virus scenario is used to frame a facilitated discussion of agencies' responses.
Agent Used in Exercise	Food contamination (Ergot - claviceps purpurea fungus)	West Nile Virus
Source of Disaster (Terrorist or Natural)	Terrorist	Natural
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	5 days	2.5 hours
	400 participants	5-10 participants
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Single jurisdiction
Materials Available	**X** Exercise plan ☐ Facilitator guide **X** Participant guide **X** Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Participant evaluation data	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE7	TE8
Lead Agency for Exercise	State Department for Health and Welfare, Division of Health	State Department of Public Health
Geographic Region	State	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goals of this exercise were to practice group problem solving and assess interagency coordination; discuss coordination/integration of forces, responsibilities, and disaster response plans; and acquaint key department personnel with one another and their mutual responsibilities.	The goals were to exercise Strategic National Stockpile and bioterrorism plans and assess familiarity with plans, policies, and procedures. Objectives included addressing surveillance, migration, roles and responsibilities, special psychosocial needs, security, alternate communication, and prophylaxis and vaccination.
Agent Used in Exercise	Salmonella	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Natural
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	3 hours	6 hours
	15 participants	96-166 participants
	Single agency activity	Multiple agency activity
	Single jurisdiction	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide **X** Participant guide ☐ Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Input cards	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE9	TE10
Lead Agency for Exercise	Not documented	Disaster Services Office, American Red Cross Chapter
Geographic Region	County	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goal was to examine public health preparedness in managing an acute public health crisis. Objectives over three phases in the exercise were: focus on assessment and implementation of contingency plans; expand the scenario to examine capability shortfalls and remedial measures; and gap assessment.	The goal was to provide an opportunity to understand problems encountered in an epidemiological incident. The focus was on overall response and the decision-making process rather than detailed response procedures. The exercise emphasizes communication, coordination, resource integration, problem identification, and resolution.
Agent Used in Exercise	Smallpox	Influenza pandemic
Source of Disaster (Terrorist or Natural)	Terrorist	Natural
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	4 hours	Approximately 1/2 day
	12 to 24 participants	Approximately 17 participants
	Multiple agency activity	Single agency activity
	Single jurisdiction	Single jurisdiction
Materials Available	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Interview guide-protocol	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE11	TE12
Lead Agency for Exercise	State Department of Health and University	State Department of Health
Geographic Region	County	State
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	This exercise aimed to identify and understand policy issues, measures to be performed, and gaps in the event of a bioterrorist attack. Objectives include recognizing roles of and building relationships among various officials and agencies.	This exercise was designed to demonstrate public health response to a major SARS outbreak. The objectives are to use the Emergency Operations Process (EOP) and SARS plan to observe methods for information analysis, decision making, and organization used by Incident Command Staff chiefs. Exercise included goals to improve the EOP and SARS plan.
Agent Used in Exercise	Shigellosis (a gastrointestinal illness)	SARS (Severe Acute Respiratory Syndrome)
Source of Disaster (Terrorist or Natural)	Terrorist	Natural
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	4 hours	3 hours
	14 participants	15-20 participants
	Multiple agency activity	Single agency activity
	Single jurisdiction	Single jurisdiction
Materials Available	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	**X** Exercise plan ☐ Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE13	TE14
Lead Agency for Exercise	State Department of Public Health	Professional Association
Geographic Region	County	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goal was to discuss with community authorities plans to respond to infectious disease and to deploy the Strategic National Stockpile. Objectives included examining plans and identifying roles of appropriate agencies to support, identify and explain the roles of the Hazmat One/Weapons of Mass Destruction Regional Response Team, and promote interagency collaboration and coordination.	This training exercise aimed to discuss each hospital/incident command personnel's role in a bioterrorism event; to describe the process for reporting unusual clusters to the health department and the health department's role in an epidemiological investigation; and to list special considerations for hospitals related to a bioterrorism attack.
Agent Used in Exercise	Widespread infectious disease	Plague (not specified) and Tularemia
Source of Disaster (Terrorist or Natural)	Natural	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	Approximately 6-8 hours	Not documented
	65-80 participants	Not documented
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE15	TE16
Lead Agency for Exercise	State Department of Health	State Department of Health
Geographic Region	State	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goal was to improve understanding of a response to a terrorist event, identify improvements, and change attitudes. The objectives were to coordinate and investigate surveillance data for identifying an agent, manage resources, provide information to the public through all media outlets, and to implement appropriate actions to protect the public.	This exercise addressed both operational and policy-level responses to a public health emergency in a hospital region. Its objectives included validation of bioterrorism preparedness plans, testing plans to identify gaps, and exploring implementation of the plan.
Agent Used in Exercise	Smallpox	Plague (Yersinia pestis)
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	1.5 days	8 hours
	25-50 participants	Not documented
	Multiple agency activity	Multiple agency activity
	Single jurisdiction	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide **X** Participant guide **X** Scenario **X** Presentation materials ☐ After-action report **X** Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report **X** Participant evaluations **X** Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	TE17	TE18
Lead Agency for Exercise	Department of Public Health	Professional Association
Geographic Region	State	County
Type of Exercise	Tabletop Exercise	Tabletop Exercise
Goals and Objectives of Exercise	The goal of the exercise was to assess the public health department's ability to conduct an epidemiological response to a suspected bioterrorism event with the ultimate goal of enhancing their response capabilities. Objectives included surveillance and detection, diagnosis and investigation, communication, and legal and law enforcement.	The goal was to raise awareness regarding the role of public health in bioterrorism prevention and response. The objectives were to identify the key roles public health plays in bioterrorism prevention and response; to describe where the public health workforce "fits in"; and to explain strategies for addressing bioterrorism in the schools of public health by turning knowledge into curricula.
Agent Used in Exercise	Smallpox	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	3 hours	2-4 hours
	14 participants (including 2 observers)	Not documented
	Multiple agency activity	Single agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations **X** Observer evaluations ☐ Observer instructions ☐ Other:	☐ Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario **X** Presentation materials ☐ After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Articles

