

Preparedness Exercises 2.0: Alternative Approaches to Exercise Design That Could Make Them More Useful for Evaluating – and Strengthening – Preparedness

Brian A. Jackson and Shawn McKay

ABSTRACT

Preparedness exercises play central roles in both the building and assessment of organizational readiness for future incidents. Though processes for designing and evaluating exercises are well established, there are opportunities to improve the value of exercises for strengthening preparedness and as tools for gathering assessment data. This article describes the application of systems analytical approach adapted from engineering that examines response operations as systems with potential failure modes that could hurt performance at future incidents. This methodology, which has been applied previously to preparedness measurement, is explored here as a tool for exercise design to focus it more tightly on key potential problem areas and to make it easier to use exercise data to explore preparedness for incidents that could differ considerably from the specific exercised scenario.

INTRODUCTION

Preparedness exercises play a significant role in the national preparedness system. In the Federal Emergency Management Agency's Comprehensive Preparedness Guide-101, exercises are identified as a central element of an area's effort to refine and execute a preparedness plan as well as contributing to red-teaming efforts to test plans against different sets of assumptions.¹ For medical institutions, periodic exercising is part of the accreditation requirements imposed by the Joint Commission on Accreditation of Healthcare Organizations. For rare types of incidents or large-scale events, use of

simulated incidents is viewed as particularly important, since emergency response and management personnel are unlikely to encounter many of the challenges associated with such incidents during their day-to-day activities.

The general term *preparedness exercise* includes activities that fall over a wide range of scale, scope, and complexity. Described in detail in the first volume of guidebooks produced by the Homeland Security Exercise and Evaluation Program (HSEEP), exercise can range from the most basic of seminar-type interactions up to full-scale response simulations where units, equipment and personnel operate as they would at a real incident and volunteers serve in the role of victims requiring treatment (Figure 1 illustrates the range of exercise types, in order of increasing complexity).

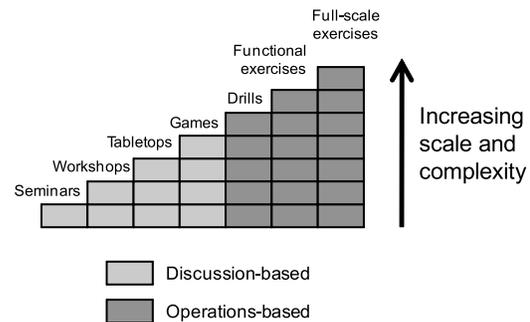


Figure 1: Varieties of Discussion-based and Operations-based Exercises²

As one component of a preparedness program, exercises of these varied types are seen as a versatile tool that can help contribute to achieving a variety of different goals. Though taxonomies of exercise objectives vary in the literature, most include the following:³

- **Planning** – Exercises provide a structure to advance planning for a particular

incident scenario, identifying problems and explore their solutions in focused way.

- **Interagency Coordination** – Exercises can act as a venue for members of different agencies to meet and interact, to build relationships that are important to effective coordination in a real event, to identify issues potentially falling in gaps of authority, jurisdiction, etc., to test mechanisms and technologies for interagency information sharing that might seldom be used in routine events, and to identify if there are agencies “missing” from plans that would be needed at a large scale disaster, accident, or terrorist attack.
- **Public Education** – Exercises can act as an “event” that, by being covered by the media and discussed publically, makes it possible to teach the public about the capabilities of response systems, creates the opportunity to educate them about preparedness actions they could take, and informs them about preparedness efforts of their local, state, or the federal government.
- **Training** – Exercises can make it possible to expose response staff to rare incidents and their unique demands – rather than their encountering them for the first time at a real emergency. Such simulations make it possible to teach responders or volunteers specific tasks, practice equipment use, and to learn or refresh other knowledge specific to an unusual incident.
- **Evaluation** – Exercises have been used to evaluate emergency preparedness activities in a variety of ways. Such evaluations range from very broad, qualitative assessments (e.g., ensuring all significant issues were considered in planning) to very detailed, quantitative studies (e.g., directly measuring the patient throughput of a medical facility). More elaborate and realistic evaluative exercises have the potential to assess not just that a preparedness plan can be executed, but how well it can be put into practice under the simulated conditions of the exercise scenario.

Given the effort and expense involved in designing them, a single exercise is sometimes expected to pursue some or all of

these goals simultaneously.⁴ This can be a challenge, since the different goals suggest different priorities and requirements for exercise design that can conflict with one another. For example, the requirement for certain types of realism could differ considerably between a training exercise and a focused evaluation drill.

The published literature contains examples of exercises focused on one or more of these goals and demonstrations, to varying extents, of how exercises can achieve them.⁵ Exercises focused on planning and training can demonstrate issues that were not addressed in existing plans and, for training efforts, pre-and post-tests of participants can show changes in their knowledge.⁶ In contrast, using exercises as tools for evaluating preparedness has been an area of active research focus,⁷ and though it is clear that the completion of an exercise does demonstrate something about preparedness, it is not always clear how much – or exactly what – has been demonstrated.⁸

Some of these challenges follow from existing shortfalls in the ability to evaluate the preparedness of an organization or jurisdiction. Though significant work has been done to develop methods and tools for assessing preparedness for specific incident types or to deliver response capabilities of interest, the ability to effectively assess whether a particular response system will perform well at a future incident is still lacking.⁹ Measurement shortfalls create challenges for designing exercises, since clear – and ideally measurable – preparedness outcomes are as important an input for designing training exercises as they are for framing exercises whose purpose is evaluation.

In a body of recent work,¹⁰ we have explored an alternative approach to preparedness assessment based on applying system reliability concepts adapted from engineering to response systems. The approach is based on taking what those systems plan to be able to do and then examining what could go wrong that might prevent them from successfully delivering the planned response capacity at a future incident. We define response reliability as the probability that a response system will be able to deliver a specific level of capability at

a future incident (e.g., the ability to deliver mass care to a population of 2,000 people for a period of time). A highly reliable system would have a high percentage chance of successfully delivering a capability level, while a low reliability one might be very unlikely to do so. Though the reliability of response systems is a characteristic that appears to be critical for understanding preparedness from the national to the local level – it answers the fundamental question the public has about response systems, “how likely is it that the system be able to respond successfully to a future incident” – it is not a factor that is currently assessed.¹¹

Examining preparedness exercises, we see the concepts that come out of thinking through factors that affect the reliability of a response system as having two potential applications:

- First, the logic underlying the analysis of a response system’s reliability can be applied in exercise design, providing a systematic design approach that can increase the insight that can be gained from exercises across the different types listed in Figure 1.
- Second, if appropriately designed, exercises could be very effective tools for *measuring* the reliability of a response system – thereby improving the ability to assess preparedness.¹²

We see both these paths as potential opportunities to improve the payoff from the substantial investment made by organizations and agencies inside and outside government in preparedness exercises every year.¹³ Extending our analogy to engineering techniques and methods, these two applications are similar to the use of these approaches at both the design stage of the construction of a technical system – to identify and correct potential problems before they occur – and in testing and evaluation of complete technical systems.

In this paper we explore these two topics, rooted in our concept of *reliable* response systems as one of the key goals for preparedness efforts and a preferred target for preparedness evaluation. After introducing response reliability analysis in

more detail, we examine the application of both its methods and results in exercise design, and examine how the logic of reliability assessment could contribute to better exercise design across the range of exercise goals. We will then turn to how exercises can be used as a primary evaluation tool to gather data to assess a response system’s reliability.