EXERCISE CODE	FU1	FU2
Lead Agency for Exercise	County Department of Health Services	County Sheriff's Office
Geographic Region	County	County
Type of Exercise	Functional Exercise	Functional Exercise
Goals and Objectives of Exercise	The exercise was designed to test and evaluate procedures for implementing a Mass Smallpox Vaccination Clinic (Dispensing Vaccination Center-DVC). The objectives were to evaluate the efficiency and effectiveness of the DVC and estimate the resources needed for implementation.	This exercise explored an attack on cattle operations with a genetically engineered zoonotic disease agent. The goals and objectives required simultaneous legal and epidemiological investigations and imposition of quarantine measures. Exercise raised issues of joint criminal, human and animal disease investigation.
Agent Used in Exercise	Smallpox	Genetically engineered Vesicular Stomatitis (VSv)
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	8 hours	Not documented
	100-200 participants	Not documented
	Single agency activity	Multiple agency activity
	Single jurisdiction	Single jurisdiction
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report **X** Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Progress report, Disease fact sheet	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	FU3	FU4
Lead Agency for Exercise	State Department of Health Office of Emergency Operations	Department of Health
Geographic Region	County	International
Type of Exercise	Functional Exercise	Functional Exercise
Goals and Objectives of Exercise	This exercise deployed teams to (1) survey impacted communities with global positioning systems and (2) develop information on health needs. The purpose was to provide an opportunity for teams to train; understand how/why teams are deployed; understand direction and control; and train on equipment.	The exercise was aimed to evaluate the communications protocols between and among a global health organization's members in the face of an outbreak of an infectious disease.
Agent Used in Exercise	Disaster	Smallpox
Source of Disaster (Terrorist or Natural)	Not specified	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	None of the scenario can be easily scaled or there isn't enough information to determine if ANY of the scenario can be easily modified.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	3 days	56 hours over 2 1/3 continuous days
	Approximately 100 participants	10 countries/multinational organizations participating. Number of participants not documented.
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Assessment information	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: News article

EXERCISE CODE	FU5	FU6
Lead Agency for Exercise	State Emergency Management Agency and State Health and Human Services System	State Department of Emergency Management
Geographic Region	State	County
Type of Exercise	Functional Exercise	Functional Exercise
Goals and Objectives of Exercise	This exercise was a focused functional test in response to a terrorist biological attack requiring federal assistance from the National Pharmaceutical Stockpile. This exercise simulates delivery of the Strategic National Stockpile Push Pack as well as breakdown and repacking and distribution to regional hubs.	This exercise was a focused functional test of receipt and distribution of supplies from the National Pharmaceutical Stockpile. Role-playing activity exercises receipt, storage, and staging functions associated with Strategic National Stockpile requests.
Agent Used in Exercise	Plague (Pneumonic)	Plague (Pneumonic)
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	2 days	11 days
	Not documented	Not documented
	Multiple agency activity	Multiple agency activity
	Single jurisdiction	Single jurisdiction
Materials Available	**X** Exercise plan **X** Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials **X** After-action report **X** Participant evaluations **X** Observer evaluations **X** Observer instructions ☐ Other:	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	FU7	FS1
Lead Agency for Exercise	Not documented	Disaster Services Office, American Red Cross Chapter
Geographic Region	City	County
Type of Exercise	Functional Exercise	Full Scale Exercise
Goals and Objectives of Exercise	The goals were to assist in training and preparedness for a chemical weapons of mass destruction attack and to identify any unexpected weaknesses not previously noted during tabletop exercises prior to actual event.	The exercise tested the county's ability to both respond in the field and coordinate activities from the County Emergency Operations Center in response to a tornado.
Agent Used in Exercise	Chemical Nerve Agent	Tornado
Source of Disaster (Terrorist or Natural)	Terrorist	Natural
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	None of the scenario can be easily scaled or there isn't enough information to determine if ANY of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	Not documented	Approximately 4.5 hours
	Not documented	Over 67 participants
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Journal article	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report ☐ Participant evaluations **X** Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	FS2	FS3
Lead Agency for Exercise	State Division of Emergency Management	County Health Department
Geographic Region	County + International Border	County
Type of Exercise	Full Scale Exercise	Full Scale Exercise
Goals and Objectives of Exercise	The purpose was to provide the community with an evaluation of the strengths and weaknesses of its responders. An action plan was produced to outline how to improve the abilities of responding agencies through training and identification of needed equipment.	The goals were to exercise the county and health department's Emergency Operations Plan, test and set protocols for emergency public notification, test critical communications infrastructure, and provide training for personnel.
Agent Used in Exercise	Improvised Explosive Devices and Industrial Chemical	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Not specified
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	None of the scenario can be easily scaled or there isn't enough information to determine if ANY of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	4 hours	11 hours
	Not documented	74 participants
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Multiple jurisdictions
Materials Available	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report ☐ Participant evaluations **X** Observer evaluations **X** Observer instructions **X** Other: Progress statement	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario ☐ Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

EXERCISE CODE	FS4	FS5
Lead Agency for Exercise	Metropolitan Medical Response System	County Department of Health and Human Services, Public Health Services
Geographic Region	State	County
Type of Exercise	Full Scale Exercise	Full Scale Exercise
Goals and Objectives of Exercise	The goals were to evaluate ability to assess bioterrorism, make decisions on Strategic National Stockpile, activate plans, and coordinate resources. Objectives were to test, evaluate, and update plans.	The goals were to test and evaluate vaccination plans. Objectives included set up clinic; time patient flow; and assess communication, knowledge of roles, response to material needs and transportation issues.
Agent Used in Exercise	Anthrax	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Not specified
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	2 days	Approximately 9 hours
	Approximately 3,500-5,500 participants, including 3,000-5,000 volunteers	Number of total participants not documented; approximately 320 volunteer patients
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Single jurisdiction
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Presentation abstract	**X** Exercise plan ☐ Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report **X** Participant evaluations ☐ Observer evaluations **X** Observer instructions **X** Other: Flier, Forms, Agenda, Maps, Participant list, Patient flow, Pocket guide; Post-vaccine and quarantine instructions, Fact sheet, Q&A

EXERCISE CODE	MU1	MU2
Lead Agency for Exercise	State Department of Health and Human Services, Office of Public Health Preparedness and Response	State Department of Health
Geographic Region	State	City
Type of Exercise	Tabletop Exercises, Functional Exercise, and Full Scale Exercise	Tabletop Exercises and Full Scale Exercises with three different agents
Goals and Objectives of Exercise	The goals were to engage organizations and identify improvement with a focus on investigation and request of federal assistance (Phase I), receipt/distribution of Strategic National Stockpile (Phase II), and quarantine (Phase III).	The goals were to educate, evaluate response, and identify needs. Smallpox exercise tested contact tracing and setting up immunization and quarantine facilities. Anthrax and tularemia exercises focused on organizing a clinic.
Agent Used in Exercise	Plague (Yersinia pestis)	Smallpox, Anthrax, Tularemia
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	9 days (3 days each phase)	4 hours (tabletop) and 5-7 hours (full-scale)
	100-300 participants (3 phases)	15-25 participants
	Multiple agency activity	Multiple agency activity
	Multiple jurisdictions	Single jurisdiction
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Newsletter articles, RAND interview	☐ Exercise plan ☐ Facilitator guide ☐ Participant guide **X** Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Scenario cards,

EXERCISE CODE	MU3	MU4
Lead Agency for Exercise	State Emergency Medical Services Authority	State Association of Directors of Health, Inc.
Geographic Region	County	City
Type of Exercise	Functional Exercise, Full Scale Exercise, and Tabletop Exercise	Orientation, Tabletop Exercise, and Functional Exercises
Goals and Objectives of Exercise	The exercise focused on healthcare and government organizations' response to manage influx of large numbers of patients and shortage of resources, including staffing, inpatient beds, equipment, supplies and medications.	The goal of this project is to be able to mobilize the community to respond to 10% affected population.
Agent Used in Exercise	Plague (Yersinia pestis)	Smallpox
Source of Disaster (Terrorist or Natural)	Terrorist	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	All of the scenario can be easily scaled or does not require modification to be used in different jurisdictions.	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	4 hours (tabletop) and 5-7 hours (full-scale)	4 hours
	15-25 participants	16 participating agencies. Number of participants not documented.
	Multiple agency activity	Multiple agency activity
	Single jurisdiction	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide **X** Participant guide **X** Scenario ☐ Presentation materials ☐ After-action report **X** Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions **X** Other: Contractor brochure

EXERCISE CODE	MU5
Lead Agency for Exercise	Department of Human Resources, Division of Public Health
Geographic Region	State
Type of Exercise	Orientation, Tabletop Exercises, Functional Exercises and Full Scale Exercises
Goals and Objectives of Exercise	This project was designed to establish a learning environment that would evaluate state and local agencies' roles, protocols, and plans for responding to a bioterrorism incident.
Agent Used in Exercise	Anthrax
Source of Disaster (Terrorist or Natural)	Terrorist
Scalable (e.g., the exercise scenario can be easily scaled to fit other settings)	Some but not all of the scenario can be easily scaled or there isn't enough information to determine if ALL of the scenario can be easily modified.
Resources Needed to Implement the Exercise (time, personnel, single/multiple agency activity, single/multiple jurisdiction)	Approximately 18 days over 7 exercises of multiple types
	Approximately 750 participants over all exercise types
	Multiple agency activity
	Multiple jurisdictions
Materials Available	**X** Exercise plan ☐ Facilitator guide ☐ Participant guide ☐ Scenario **X** Presentation materials **X** After-action report ☐ Participant evaluations ☐ Observer evaluations ☐ Observer instructions ☐ Other:

APPENDIX C:
SUMMARY OF AVERAGE SCORES FOR EVALUATED EXERCISES
(individual scores are noted in parentheses)