OVERVIEW OF RESPONSE SYSTEM RELIABILITY ANALYSIS

The fundamental approach of reliability analysis is to evaluate a response system and plans by systematically and, if possible, quantitatively analyzing the events and faults that could prevent it from performing as planned. To evaluate response system reliability, we adapted analytical techniques developed in engineering, specifically fault tree analysis and failure mode, effects, and criticality analysis (FMECA).¹⁴ The basic steps of a FMECA analysis are:

- **Defining and mapping the response system**, to identify the different parts of the response operation and articulate what it means for them to function well. For example, incident command at a response could be mapped as made up of several parts, including building situational awareness about the incident, making decisions about resource allocation among response functions, and dispatching response resources. The system diagrams used for reliability analysis are similar to process maps applied in other preparedness evaluation efforts.¹⁵ To illustrate the types of diagrams involved and provide an anchor for later discussion of their application in exercise design, Figure 2 includes an example of a system model for the incident command elements of a generic response based on recent analyses we have performed. Complete examples are available in works previously cited.
- **Identifying failure modes that could hurt system performance**. Failure modes are defined as “the observable manners in which a component fails,”¹⁶ which, for a response system, would be the ways performance in different parts of the

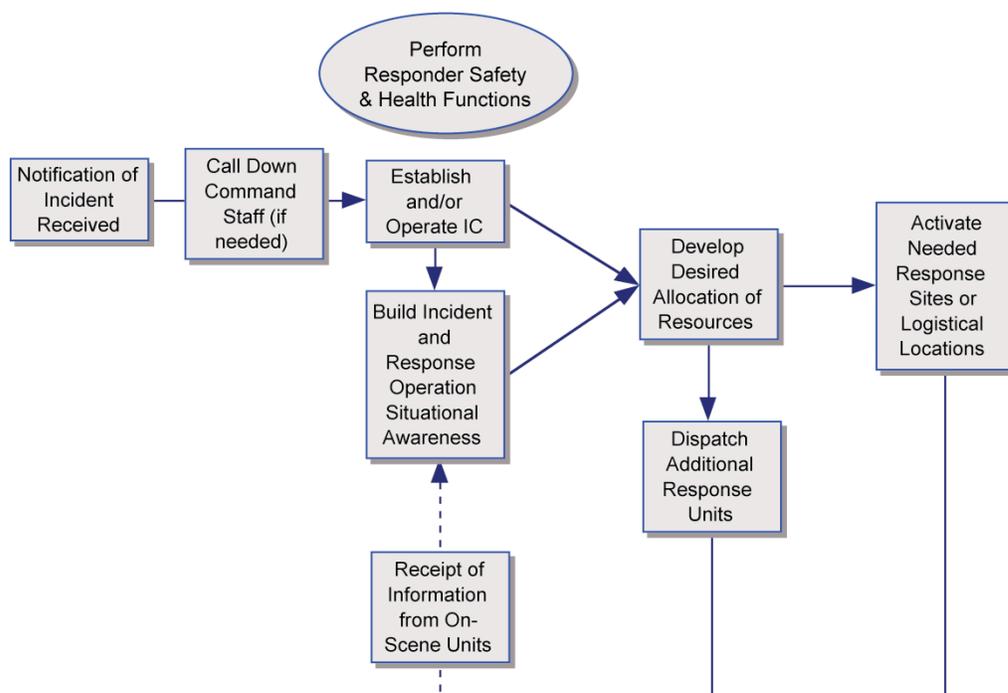


Figure 2: Example Portion of Response System Model, Incident Command Components

response system would break that would hurt overall system performance. Identifying failure modes involves systematically inventorying what might go wrong in each part of the system. In our past work, we used “classes” of failures that might produce the same end result to organize the analysis. For each part of the system we looked for potential failures in four main classes: (a) planning and organization, (b) equipment and technology, (c) personnel shortfalls and human error, and (d) external environmental causes. Some events or breakdowns could hurt response performance directly, while others might only do so in combination with other failures. Failure modes can be presented in fault trees that show the range of breakdowns that could affect the functioning of an individual part of a response system. Figure 3 provides an example failure tree for communications between on-scene response units and incident command based on our recent work, which illustrates the different classes of failure modes. The example is included to

provide a foundation for later discussion of how such trees could be used in designing or evaluating exercises.

- **Estimate the probability of different failure modes.** Because there are many events that could hurt the functioning of a response system, one way to differentiate among potential failure modes is their probability of occurrence. All other things equal, failure modes that are more common will be of greater concern than less common breakdowns. There are a variety of ways that failure mode likelihoods can be estimated, ranging from use of real world data to the systematic elicitation of expert opinion.
- **Assess the effects and severity of different failure modes.** The other characteristic that differentiates among failure modes is their severity. For FMECA, severity is assessed by asking what the effect of the failure is on overall system functioning – which for a response system would be the ability of to deliver the response capability it is designed to provide at an incident. Failure modes can have a

variety of effects, ranging from complete failure of the system (i.e., termination of a response operation) to minor reductions in capability or effectiveness. For example, destruction of an emergency operation center in the course of an incident might have very significant or even response terminating effect, while loss of a few response vehicles to breakdowns would cut into response capacity to a smaller extent.

Because the focus of this discussion is not the details of the analysis process but on the relationship with exercise design and evaluation, we will not describe the details of each of the steps.¹⁷ Instead, we will walk through the broader outcomes of reliability analysis that relate most closely to the later sections discussing exercise design.

As illustrated in Figure 3, a realistic response system will have a variety of failure modes that might occur – e.g., communications problems, staff shortages, traffic delays that limit dispatching of

resources – whose effects on response operations and on the effectiveness of the activities those operations are tasked with carrying out will vary in magnitude. For example, for delivering food aid to victims evacuated from a disaster, an area’s plans and preparedness efforts will have some theoretical maximum capacity to provide care. For ease of discussion, assume a maximum capacity to feed 1000 people for one week. But that same system has a set of failure modes that could reduce its performance when its response is activated. Coordination problems between aid organizations and response agencies might reduce efficiency, cutting the maximum by 10 percent. Damage to key infrastructures during the incident (e.g., an aid storage or staging area) might cause more significant reductions. As a result, at an incident where coordination broke down but everything else went as planned, the system would be able to feed 900 people for the required week. At another incident, where multiple failures

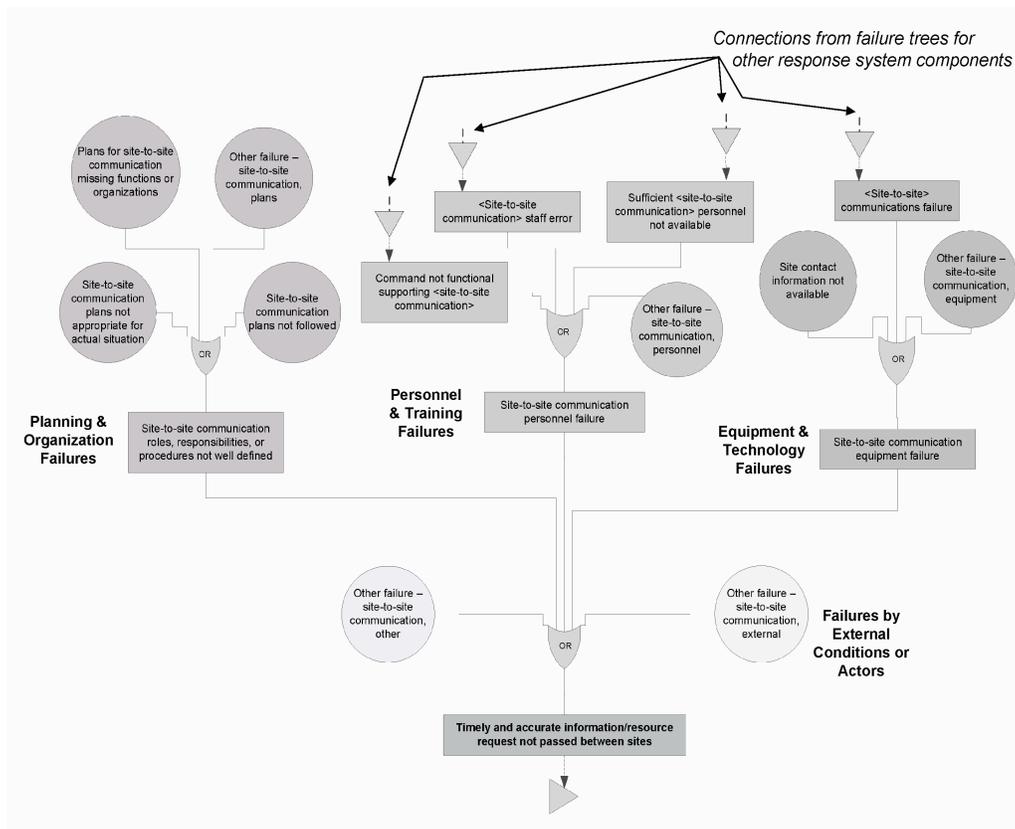


Figure 3: Example Failure Tree for Communications Between On-Scene Units & Incident Command

occurred, its performance would be lower – e.g., 600 people fed – while at incidents where everything went perfectly it could hit its designed capacity of 1,000.

This type of “what-if” or “what might go wrong” analysis is consistent with guidance in documents such as FEMA’s Comprehensive Preparedness Guide that plans be evaluated for adequacy, feasibility, and acceptability.¹⁸ In particular, systematic assessment of potential failure modes, their likelihood, and consequences captures the CPG direction that planners “assess whether their organization can accomplish the assigned mission and critical tasks by using available resources within the time contemplated by the plan” as well as that

Planners use both acceptability and feasibility tests to ensure that the mission can be accomplished with available resources, without incurring excessive risk regarding personnel, equipment, materiel, or time. They also verify that risk management procedures have identified, assessed, and applied control measures to mitigate operational risk (i.e., the risk associated with achieving operational objectives).¹⁹

This later guidance captures not only the importance of assessing the probability and consequences of different failure modes, but approaches to deal with them if they occur.

For a real system with many failure modes (e.g., the failure tree include in Figure 3 was one of more than twenty crafted to describe a realistic response system), estimates of the probability and consequences of each failure can provide the basis for simulations of response performance. Rather than walking through single cases and examining how the effects of one failure mode or another might cut into theoretical response performance, Monte Carlo simulations of many cases can be done to better reflect how possible failure modes affect the distribution of its performance. The results can be used to calculate the probability that the system will perform at or above particular capacity levels

or, put in our terms, its reliability for responding to incidents of different sizes.

Illustrative reliability curves that show those probabilities of success for several response systems at increasingly demanding incidents are shown in Figure 4. The dotted line shows a perfectly reliable system – since nothing would ever go wrong with the system’s functioning, it will perform with 100 percent reliability for any incident up to its maximum capacity. The light line shows a relatively unreliable system, which, although *designed* to deliver the same level of capacity, in fact would routinely perform much worse. The heavy line shows a more robust system, which, though its reliability drops off as incident size approaches its maximum capacity, is likely to perform well over a wide range of incidents. All three systems are of comparable reliability at very small incidents, where each has so much slack capacity that even the system with the most problems is still likely to hit the low required level of performance.

The number of failure modes a system has, the probability of failure, and the scale of failure’s effect on response performance determine the shape of these curves. Different types of failures (e.g., those with the potential to halt all response operations versus capacity reducing failures) have different effects on the shape of the curves, with their probability of occurring affecting the scale of their effect. In previous work, we have demonstrated how these types of curves – including their shape and the area under them – can provide a composite measure of preparedness (since they reflect likely performance at the full range of incidents a system might be called on to address) and as yardsticks for comparing different potential preparedness improvements. Since the area under these curves provide a measure of aggregate performance across different scale incidents, the amount that different preparedness interventions are predicted to change that area can be used to anchor cost effectiveness comparisons among them.²⁰

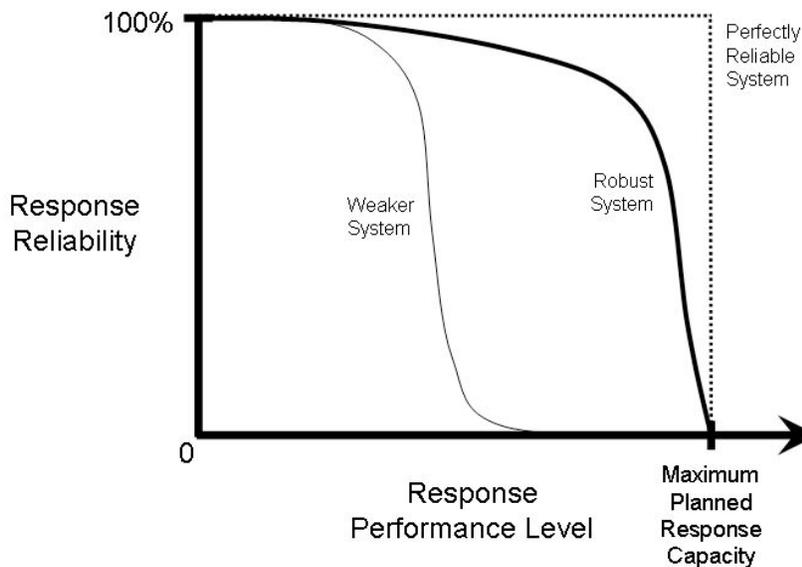


Figure 4: Illustrative Reliability Curves for Response Systems with Different Performance Characteristics

USING THE LOGIC UNDERLYING RESPONSE RELIABILITY ASSESSMENT TO DESIGN BETTER EXERCISES

In programmatic guidance from sources such as HSEEP, structured approaches for developing exercises and designing multi-year, multi-exercise programs are laid out to help frame how goals should be chosen, scenarios designed, and the actual tasks of exercise development and execution carried out. For example, in HSEEP, sequential steps and conferences are defined that identify and later flesh out the “type, scope, objectives and purpose of the exercise,”²¹ and the write and assemble the materials required for the exercise itself and evaluating its outcomes.

Key steps in this process – which shape the potential value that will be gained from actually planning and running the exercise – include selection of the portions of the response system and the response capabilities that will be exercised, the nature of the scenario that will provide the foundation for exercising those capabilities, and how different injects or challenges throughout the course of the exercise will either test specific response functions or shape the educational or training experience

of the participants. Though exercise design doctrine (e.g., the HSEEP documents referenced previously) provide processes for carrying out these design tasks, good choices of such key exercise parameters often depend on the expertise and experience of the planners involved. Furthermore, particularly if multiple organizations are involved whose priorities for the exercise differ, there may be significant divergence among participants about the correct balances to strike in the design process.

In considering the exercise design process – whether the central purpose of the exercise is training, planning, or one of the others identified previously – the logical elements of response reliability analysis could help to inform the choices made during that process and could help to provide a more common baseline to design scenarios and exercise events to meet the needs of all involved organizations.

- A **system model** for the response operation of interest (e.g., Figure 2) – by mapping out the “moving parts” in the response and capturing the full range of agencies or other organizations which would be involved in responding successfully – can help make sure that key response elements (or the agencies responsible) are not left out of an exercise.