			OR1	DR1	TE1	TE2	TE3	TE4	TE5	TE6	TE7	TE8	TE9	TE10	TE11
#	CRITERIA	Metric (Low to High)													
1	The goals of the exercise are clearly stated.	1 = No 2 = Yes	1.4 (1,2,1,1,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	1 (1,1,1,1)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2,2)
2	The objectives of the exercise are clearly stated.	1 = No 2 = Yes	2 (2,2,2,2,2)	1.6 (2,1,1,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	1 (1,1,1,1)	2 (2,2,2,2)	1.5 (2,1,2,1)	2 (2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,2,1)	2 (2,2,2,2,2)	2 (2,2,2,2,2)
3	Exercise objectives are appropriate given the goals of the exercise.	0-3	1 (0,3,0,0,2)	1.6 (3,1,1,0,3)	2.8 (2,3,3,3)	3 (3,3,3,3)	2.8 (3,3,2,3,3)	0.5 (0,0,0,2)	2.5 (2,3,2,3)	1.8 (3,1,2,1)	3 (3,3,3,3)	2.8 (3,3,2,3)	2 (2,3,3,0)	3 (3,3,3,3,3)	2.8 (3,3,3,2,3)
4	The exercise addresses each of its objectives.	0-3	3 (3,3,3,3,3)	2 (3,1,1,3,2)	2.8 (3,3,3,2)	3 (3,3,3,3)	2.6 (3,3,2,3,2)	0.5 (0,0,2,0)	2.5 (2,3,2,3)	1.5 (3,0,3,0)	2.5 (2,3,3,2)	2.3 (2,3,2,2)	2 (2,3,3,0)	2.4 (3,2,3,2,2)	2.8 (3,3,3,2,3)
5	Exercise objectives are measurable within the context of the exercise.	1-3	3 (3,3,3,3,3)	2 (3,1,1,3,2)	2.8 (3,3,2,3)	2.8 (3,3,3,2)	2 (3,2,2,2,1)	1.5 (2,1,1,2)	2.5 (2,3,2,3)	1.5 (2,1,2,1)	2.8 (3,3,3,2)	2.3 (2,3,2,2)	2.3 (2,3,3,1)	2.2 (3,2,2,2,2)	2.8 (3,3,3,2,3)
6	The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3	2.8 (3,3,3,2,3)	3 (3,3,3,3,3)	3 (3,3,3,3)	3 (3,3,3,3)	2.6 (3,3,2,3,2)	1.8 (2,0,3,2)	2.8 (2,3,3,3)	3 (3,3,3,3)	3 (3,3,3,3)	2.5 (2,3,3,2)	3 (3,3,3,3)	3 (3,3,3,3,3)	3 (3,3,3,3,3)
7	The exercise scenario is internally consistent.	0-2	1.8 (1,2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,2,1)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2,2)

			OR1	DR1	TE1	TE2	TE3	TE4	TE5	TE6	TE7	TE8	TE9	TE10	TE11
8	The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2	1.6 (2,2,2,0,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,2,1)	1 (2,0,1,2,0)	2 (2,2,2,2)	1.8 (2,2,2,1)	2 (2,2,2,2)	1.8 (2,2,2,1)	2 (2,2,2,2)	1.5 (2,2,2,0)	1.4 (2,2,0,1,2)	2 (2,2,2,2,2)
9	The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3	1.6 (3,1,1,1,2)	2 (3,2,2,1,2)	2.5 (3,2,3,2)	2.3 (2,2,3,2)	2.2 (2,2,3,2,2)	2.3 (2,2,2,3)	2.5 (2,2,3,3)	2 (2,2,2,2)	3 (3,3,3,3)	2.3 (2,2,3,2)	1.5 (1,2,2,1)	1.6 (1,2,1,2,2)	3 (3,3,3,3,3)
10	The exercise is designed to engage all invited participants.	0-3	1.8 (3,3,0,0,3)	1 (2,0,0,0,3)	2.8 (3,3,3,2)	3 (3,3,3,3)	1.4 (2,0,3,0,2)	2.3 (2,2,3,2)	3 (3,3,3,3)	3 (3,3,3,3)	3 (3,3,3,3)	2 (2,3,3,0)	2 (3,3,2,0)	1 (0,3,0,2,0)	3 (3,3,3,3,3)
11	Exercise guidance and materials are adequate to allow others to easily replicate the exercise	1-3	2.2 (2,3,2,2,2)	1.4 (2,1,1,2,1)	2 (3,2,2,1)	3 (3,3,3,3)	2.6 (3,3,3,2,2)	2.3 (2,2,3,2)	2.3 (2,2,3,2)	2.3 (2,2,3,2)	2.5 (2,2,3,3)	2.3 (2,3,2,2)	2.8 (2,3,3,3)	2.2 (2,3,2,3,1)	3 (3,3,3,3,3)
12	The exercise is designed to result in action items.	1 = No 2 = Yes	1.8 (2,2,1,2,2)	2 (2,2,2,2,2)	1 (1,1,1,1)	2 (2,2,2,2)	2 (2,2,2,2,2)	1.3 (1,1,2,1)	2 (2,2,2,2)	1.3 (1,1,2,1)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.4 (1,2,1,2,1)	2 (2,2,2,2,2)
13	The exercise is designed to solicit feedback from participants.	1 = No 2 = Yes	1.2 (1,1,1,2,1)	2 (2,2,2,2,2)	1.3 (1,1,1,2)	1 (1,1,1,1)	1.6 (2,2,1,1,2)	1 (1,1,1,1)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.5 (2,1,1,2)	1 (1,1,1,1,1)	1.8 (2,2,2,1,2)
14	The exercise, as designed, can be completed within the scheduled timeframe.	0-2	2 (2,2,2,2,2)	2 (2,2,2,2,2)	0 (0,0,0,0)	0.5 (0,2,0,0)	1.6 (2,2,1,1,2)	0 (0,0,0,0)	2 (2,2,2,2)	1.8 (2,1,2,2)	2 (2,2,2,2)	1.5 (2,2,2,0)	2 (2,2,2,2)	0.4 (0,0,2,0,0)	2 (2,2,2,2,2)

SUMMARY OF AVERAGE SCORES FOR EVALUATED EXERCISES (CONTINUED)
(Individual scores are noted in parentheses)