By providing a “map” on which participating organizations can locate their functions within the response, such a model can help provide a common basis for developing a scenario that meets the needs of all participants.

- Clear **failure trees** that identify what might go wrong with specific parts of the response (e.g., Figure 3) – whether because a particular response activity inherently has many failure modes or because it depends on many other parts of the response system (and is therefore subject to problems that might arise elsewhere) – can help guide choices of what functions to exercise.²² Having failure trees available to exercise designers can also help to ensure that key failure modes aren’t left out when exercise scenarios are designed, the injects written for scripted exercises, or exercise evaluation guides prepared to help record the key informational outcomes from the exercise. In combination with a response system model, failure trees for a response can also assist in accurately interpreting exercise results. Failure modes can also provide a common language and structure for planners and organizations to negotiate about the specifics of exercise injects or events to ensure the organizational “pressure points” of interest to all participants are covered.
- Data on **specific failure modes** observed in the jurisdiction’s past responses can similarly help identify key functions that might benefit from focused exercising. Though using past experience to guide such choices is already prominent in exercise design doctrine, looking at what failure modes occurred (rather than focusing on response functions that encountered problems in past incidents) can help designers look across incident types and build more valuable future exercises. For example, if in a past chemical weapons response exercise serious problems were observed in hazmat response, a planner might conclude that another hazardous materials exercise was needed. But, if the root failure mode that caused those problems was in communications or incident command, it might be possible to cover the necessary material in a very

different exercise type, creating the opportunity to explore a very different incident and advance training goals more broadly.

None of the elements from response reliability analysis would replace steps in existing exercise design processes. However, the structure they provide for systematically thinking through what is involved in a response operation and what could disrupt it can help to make earlier steps in the exercise design process more straightforward. At the same time, by laying out a “menu” of the choices faced by exercise designers – from what to test to the specific challenges exercise participants could be presented with as a scenario evolved – these tools could help to ensure that potentially valuable details are not missed in the course of exercise design.

USING EXERCISES TO EVALUATE A RESPONSE SYSTEM’S RELIABILITY

Though the logic of reliability analysis could contribute to exercise design, using exercises as evaluation tools to assess the probability and consequences of individual failure modes could contribute to actually measuring a response system’s reliability – and, therefore, make exercises more effective as preparedness evaluation tools.

Actually assessing the reliability of a real response system requires identifying what failure modes could affect its performance and estimating their probability and consequences for response performance. A variety of strategies might be used to do so, ranging from practitioner estimation to analysis of performance in past response operations.²³ However, for assessing levels of preparedness, approaches that do not rely either on simple estimates or require waiting for a disaster to occur and a response breakdown to happen are more attractive. As a result, preparedness exercises represent a potentially attractive opportunity to gather information on response systems’ reliability characteristics. However, the ability of exercises to support this type of assessment depends on whether they are designed to measure the information needed.

The Design of an Exercise Shapes the Information It Can Provide

In thinking about exercise design, an analogy to the kinds of testing and assessment used in engineering – the field where this type of analysis was developed – is useful. Just as we would like exercises to discover or assess problems that might affect future responses, engineers want to discover problems in the technologies they design and build so steps can be taken to correct them before it becomes too costly to do so. As a result, different types of testing and experiments are performed on components of such systems or on models or prototypes to identify failure modes, assess their probability, determine their consequences – and, in some cases, determine how to maintain or service the system to prevent known failure modes from affecting performance once it is put into operation.

For some tests, systems are evaluated under conditions that are very similar to what they will be expected to face when they are actually used. Put in a language more relevant to exercise design, the scenarios the systems are subjected to are very realistic. In others tests, conditions are unrealistic by design. Tests subject technologies to very high stress to try to cause failures more quickly (to limit the amount of time and money that must be spent testing). In other tests, specific failures are caused directly – and the focus of the test is on assessing the consequences.

It is intuitive that exactly how a test is designed drives what information it can provide. Tests that cause failures directly can tell you nothing about the probability that a failure will occur, but may provide very good information on what happens when it does. To get information about the probability of a breakdown from a test that uses highly stressful conditions (e.g., testing something at a very high temperature to make it fail more quickly), the tester needs to know how to relate those conditions to what might exist under realistic circumstances. Tests that are done under entirely realistic conditions might provide both probability and consequence information, but might be very expensive to carry out (e.g., requiring testing a prototype

computer for months of continuous use until it begins to break.)

Exactly the same logic applies to exercises and the sorts of tradeoffs that exist between design choices and the information content of their results. For example, there might be design choices that are advantageous from some perspectives (e.g., reduce the cost of the exercise), but might also reduce the information the exercise produces. Just as different diagnostic tests are performed on technical systems to get different types of information, there could be very good reasons to run exercises that might not provide a complete picture of a response system – but, in that case, that understanding must be carried through to evaluating the outcomes the exercise does produce. To help illustrate these tradeoffs, we will use exercise realism as a way of working through some key design choices and exploring how they affect the information content of the “test results” provided by the exercise.

The “Realistic Exercise” as the Gold Standard

In policy discussion about preparedness exercises, “realistic” exercises are frequently put forward as the standard that exercise designers should target. For example, in testimony and analyses by the Government Accountability Office (GAO) in the years since 2001,²⁴ the need for realistic exercises has been emphasized, a finding echoed in outside analyses of national preparedness efforts.²⁵ The requirement for realism has even been embedded in legislation, with PKEMRA stating that exercises “as realistic as practicable.”²⁶ Unsurprisingly, this call for realism is embedded in exercise design doctrine.²⁷

From the perspective of response reliability analysis, realistic exercises are attractive. Most importantly, the more realistic an exercise’s scenario the more likely it will reflect the full range of possible failure modes that could affect response performance (see Figure 3). The concern that unrealistic exercises omit important elements that could affect performance has been raised as a general issue in previous policy analyses,²⁸ and in the specific examination of individual exercises and their results.²⁹ Furthermore, analogous to testing a piece of

technology under its real-life “operating conditions,” the occurrence of failures during a realistic exercise is easy to relate to their potential to occur at an actual incident and their consequences should similarly provide direct insight into how they would reduce performance at a future response.

Roles for Less-Than-Realistic Exercises

Even as policy and doctrinal sources advocate for realistic exercises, they acknowledge the potential utility of exercises that are less realistic as well. For example, both GAO and Congress (in published analyses and testimony) suggest scenarios should be intense enough, in the words of the GAO, to stress response systems “to the breaking point if possible.”³⁰ Returning to the exercise goals discussed in the introduction, there may also be good reasons why a training exercise would intentionally be unrealistic in some ways to better focus training efforts on the key points participants are expected to take away from the experience. Variations in realism are unavoidable if an exercise program includes activities ranging from seminars to full-scale exercises (Figure 1) and often realism is relaxed for valid cost and other concerns. To oversimplify somewhat, more realism usually corresponds to a more expensive exercise.

Though both exercise design doctrine and the literature include discussion of exercises of varying levels of realism, we could not find a systematic examination of the different ways the realism of an exercise is “relaxed” and how the different options affect the potential value of the information coming out of an exercise. Particularly if one goal of an exercise is assessing preparedness, a clear understanding of what types of evaluation information can be obtained from exercises designed for different levels of realism.

Examining both the literature on exercises and drawing on our earlier discussion of test design in engineering, we will look at four different ways that realism is relaxed in exercises and explore what evaluation information – specifically data for reliability assessment – can be obtained in each case.