			TE12	TE13	TE14	TE15	TE16	TE17	TE18	FS1	FS2	FS3	FS4	FS5
#	CRITERIA	Metric (Low to High)												
1	The goals of the exercise are clearly stated.	1 = No 2 = Yes	2 (2,2,2,2)	2 (2,2,2,2)	1 (1,1,1,1,1)	2 (2,2,2,2)	1.8 (2,2,2,2,1)	2 (2,2,2,2)	1.8 (2,2,2,2,1)	2 (2,2,2,2)	1.8 (2,2,2,1)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)
2	The objectives of the exercise are clearly stated.	1 = No 2 = Yes	1.5 (2,1,1,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,2,2,1)	2 (2,2,2,2)	1.3 (1,1,1,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)
3	Exercise objectives are appropriate given the goals of the exercise.	0-3	1.8 (2,0,2,3)	2.5 (3,3,2,2)	0.4 (0,0,2,0,0)	2.8 (3,2,3,3)	2.8 (3,3,3,2,3)	3 (3,3,3,3)	2.6 (3,3,2,3,2)	3 (3,3,3,3)	0 (0,0,0,0)	2.4 (2,3,3,1,3)	2.3 (3,2,2,2)	2.8 (3,3,3,2)
4	The exercise addresses each of its objectives.	0-3	1 (2,0,2,0)	2 (2,3,0,3)	2.8 (2,3,3,3,3)	3 (3,3,3,3)	2.8 (3,3,3,2,3)	2.8 (3,3,3,2)	2.2 (2,3,2,2,2)	3 (3,3,3,3)	0 (0,0,0,0)	1.8 (2,2,3,2,0)	2.5 (3,2,2,3)	2 (3,2,3,0)
5	Exercise objectives are measurable within the context of the exercise.	1-3	1.5 (2,1,2,1)	2.3 (2,3,2,2)	2.6 (3,3,3,3,1)	2.5 (2,3,3,2)	2.6 (3,3,3,2,2)	2.5 (2,3,2,3)	2.4 (3,3,2,3,1)	3 (3,3,3,3)	1 (1,1,1,1)	2.2 (2,2,3,3,1)	2 (2,2,2,2)	2.3 (2,3,3,1)
6	The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3	1.8 (2,0,2,3)	2.5 (2,3,2,3)	2.4 (3,3,2,2,2)	2.8 (3,3,2,3)	2.8 (3,3,3,2,3)	2.8 (3,3,3,2)	2.8 (3,3,3,3,2)	3 (3,3,3,3)	1.5 (0,3,2,1)	2 (2,3,3,0,2)	2.5 (3,3,2,2)	2.3 (3,3,3,0)
7	The exercise scenario is internally consistent.	0-2	1.5 (2,0,2,2)	2 (2,2,2,2)	1.2 (2,2,2,0,0)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	1.6 (2,2,2,0,2)	2 (2,2,2,2)	1.5 (0,2,2,2)	1.4 (1,2,2,0,2)	2 (2,2,2,2)	1.8 (2,2,2,1)
8	The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2	1 (2,0,2,0)	1.8 (2,2,1,2)	1.4 (2,2,1,2,0)	1.8 (2,2,1,2)	1.8 (2,2,2,1,2)	2 (2,2,2,2)	1.6 (2,2,2,2,0)	1.8 (2,2,1,2)	1 (2,2,0,0)	1.6 (2,2,2,0,2)	2 (2,2,2,2)	2 (2,2,2,2)

			TE12	TE13	TE14	TE15	TE16	TE17	TE18	FS1	FS2	FS3	FS4	FS5
9	The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3	1.3 (1,1,2,1)	1.8 (2,2,1,2)	1.4 (1,1,2,2,1)	2.8 (2,3,3,3)	1.8 (2,2,2,2,1)	3 (3,3,3,3)	2.8 (3,3,3,3,2)	2.8 (2,3,3,3)	1.3 (1,2,1,1)	2 (2,2,2,2,2)	2 (2,2,2,2)	2.5 (2,3,3,2)
10	The exercise is designed to engage all invited participants.	0-3	0 (0,0,0,0)	1.8 (2,3,0,2)	1.2 (0,0,3,3,0)	3 (3,3,3,3)	2.8 (2,3,3,3,3)	3 (3,3,3,3)	2.4 (3,3,3,3,0)	2.8 (2,3,3,3)	0.8 (0,3,0,0)	0.8 (2,0,0,0,2)	0.8 (3,0,0,0)	1 (2,0,2,0)
11	Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1-3	1.5 (1,1,2,2)	2 (2,2,2,2)	1.4 (1,2,1,2,1)	2.5 (2,3,3,2)	2.8 (3,3,3,3,2)	3 (3,3,3,3)	3 (3,3,3,3,3)	2.5 (2,3,3,2)	1.5 (2,2,1,1)	1.4 (1,2,1,1,2)	2 (2,2,2,2)	2.3 (2,3,2,2)
12	The exercise is designed to result in action items.	1 = No 2 = Yes	1.8 (2,1,2,2)	1.8 (2,2,1,2)	1.2 (1,1,2,1,1)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	1 (1,1,1,1,1)	1.8 (1,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)
13	The exercise is designed to solicit feedback from participants.	1 = No 2 = Yes	2 (2,2,2,2)	2 (2,2,2,2)	1 (1,1,1,1,1)	2 (2,2,2,2)	1.8 (2,2,2,2,1)	2 (2,2,2,2)	1 (1,1,1,1,1)	2 (2,2,2,2)	1.5 (2,1,1,2)	2 (2,2,2,2,2)	1.8 (2,2,1,2)	2 (2,2,2,2)
14	The exercise, as designed, can be completed within the scheduled timeframe.	0-2	1 (2,0,2,0)	2 (2,2,2,2)	0 (0,0,0,0,0)	2 (2,2,2,2)	2 (2,2,2,2,2)	1.8 (2,2,1,2)	1.4 (0,2,2,2,1)	1.5 (0,2,2,2)	1 (0,2,2,0)	2 (2,2,2,2,2)	1.8 (2,2,1,2)	2 (2,2,2,2)

SUMMARY OF AVERAGE SCORES FOR EVALUATED EXERCISES (CONTINUED)
(Individual scores are noted in parentheses)

			FU1	FU2	FU3	FU4	FU5	FU6	FU7	MU1	MU2	MU3	MU4	MU5
#	CRITERIA	Metric (Low to High)												
1	The goals of the exercise are clearly stated.	1 = No 2 = Yes	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.5 (2,2,1,2)	2 (2,2,2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,2,1)	1.5 (1,2,2,1)	2 (2,2,2,2)	2 (2,2,2,2,2)
2	The objectives of the exercise are clearly stated.	1 = No 2 = Yes	1.5 (2,1,2,1)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1 (1,1,1,1,1,1)	2 (2,2,2,2)	1.8 (2,1,2,2)	2 (2,2,2,2)	1.5 (2,1,2,1)	2 (2,2,2,2,2)
3	Exercise objectives are appropriate given the goals of the exercise.	0-3	1.8 (2,2,3,0)	3 (3,3,3,3,3)	3 (3,3,3,3)	2.8 (3,2,3,3)	2.8 (2,3,3,3)	2.8 (3,2,3,3)	0.7 (0,0,0,3,1,0)	3 (3,3,3,3)	1.8 (3,2,2,0)	1.3 (0,2,3,0)	1.5 (3,0,3,0)	3 (3,3,3,3,3)
4	The exercise addresses each of its objectives.	0-3	0.8 (3,0,0,0)	1.6 (0,2,3,3,0)	1.5 (2,2,0,2)	2.3 (3,0,3,3)	2.3 (2,2,3,2)	2.5 (3,3,2,2)	0.5 (0,0,0,3,0,0)	2.3 (2,3,2,2)	2.8 (3,2,3,3)	2.5 (2,3,3,2)	1.3 (3,0,2,0)	2.4 (3,3,2,2,2)
5	Exercise objectives are measurable within the context of the exercise.	1-3	1.8 (3,2,1,1)	2.4 (2,3,2,3,2)	2.3 (2,3,2,2)	3 (3,3,3,3)	2.3 (2,3,3,1)	2.8 (3,3,2,3)	1.3 (1,1,1,3,1,1)	2.3 (2,3,2,2)	2.5 (3,2,2,3)	2.3 (2,2,3,2)	2 (3,1,2,2)	2.6 (2,3,2,3,3)
6	The scenario used in the exercise is appropriate given the goals and/or objectives of the exercise.	0-3	2 (3,3,0,2)	2 (2,3,3,2,0)	2 (3,2,3,0)	1.5 (0,3,0,3)	2.8 (3,3,3,2)	2.8 (3,3,2,3)	1.7 (2,3,0,2,3,0)	3 (3,3,3,3)	3 (3,3,3,3)	2.3 (3,2,2,2)	2.5 (3,2,3,2)	3 (3,3,3,3,3)
7	The exercise scenario is internally consistent.	0-2	1.5 (2,2,0,2)	0.8 (0,2,2,0,0)	1.5 (2,2,2,0)	0.5 (0,2,0,0)	2 (2,2,2,2)	1.5 (2,2,0,2)	1 (2,2,0,0,2,0)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	0.8 (2,1,0,0)	1.6 (2,2,0,2,2)
8	The exercise scenario is a realistic depiction of the capabilities and resources likely to be available to a participating health jurisdiction.	0-2	1 (2,2,0,0)	1.2 (0,2,2,2,0)	1.8 (2,1,2,2)	0.5 (0,2,0,0)	2 (2,2,2,2)	1.5 (2,2,0,2)	1.5 (1,2,0,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.5 (2,2,0,2)	1.6 (2,2,0,2,2)

			FU1	FU2	FU3	FU4	FU5	FU6	FU7	MU1	MU2	MU3	MU4	MU5
9	The exercise documentation gives clear guidance as to who should participate in the exercise, and which other organizations or functions need to be simulated.	1-3	1 (1,1,1,1)	2.8 (3,3,3,3,2)	2.3 (2,2,3,2)	2 (2,3,1,2)	2.5 (3,3,3,1)	2.3 (2,2,3,2)	1.2 (1,2,1,1,1,1)	3 (3,3,3,3)	2.5 (3,2,3,2)	2.8 (3,3,3,2)	2.3 (2,2,3,2)	2.8 (3,3,2,3,3)
10	The exercise is designed to engage all invited participants.	0-3	0 (0,0,0,0)	1.8 (0,3,3,3,0)	3 (3,3,3,3)	1.5 (3,0,0,3)	2 (2,3,3,0)	2.3 (3,3,0,3)	0 (0,0,0,0,0,0)	3 (3,3,3,3)	2.5 (3,2,3,2)	2.5 (3,2,3,2)	1 (2,0,0,2)	2 (2,3,2,3,0)
11	Exercise guidance and materials are adequate to allow others to easily replicate the exercise.	1-3	1.5 (1,2,2,1)	1.4 (1,1,2,2,1)	2 (2,2,2,2)	1 (1,1,1,1)	2 (2,2,2,2)	2 (2,3,1,2)	1.2 (1,1,1,1,2,1)	2.8 (3,3,3,2)	2.5 (3,2,2,3)	3 (3,3,3,3)	1.8 (2,1,2,2)	1.8 (2,1,1,3,2)
12	The exercise is designed to result in action items.	1 = No 2 = Yes	1 (1,1,1,1)	2 (2,2,2,2,2)	2 (2,2,2,2)	1.8 (2,1,2,2)	1.8 (2,1,2,2)	2 (2,2,2,2)	1.7 (2,2,2,2,1,1)	2 (2,2,2,2)	1.8 (2,1,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)
13	The exercise is designed to solicit feedback from participants.	1 = No 2 = Yes	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.5 (2,1,1,2)	2 (2,2,2,2)	1 (1,1,1,1,1,1)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1.8 (2,2,1,2)	1.8 (2,2,1,2,2)
14	The exercise, as designed, can be completed within the scheduled timeframe.	0-2	2 (2,2,2,2)	2 (2,2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	1 (2,0,2,0)	2 (2,2,2,2)	0.8 (1,0,2,2,0,0)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2)	2 (2,2,2,2,2)

Exercise code abbreviations: OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises.