- ***Different Exercise Types.*** Exercises conducted around a conference table or in a seminar room (Figure 1)³¹ are, by

definition, less realistic than full-scale operational exercises. From the perspective of viewing exercises as preparedness assessment tools, the concern is how different designs may – intentionally or unintentionally – foreclose the possibility of failure modes occurring that could significantly affect performance at a real incident.

A very tangible example of this possibility is that exercises that do not involve physically deploying people or response resources may not cover failure modes like vehicles breaking down or identify differences between assumed deployment rates and what is actually realistically achievable. Another, subtler, example of this issue is addressed in HSEEP guidance: “The level of detail provided in a scenario should reflect real-world uncertainty. Inclusion of superfluous information, or ‘white noise,’ is a variable that should be discussed and agreed upon by the exercise planning team.”³² It would presumably be easy for an exercise planner who was focused on building a clear and high quality tabletop exercise scenario to leave out the fact that real-world information flows are often confusing. If omitted, then the possibility of incorrect information affecting command decision-making would not be addressed – and the exercise would be blind to an important command-level failure mode.

This is not to say, however, that exercises that only capture a subset of failure modes cannot provide some data to support assessment of a response system’s reliability. Though such scenarios might omit some ways performance could break down, as long as their omission is unlikely to distort the occurrence of other failure modes (or those effects can be understood and taken into account), such exercises can provide useful data on the failure modes that are included. Failure modes that are particularly high probability (e.g., major planning shortfalls) may be quite likely to be identified in such exercises. When interpreting the results of such an exercise, it is important not to fall into the trap of drawing conclusions about failure

modes that were not included in the scenario (or the overall performance of the response system) – since the fact that they were not observed when the exercise was run was a result of its design, and had little to do with the characteristics of the response system being evaluated.

With respect to assessing the potential consequences of individual failure modes, less realistic exercises can provide more limited information. For example, a table top exercise might be used as a venue to explore the consequences of identified failure modes – though doing so would require additional analysis or assessment since they could not simply be observed directly (as might be the case in a full scale operational exercise).³³

- **High Stress Exercises.** In both the GAO report and legislation cited above, exercises that stressed response systems, potentially to their breaking point, were highlighted. Such exercises – similar to a “high temperature” test of a technological system – increase the chance that failure modes will be detected because of the demanding and hostile testing conditions. From the perspective of observing failure modes, such demanding exercises are very attractive. If an exercise was run under entirely realistic conditions, it is entirely possible that no failure modes (or at least none of much consequence) would be observed. Though that observation would still provide useful data about the response system’s performance, it would be less useful from the perspective of considering future strategies to strengthen preparedness.

Highly stressful exercise scenarios increase the chance that breakdowns will be observed in the course of the exercise either because the demands of the scenario make failure modes more likely or because the consequences of their occurring are more obvious. Identifying what breaks (and when) under stressful conditions can help to identify “weaker links” in the response system (Figure 2). This can be advantageous for identifying failures that might not have been anticipated before the exercise was run (e.g., planners believed incident

command activities would function well, but during a stressful scenario situational awareness broke down).

Relating the results of a high stress exercise back to an understanding of system performance under more normal conditions requires breaking out the nature of the stress and how it likely affects the occurrence of system failure modes. If the scale of the exercise scenario is within the system’s expected capabilities – i.e., looking at Figure 4, it is an incident whose requirements fall to the right of the graph, but do not exceed the maximum planned capacity of the system – then no further analysis is needed.³⁴

However, if the exercise scenario was intentionally selected to fall above (or far above) the maximum planned capacity of the system,³⁵ then it is likely that the overwhelming nature of the event could change the probability and consequences of failure modes occurring – and so relating the observations back to the system’s expected performance at smaller scale incidents would have to be done with care. It may be the case that the fact that particular failures were observed in these scenarios indicates they would be problems for smaller incidents as well, but before drawing that conclusion a critical examination must be done to determine if the stress level of the exercise makes doing so impossible.

- **Exercise Scenarios Which Force Failures to Occur.** In the design of exercises, scripted series of events are frequently used to ensure that the participating response organizations test the response capabilities or explore the policy issues the exercise was planned to address. The “injects” that are included as part of an exercise scenario could include potential failure modes occurring in a scripted way – e.g., an exercise inject stating that the communications system has broken down, certain supplies are running out, etc. In such situations, part of what the exercise is testing is the ability of the participants to adapt or improvise and prevent the *potential* failure from

becoming an *actual* failure with an impact on response performance.

There are real limits to conclusions that can be drawn about the probabilities of different failures based on the outcome of such an exercise. Since a failure was “forced” in the scenario, its occurrence has no information content.³⁶ However, the extent to which the response system was able to adapt to the failure and mitigate its impact on response performance is informative. The ability of a response system to prevent a potential failure from affecting performance does provide some evidence that the system’s performance is less likely to be affected by that failure mode at an actual incident.³⁷ Furthermore, even if the response system cannot adapt to a forced failure, such an approach could be very useful for assessing the consequences of individual failure modes since evaluation could focus on gathering data of interest immediately after the scripted (and therefore anticipated by the evaluation team) failure occurred (discussed below)

- ***Exercises Testing Parts of Response Systems in Isolation.*** Finally, some exercises – notably drills of specific response functions – are designed to test pieces of the response system individually. Though such focus can allow more detailed examination of parts of the response,³⁸ understanding how what is observed in a drill relates to response performance overall depends on understanding the effects of isolating it. For example, such isolation will leave out failure modes caused by linkages to other pieces of the response system (Figure 3, arrows entering the tree at the top). If the effects of these failures are understood and potentially simulated as part of the focused exercise, then it might be possible to make inferences about system performance from problems observed in the drill. For assessing the consequences of individual failure modes, the “linkages out” – how performance in the function being exercised relates to the rest of the system – must be taken into account. All other things being equal, a breakdown that has the potential to affect many other

parts of the response system will be more consequential than one that does not.

Though potentially not as comprehensive as realistic response exercises, those in which reality has been relaxed in various ways can still provide insights useful for assessing a response system’s reliability characteristics. Furthermore, the systematic thought process about what an individual exercise is testing, which reliability analysis provides, helps to interpret exercise results more generally and draw conclusions about what specific exercises can reveal about preparedness. Particular exercise designs – e.g., those that look at “forced failures” or individual parts of the response systems – may be very attractive for some exercise goals. But by helping to identify what types of evaluation data can still be extracted from the results of such exercises, this approach can help to increase the potential value of individual exercises and their ability to pursue multiple exercise goals simultaneously.

GENERALIZING FROM EXERCISE RESULTS TO DESCRIBE RESPONSE RELIABILITY

Exercises of various designs can clearly provide an approach to collecting data on a response system’s reliability characteristics. Either in a realistic or an appropriately crafted less-than-realistic exercise, the observation of a failure mode with non-trivial consequences for response performance provides one data point for reliability analysis. Comparison of absolute performance in an exercise (e.g., the number of people evacuated in a given time) with planning assumptions for that capability can both help characterize the consequences of observed failures and suggest the presence of failure modes that may have occurred but not been recorded in exercise evaluation. Both these approaches can contribute to identifying failure modes that affect a system, observe the occurrence of some of those modes during a particular scenario, and explore their consequences.

In considering the overall reliability of a response system, there is still a ways between the results of one exercise and the type of

reliability curves shown at the beginning of this article. Just as we drew a contrast between a single observation of a response system and simulation of many cases – which made it possible to draw the type of reliability curves shown in Figure 4 – one exercise represents essentially one observation of a system. Because the occurrence of failure modes is a probabilistic process – a system failure of modest probability will occur in some responses but not others – drawing more general conclusions about system reliability requires more than one data point.