APPENDIX D:

OVERALL PERFORMANCE OF EXERCISES AND INDIVIDUAL EXERCISE PERFORMANCE ACROSS CRITERIA

			Criteria														
			Goals and Objectives					Scenario			Participation		Materials	Execution and Feedback		
Exercise Code[a]	Exercise Performance	Exercise Consistency	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
TE11	0.97	0.34	1.0	0.5	0.6	0.3	0.3	0.6	0.8	0.5	0.1	0.0	0.3	0.8	1.0	0.5
TE17	0.96	0.34	0.8	0.3	0.0	0.0	0.0	0.5	0.8	0.5	0.1	0.3	0.3	1.0	0.5	0.5
TE7	0.95	0.33	1.0	0.5	0.5	0.4	0.5	0.8	0.4	0.8	0.6	0.3	0.4	1.0	0.8	1.0
FU5	0.95	0.29	1.0	0.0	0.2	0.2	0.2	0.6	0.5	0.8	0.1	0.0	0.1	0.7	0.0	0.4
TE15	0.93	0.29	1.0	0.8	0.7	0.7	0.6	1.0	1.0	0.8	0.3	0.7	0.9	1.0	0.5	1.0
FS1	0.92	0.28	0.4	1.0	0.3	1.0	1.0	0.9	0.9	0.8	0.3	0.6	0.6	0.8	0.2	1.0
TE5	0.90	0.27	0.8	0.8	0.9	0.7	0.7	0.9	0.8	0.8	0.9	0.8	1.0	0.0	0.0	0.7
TE16	0.88	0.27	1.0	0.5	0.6	0.3	0.4	0.7	0.8	0.5	0.0	0.0	0.3	0.0	1.0	1.0
MU5	0.85	0.27	1.0	1.0	0.9	0.8	1.0	0.5	0.3	0.3	0.5	0.5	0.0	0.8	1.0	1.0
MU2	0.85	0.26	0.0	1.0	0.1	0.9	0.8	0.8	0.6	0.7	0.2	0.4	0.2	0.2	0.0	0.0
MU3	0.85	0.26	1.0	1.0	0.9	0.8	0.6	0.9	1.0	1.0	0.8	0.7	0.5	0.8	0.5	0.5
MU1	0.83	0.25	0.8	1.0	0.9	0.8	0.9	0.9	0.8	0.8	0.6	0.8	0.5	1.0	1.0	1.0
TE8	0.83	0.24	1.0	1.0	1.0	0.5	0.7	0.7	0.4	0.6	0.9	0.6	0.2	1.0	1.0	1.0
TE2	0.83	0.22	0.8	0.8	0.6	0.9	0.8	1.0	1.0	1.0	0.8	0.8	0.8	0.8	1.0	1.0
FS5	0.82	0.22	1.0	1.0	0.9	0.9	0.5	0.9	1.0	0.5	0.6	0.5	0.8	1.0	0.6	0.8
FU3	0.82	0.22	1.0	1.0	0.8	0.6	0.6	0.7	0.7	0.8	0.5	0.3	0.2	1.0	1.0	1.0
TE13	0.79	0.22	1.0	1.0	1.0	0.8	0.6	1.0	1.0	0.7	0.3	0.3	0.6	0.4	0.0	0.2
TE3	0.78	0.21	1.0	0.6	0.5	0.7	0.5	1.0	1.0	1.0	0.5	0.3	0.2	1.0	1.0	1.0
FU6	0.78	0.20	1.0	1.0	1.0	0.8	0.8	1.0	0.8	0.8	0.9	0.7	0.4	1.0	0.8	1.0
FS4	0.77	0.20	1.0	1.0	0.9	0.7	0.6	0.8	0.9	1.0	0.8	0.3	0.6	1.0	1.0	1.0
TE9	0.77	0.20	1.0	0.5	0.6	0.5	0.3	1.0	1.0	1.0	0.5	1.0	0.6	0.3	1.0	0.9
FU2	0.76	0.18	0.0	0.0	0.2	0.2	0.3	0.6	1.0	1.0	0.6	0.8	0.6	0.3	0.0	0.0
DR1	0.74	0.18	1.0	1.0	0.8	0.7	0.6	0.8	1.0	0.9	0.4	0.6	0.5	0.8	1.0	1.0
FS3	0.72	0.18	1.0	1.0	0.9	0.8	0.6	0.8	0.9	1.0	0.6	0.7	0.6	1.0	1.0	0.8
TE6	0.72	0.15	0.5	1.0	0.4	0.8	0.6	0.8	1.0	1.0	0.9	0.8	1.0	1.0	1.0	1.0