As a result, an exercise driven assessment of a response system's reliability is best considered a target of an entire exercise program, where the results of multiple exercises – and the experience in actual incidents as well – are combined to identify failure modes that occur more frequently than others or ones whose consequences appear more consistently serious. To build that composite picture, individual failure modes provide a common framework to link the occurrence of – for example, command and control problems – from tabletop exercises on unconventional weapons scenarios to an operational exercise on responding to a hurricane. Such an analytical model would fit readily into existing guidance on corrective action programs that are designed to capture the improvements identified in past exercises and assist in prioritizing among those improvements and allocating time and resources to those that are more common, more consequential, or a combination of both.

CONCLUSIONS

In efforts both to assess and improve national preparedness, exercises occupy a prominent place in both policy and practitioner thinking. When an exercise is successfully completed, the results are cited as a demonstration of preparedness. When a response operation does not go as expected, more – or different – exercise regimens are often cited as part of the solution to improve preparedness. In some cases, past exercise experiences are cited as evidence that lessons that “should have been learned” were not absorbed, and

changes in exercising and preparedness efforts demanded as a result.

The weight given to exercises, and the frequency with which they are run, make understanding how to get the most value out of these activities an important topic that merits analytic attention. Numerous programs require exercising as part of preparedness efforts for a range of hazards, and the amount of money devoted to meeting those requirements is considerable. Improvements to exercise design that strengthen their effectiveness for evaluation or other goals – particularly improvements that can be made that do not increase the costs of individual exercises – will increase the return on that substantial investment.

In this article, we have discussed how response reliability analysis – systematically looking at response operations at systems and assessing what might go wrong that would hurt their performance – can be applied to exercise design. Because of the need to improve efforts to measure emergency preparedness, we focused primarily on how exercises can be used to collect the information needed to assess the probability and consequences of different failure modes that could prevent a response system from responding effectively to a future incident. Though highly realistic exercises are potentially very useful for that assessment, exercises that relax realism in a different ways can produce useful information as well. Just as tests of technological systems are done under both realistic and intentionally unrealistic circumstances, exercise tests of response systems can – and should – be conducted in this manner as well. However, when exercises are run where realism has been relaxed in different ways, the limits on the information obtained must be recognized – and the framework provided by response reliability analysis can help to ensure that the conclusions drawn from results neither over- nor understate what the exercise revealed about the preparedness of the response system being tested.

About the Authors

Brian A. Jackson is a senior physical scientist at the RAND Corporation, where he has managed and carried out homeland security research efforts since 2000. His research has focused on preparedness, response operations management, and organizational behavior associated with both emergency management and terrorism. He also teaches on homeland security technology issues in Georgetown University's Security Studies Program. Mr. Jackson may be contacted at bjackson@rand.org.

Shawn McKay joined the RAND Corporation January 2010 after receiving a PhD from Purdue University. At Purdue, Shawn was awarded the Homeland Security Science, Technology, Engineering, and Mathematics Fellowship (HS-STEM), participated in many DHS events on and off campus, and conducted DHS research in fusion of sensor technologies. Throughout his career, Shawn has conducted a variety of reliability analyses for government clients and while employed at Intel and Honeywell.

ACKNOWLEDGEMENT

We would like to gratefully acknowledge our colleague Kay Sullivan Faith who was involved in building the Figures 2 and 3 we use as examples.

¹ Federal Emergency Management Agency, *Developing and Maintaining State, Territorial, Tribal, and Local Government Emergency Plans: Comprehensive Preparedness Guide-101* (FEMA, 2009), 3.17, 3.20-3.22.

² Adapted from U.S. Department of Homeland Security, *Homeland Security Exercise and Evaluation Program*, Volume I (DHS, February 2007), 5. Hereafter referred to as HSEEP.

³ This set of five exercise goals is based on the HSEEP Exercise Guidance, Sample Objectives for Operations-Based; Sample Objectives for Discussion-Based Exercises; HSEEP, Vol.1; and IS-120.A, *An Introduction to Exercises*, all available at <https://hseep.dhs.gov>.

⁴ Some sources draw distinctions between the types of exercises best suited to different goals. For example, distinguishing between larger, full-scale exercises for evaluation and more focused drills for training; see discussion in International Atomic Energy Agency (IAEA), "Method for Developing Arrangements for Response to a Nuclear or Radiological Emergency," *EPR-METHOD 2003* (October 2003), 24.

⁵ One challenge in examining preparedness exercises and their results are dissemination restrictions placed on documents describing exercises and their outcomes. For example, the Department of Homeland Security's LLIS – one centralized (though admittedly non-comprehensive) source of data on past preparedness exercises – includes many exercise documents and after-action reports (AAR) designated "For Official Use Only." Though potentially available for analysis, such a designation forbids quotation of their content in publication in the open literature. Though in the course of the research resulting in this article we examined a broader set of exercise AAR, in the text we reference only exercise descriptions or analyses published in the open literature without restriction. It is also worth noting that, paradoxically, even some AAR posted on the internet, apparently by the organizations authoring them, are still marked as For Official Use Only with its associated dissemination restrictions.

⁶ See, for example, published examples of exercises that focused on training goals as described in the following: D.M. Peterson and R.W. Perry, "The impacts of disaster exercises on participants," *Disaster Prevention and Management* 8, no. 4 (1999): 241-254; R.W. Perry, "Disaster Exercise Outcomes for Professional Emergency Personnel and Citizen Volunteers," *Journal of Contingencies and Crisis Management* 12, no. 4 (2004): 64-75; S.A. Sarpy, C.R. Warren, S. Kaplan, J. Bradley and R. Howe, "Simulating Public Health Response to a Severe Acute Respiratory Syndrome (SARS) Event: A Comprehensive and Systematic Approach to Designing, Implementing, and Evaluating a Tabletop Exercise," *Journal of Public Health Management Practice*, Supplement (November 2005): S75-S82; B.H. Bartley, J.B. Stella, and L.D. Walsh, "What a Disaster?! Assessing Utility of Simulated Disaster Exercise and Educational Process for Improving Hospital Preparedness," *Prehospital and Disaster Medicine* 21, no. 4 (2006): 249-255; E. Savoia, P.D. Biddinger, P. Fox, D.E. Levin, L. Stone, and M.A. Stoto, "Impact of Tabletop Exercises on Participants' Knowledge of and Confidence in Legal Authorities for Infectious Disease Emergencies," *Disaster Medicine and Public Health Preparedness* 3 (2009): 104-110; and R. Silenas, R. Akins, A.R. Parrish, and J.C. Edwards, "Developing Disaster Preparedness Competence: An Experiential Learning Exercise for Multiprofessional Education," *Teaching and Learning in Medicine* 20, no. 1 (2008): 62-68. See also, J. Dudte, "Planning to Train: What is the Objective?" *FireEMS*, January 2005, 19-24. Examples of planning exercises include E.H. High, K.A. Lovelace, B.M. Gansneder, R.W. Strack, B. Callahan, and P. Benson (2010) "Promoting Community Preparedness: Lessons Learned from the Implementation of a Chemical Disaster Tabletop Exercise," *Health Promotion Practice* 11, no. 3, (2010): 310-9H; N. Lurie, R.B. Valdez, J. Wasserman, M. Stoto, S. Myers, R. Molander, S. Asch, B.D. Mussington, and V. Solomon, *Public Health Preparedness in California: Lessons from Seven Jurisdictions*, TR-181 (Santa Monica, CA: RAND Corporation, 2004); D.J. Dausey, J.W. Buehler, and N. Lurie, "Designing and conducting tabletop exercises to assess public health preparedness for manmade and naturally occurring biological threats," *BMC Public Health* 7, no. 92 (2007), <http://www.biomedcentral.com/1471-2458/7/92>; B. A. Jackson, J.W. Buehler, D. Cole, S. Cookson, D.J. Dausey, L. Honess-Morreale, S. Lance, R.C. Molander, P. O'Neal, and N. Lurie, (2006) "Bioterrorism with Zoonotic Disease: Public Health Preparedness Lessons from a Multiagency Exercise," *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 4, no. 3 (2006): 287-292; and D. Jarrett, "Lessons Learned: The 'Pale Horse' Bioterrorism Response Exercise," *Disaster Management & Response* 1, no. 4 (2003): 114-118.

⁷ See, for example, discussion in V.J. Doherty, “Metrics for Success: Using Metrics in Exercises to Assess the Preparedness of the Fire Service in Homeland Security” (master’s thesis, Naval Postgraduate School, Monterey, California, 2004); K.M. Gebbie, J. Valas, J. Merrill, and S. Morse, “Role of Exercises and Drills in the Evaluation of Public Health in Emergency Response,” *Prehospital and Disaster Medicine* 21, no. 3 (2006): 173-182; L. Sugarman, E. Eiseman, A. Jain, N. Nicosia, S. Stern, and J. Wasserman, *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*, TR-249-DHHS (Santa Monica, CA: RAND Corporation, 2005); C.C. Bradshaw and T.A. Bartenfeld, “Exercise Evaluation Guides for Public Health Emergency Preparedness,” *Homeland Security Affairs* V, no. 3 (2009), <http://www.hsaj.org>; S.A. Sarpy, C.R. Warren, S. Kaplan, J. Bradley and R. Howe, “Simulating Public Health Response to a Severe Acute Respiratory Syndrome (SARS) Event: A Comprehensive and Systematic Approach to Designing, Implementing, and Evaluating a Tabletop Exercise,” *Journal of Public Health Management Practice*, Supplement (November 2005): S75–S82; K.R. Klein, D.C. Brandenburg, J.G. Atas, and A. Maher, “The Use of Trained Observers as an Evaluation Tool for a Multi-Hospital Bioterrorism Exercise,” *Prehospital and Disaster Medicine* 20, no. 3 (2005): 159-163; E. Savoia, P.D. Biddinger, J. Burstein, and M.A. Stoto, “Inter-Agency Communication and Operations Capabilities during a Hospital Functional Exercise: Reliability and Validity of a Measurement Tool,” *Prehospital and Disaster Medicine* 25, no. 1 (2010): 52-58; A.H. Kaji, V. Langford, and R.J. Lewis, “Assessing hospital disaster preparedness: a comparison of an on-site survey, directly observed drill performance, and video analysis of teamwork,” *Annals of Emergency Medicine* 52, no. 3 (2008): 195-201; and E.A. Prebles, A.D. Sayhir, D.C. Brandenburg, and E.C. Mather, “Longitudinal Evaluation of Food Safety Discussion-Based Exercises: Tool Development and Initial Validation,” *Journal of Homeland Security and Emergency Management* 5, no. 1 (2008). Examples are available of exercises seeking to evaluate very specific things (e.g., S. Phelps, “Mission Failure: Emergency Medical Services Response to Chemical, Biological, Radiological, Nuclear, and Explosive Events,” *Prehospital and Disaster Medicine* 22, no. 4 (2007): 293-296; T. Zerwekh, J. McKnight, N. Hupert, D. Wattson, L. Hendrickson, and D. Lane, “Mass Medication Modeling in Response to Public Health Emergencies: Outcomes of a Drive-thru Exercise,” *Journal of Public Health Management Practice* 13, no. 1 (2007): 7-15; A. Brody, J.L. Kashuk, E.E. Moore, C. Barnett, W.L. Biff, C.C. Burlew, J.L. Johnson, A. Sabel, and C. Colwell, “Live Victim Volunteers Enhance Performance Improvement (PI) in Mass Casualty Incident (MCI) Drills,” *Journal of Surgical Research* 158, no. 2 (2010): 253; Stergachis, A., et al., 2007; C. Nelson, E.W. Chan, C. Fan, D. Lotstein, L.B. Caldarone, S.R. Shelton, A.L. Maletic, A.M. Parker, A. Felton, A. Pomeroy, and E.M. Sloss, *New Tools for Assessing State and Local Capabilities for Countermeasure Delivery*, TR-665-DHHS (Santa Monica, CA: RAND Corporation, 2009) and to assess preparedness much more broadly (e.g., E. Jasper, M. Miller, B. Sweeney, D. Berg, E. Feuer, and D. Reganato, “Preparedness of Hospitals to Respond to a Radiological Terrorism Event as Assessed by a Full-Scale Exercise,” *Journal of Public Health Management Practice*, Supplement (November 2005): S11–S16; High, E.H., et al., “Promoting Community Preparedness”; J.L. Taylor, B.J. Roup, D. Blythe, G.K. Reed, T.A. Tate, and K.A. Moore, “Pandemic Influenza Preparedness in Maryland: Improving Readiness Through a Tabletop Exercise,” *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 3, no. 1 (2005): 61-69; D.J. FitzGerald, M.D. Sztajnkrzyer, and T.J. Crocco, “Chemical Weapon Functional Exercise—Cincinnati: Observations and Lessons Learned from a ‘Typical Medium-Sized’ City’s Response to a Simulated Terrorism Utilizing Weapons of Mass Destruction,” *Public Health Reports* 118 (2003): 205-214; S.E. Cosgrove, M.W. Jenckes, L.M. Wilson, E.B. Bass, and E.B. Hsu, *Tool for Evaluating Core Elements of Hospital Disaster Drills* (AHRQ Publication 08-0019, June 2008).

⁸ For example, extrapolating from the specifics of the exercise scenario to preparedness more generally. The written AAR produced from exercises routinely include both strengths and areas for improvement (described at <https://hseep.dhs.gov>), which have obvious value, but are difficult to use as a basis for broader generalizations about the area’s preparedness. In discussions with response practitioners in the course of this and related work, an “art” was described about what is and is not included in exercise AAR that reinforced concerns about using them as a data source for drawing conclusions about an area’s level of preparedness.

⁹ See discussion and references cited in B.A. Jackson, *The Problem of Measuring Emergency Preparedness: The Need for Assessing “Response Reliability” as Part of Homeland Security Planning*, OP-234-RC (Santa Monica, CA: RAND Corporation, 2008), 5–10; C. Nelson, E.W. Chan, C. Fan, D. Lotstein, L.B. Caldarone, S.R. Shelton, A.L. Maletic, A.M. Parker, A. Felton, A. Pomeroy, and E.M. Sloss, *New Tools for Assessing State and Local Capabilities for Countermeasure Delivery*, TR-665-DHHS (Santa Monica, CA: RAND Corporation, 2009); C. Nelson, N. Lurie, and J. Wasserman, “Assessing Public Health Emergency Preparedness: Concepts, Tools, and Challenges,” *Annual Review of Public Health* 28 (2007): 12.1–12.18; H.H. Willis, C. Nelson, S. R. Shelton, A. M. Parker, J. A. Zambrano, E. W. Chan, J. Wasserman, and B. A. Jackson, *Initial Evaluation of the Cities Readiness Initiative*, TR-640-CDC (Santa Monica, CA: RAND Corporation, 2009). Federal Emergency Management Agency, *The Federal Preparedness Report* (FEMA, January 13, 2009), 111–113, includes a list of current preparedness-assessment-related systems and some description of their content.