			Criteria														
			Goals and Objectives					Scenario			Participation		Materials	Execution and Feedback		
Exercise Code[a]	Exercise Performance	Exercise Consistency	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
TE1	0.72	0.17	0.8	1.0	0.9	0.9	0.8	0.9	1.0	0.9	0.4	0.9	0.9	1.0	0.8	1.0
OR1	0.70	0.15	1.0	1.0	1.0	0.5	0.6	0.7	0.8	0.9	0.6	1.0	0.5	1.0	1.0	1.0
TE18	0.70	0.14	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.9	0.8	0.8	1.0	0.8
FU4	0.67	0.14	1.0	1.0	0.9	0.9	0.9	1.0	1.0	1.0	0.8	0.9	0.5	0.0	0.3	0.0
MU4	0.64	0.13	1.0	1.0	0.8	0.8	0.5	0.8	1.0	1.0	0.5	0.3	0.5	1.0	0.8	0.9
TE10	0.64	0.12	1.0	1.0	0.8	0.8	0.8	0.9	1.0	0.9	0.8	1.0	0.6	1.0	1.0	1.0
TE12	0.51	0.10	1.0	1.0	0.9	1.0	0.8	0.9	1.0	0.9	0.9	1.0	0.8	1.0	1.0	1.0
FU1	0.49	0.09	1.0	1.0	1.0	1.0	0.9	1.0	1.0	0.9	0.6	1.0	1.0	1.0	0.0	0.3
TE14	0.43	0.07	1.0	1.0	1.0	0.8	0.9	1.0	1.0	0.9	1.0	1.0	0.8	1.0	1.0	1.0
TE4	0.39	0.07	1.0	1.0	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.0
FS2	0.38	0.06	1.0	1.0	1.0	0.9	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
FU7	0.33	0.05	1.0	1.0	1.0	0.8	0.6	1.0	1.0	1.0	1.0	1.0	0.9	1.0	1.0	1.0

[a] Exercise code abbreviations: OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises.

APPENDIX E:
EXAMPLE REPORT CARD WITH PERFORMANCE AND AGREEMENT TERTILES

		Criteria													
		Goals and Objectives					Scenario			Participation		Materials	Execution and Feedback		
Exercise Code[a]	Exercise Performance	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
TE11	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH
TE17	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM
TE7	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
FU5	HIGH	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
TE15	HIGH	HIGH	HIGH	MEDIUM	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH
FS1	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	HIGH	MEDIUM
TE5	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM	HIGH	HIGH	HIGH
MU3	HIGH	LOW	HIGH	LOW	MEDIUM	MEDIUM	LOW	HIGH	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH	HIGH
MU5	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM	LOW	HIGH	MEDIUM	HIGH
TE16	HIGH	MEDIUM	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH	MEDIUM	LOW	HIGH	HIGH	HIGH	MEDIUM	HIGH
MU2	HIGH	MEDIUM	MEDIUM	LOW	HIGH	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM	HIGH	HIGH
TE8	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM	HIGH	HIGH	MEDIUM
MU1	HIGH	MEDIUM	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM	MEDIUM	HIGH	HIGH	HIGH
TE2	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	HIGH	HIGH	LOW	LOW
FU3	MEDIUM	HIGH	HIGH	HIGH	LOW	MEDIUM	LOW	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	HIGH	HIGH	HIGH
FS5	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	LOW	MEDIUM	HIGH	MEDIUM	LOW	MEDIUM	HIGH	HIGH	HIGH
TE13	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	LOW	MEDIUM	MEDIUM	MEDIUM	HIGH	HIGH
TE3	MEDIUM	HIGH	HIGH	HIGH	HIGH	LOW	MEDIUM	HIGH	LOW	MEDIUM	LOW	HIGH	HIGH	MEDIUM	MEDIUM
FU6	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	LOW
FS4	MEDIUM	HIGH	HIGH	MEDIUM	MEDIUM	LOW	MEDIUM	HIGH	HIGH	LOW	LOW	MEDIUM	HIGH	MEDIUM	MEDIUM
TE9	MEDIUM	HIGH	MEDIUM	LOW	MEDIUM	MEDIUM	HIGH	HIGH	LOW	LOW	MEDIUM	HIGH	HIGH	MEDIUM	HIGH
FU2	MEDIUM	HIGH	HIGH	HIGH	LOW	MEDIUM	LOW	LOW	LOW	HIGH	MEDIUM	LOW	HIGH	HIGH	HIGH
DR1	MEDIUM	HIGH	MEDIUM	LOW	MEDIUM	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH
FS3	MEDIUM	HIGH	HIGH	MEDIUM	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW	LOW	HIGH	HIGH	HIGH
TE6	MEDIUM	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	MEDIUM	LOW	HIGH	MEDIUM
TE1	LOW	HIGH	HIGH	MEDIUM	HIGH	HIGH	HIGH	HIGH	HIGH	MEDIUM	HIGH	MEDIUM	LOW	LOW	LOW

		Criteria													
		Goals and Objectives					Scenario			Participation		Materials	Execution and Feedback		
Exercise Code[a]	Exercise Performance	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
OR1	LOW	LOW	HIGH	LOW	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM	MEDIUM	LOW	HIGH
TE18	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	HIGH	LOW	LOW	MEDIUM
FU4	LOW	HIGH	HIGH	MEDIUM	MEDIUM	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	MEDIUM	HIGH	HIGH
MU4	LOW	HIGH	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW	MEDIUM	LOW	LOW	HIGH	MEDIUM	HIGH
TE10	LOW	HIGH	HIGH	HIGH	MEDIUM	LOW	HIGH	HIGH	LOW	LOW	LOW	MEDIUM	LOW	LOW	LOW
TE12	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW	LOW	LOW	MEDIUM	HIGH	LOW
FU1	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH
TE14	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
TE4	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	MEDIUM	MEDIUM	MEDIUM	LOW	LOW	LOW
FS2	LOW	MEDIUM	LOW	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW	LOW	LOW	HIGH	MEDIUM	LOW
FU7	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	MEDIUM	LOW	LOW

[a] Exercise code abbreviations: OR: Orientation; DR: Drill; TE: Tabletop Exercise; FU: Functional Exercise; FS: Full Scale Exercise; MU: Mixed Exercises.