¹⁰ This body of work includes: B.A. Jackson, K. Sullivan Faith, and H.H. Willis, *Evaluating the Reliability of Emergency Response Systems for Large-Scale Incident Operations*, MG-994-FEMA (Santa Monica, CA: RAND Corporation, 2010); B.A. Jackson, K. Sullivan Faith, and H.H. Willis, “Are We Prepared? Using Reliability Analysis to Evaluate Emergency Response Systems,” *Journal of Contingencies and Crisis Management* (forthcoming); and K. Sullivan Faith, B.A. Jackson, and H.H. Willis, “Text Analysis of After Action Reports to Support Improved Emergency Response Planning,” (manuscript in preparation).

¹¹ Elements of what we call response reliability are addressed in operations research analyses of response systems, though the focus of such work is generally response to everyday emergencies (e.g., distributions of response times to fires from defined fire station locations) rather than for large-scale emergencies served by more complex and ad hoc response networks.

¹² This potential has been highlighted explicitly by the Government Accountability Office in a recent review of national preparedness: “Exercises that stress the preparedness system in a realistic way are key to testing the prospective reliability of a response and determining whether plans have accounted for potential breakdowns with relatively greater consequences.” (Government Accountability Office, “National Preparedness: FEMA Has Made Progress, but Needs to Complete and Integrate Planning, Exercise, and Assessment Efforts,” GAO-09-369 (Washington, DC: Government Accountability Office, April 2009), 48-49, emphasis added)

¹³ See, for example, R.E. Peterson, B.R. Lindsay, L. Kapp, E.C. Liu, and D.R. Peterman, “Homeland Emergency Preparedness and the National Exercise Program: Background, Policy Implications, and Issues for Congress,” RL34737 (Washington, DC: Congressional Research Service, 2008) for a review of a variety of federal level efforts and discussion in Government Accountability Office, “National Preparedness: FEMA Has Made Progress, but Needs to Complete and Integrate Planning, Exercise, and Assessment Efforts,” GAO-09-369 (Washington, DC: Government Accountability Office, April 2009) describing the challenges and concerns from the federal perspective about exercise program implementation.

¹⁴ This technique is described in a variety of sources, including C.E. Ebeling, *An Introduction to Reliability and Maintainability Engineering* (New York, N.Y.: McGraw Hill, 1997), 166–173; Mohammad Modarres, Mark Kaminskiy, and Vasilij Krivtsov, *Reliability Engineering and Risk Analysis: A Practical Guide* (New York, N.Y.: Marcel Dekker, 1999), 262–267; U.S. Department of Defense, *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*, MIL-STD-1629A (DHS, November 24, 1980); United States Army, *Failure Modes, Effects and Criticality Analysis (FMECA) for Command, Control, Communications, Computer, Intelligence, Surveillance, and Reconnaissance (C4ISR) Facilities*, TM 5-698-4 (U.S. Army, September 29, 2006); U.S. Nuclear Regulatory Commission, *Fault Tree Handbook*, NUREG-0492 (January 1981); Federal Aviation Administration, *FAA System Safety Handbook*, Chapter 9: Analysis Techniques (FAA, December 30, 2000).

¹⁵ See, for example, D. Lotstein, K. J. Leuschner, K. A. Ricci, J. S. Ringel, and N. Lurie, *PREPARE for Pandemic Influenza: A Quality Improvement Toolkit*, TR-598-RWJ (Santa Monica, CA: RAND Corporation, 2008).

¹⁶ Ebeling, *An Introduction to Reliability and Maintainability Engineering*, 168.

¹⁷ The analysis process and detailed examination of some example cases are published in Jackson, et al, *Evaluating the Reliability of Emergency Response Systems*.

¹⁸ FEMA, *Comprehensive Preparedness Guide-101*, 3.21.

¹⁹ Ibid.

²⁰ Jackson, et al., *Evaluating the Reliability of Emergency Response Systems*; and Jackson, et al., “Are We Prepared?”

²¹ DHS, *Homeland Security Exercise and Evaluation Program*, Volume I (DHS, February 2007), 15.

²² This logic matches that in analyses relating to exercises for public health preparedness, where a focus in the choice of what functions to drill include focusing on the most “failure prone” activities (see, Nelson, C. et al., *New Tools for Assessing State and Local Capabilities*). In a conversation about the technique with one local response practitioner, the failure trees were likened to a “preventive maintenance guide,” by flagging key parts of the response system whose potential failure modes meant they were worth exercising more frequently – or more intensively.

²³ For example, in Jackson, et al., *Evaluating the Reliability of Emergency Response Systems*, we discuss an effort to use response after-action reports as a data source for reliability analysis.

²⁴ See, for example, W.O. Jenkins, Jr., *Emergency Preparedness and Response: Some Issues and Challenges Associated with Major Emergency Incidents*, GAO-06-467T (Washington, DC: Government Accountability Office, February 23, 2006) Jenkins, Jr., W.O., 2006; Government Accountability Office (GAO), “National Preparedness: FEMA Has Made Progress, but Needs to Complete and Integrate Planning, Exercise, and Assessment Efforts,” GAO-09-369 (Washington, DC: Government Accountability Office, April 2009), 48.

²⁵ For example, R. Clarke and R. Beers, eds., *The Forgotten Homeland* (New York, NY: The Century Foundation Press, 2006).

²⁶ 6 U.S.C. § 748(b)(2)(A)(i), as cited in GAO, “National Preparedness.”

²⁷ For example, DHS, *Homeland Security Exercise and Evaluation Program*, Volume II (DHS, February 2007), 13.

²⁸ Clarke and Beers, *The Forgotten Homeland*, 21, 31-2.

²⁹ For example, R.R. Ferrer, M. Ramirez, K. Sauser, E. Iverson, and J.S. Upperman, in “Emergency drills and exercises in healthcare organizations: assessment of pediatric population involvement using after-action reports,” *American Journal of Disaster Med* 4, no. 1 (2009): 23-32, examine a variety of hospital preparedness exercises and discuss the fact that most leave out pediatric populations as considerations in the exercise and as “player-victims” when run. The concern is that such populations pose unique challenges and have specific needs, and their omission produces a blind spot with respect to – in our terminology – associated system failure modes.

³⁰ GAO, “National Preparedness.”

³¹ See also DHS, HSEEP, Volume I, 9.

³² DHS, HSEEP, Volume II, 13.

³³ Another way of considering the application of exercises of varying realism levels is that different types of exercises may be more applicable to different steps of response reliability analyses, with less realistic or complete ones (left, Figure 1) being potentially more valuable for the first two steps (mapping out the structure of a response system and identifying potential failure modes) and the more realistic ones (right, Figure 1) being more useful for likelihood and consequence analysis.

³⁴ That is, there is no reason to believe that the probabilities or consequences of failure modes would have changed because the nature of the scenario was very different from what the system was designed to do.

³⁵ Which might be done, for example, in a training exercise regarding truly catastrophic events – or at an interagency planning activity to explore how response operations might be staged for a true outlier incident.

³⁶ J. Dudte, “Planning to Train: What is the Objective?” *FireEMS* (January 2005), 19-24, discusses the potential problem of exercise designers injecting many such failures – especially ones that were not examined and assessed during exercise planning – in part because of how doing so complicates evaluating the outcomes of the exercise.

³⁷ In spite of the effect on the information the exercise can provide, there are a number of reasons exercises are designed this way. For training purposes, such scenarios focus attention on specific problems of interest to focus learning. Furthermore, exercise scenarios that try to force specific failure modes could be used to evaluate past investments intended to address those breakdown paths.

³⁸ See, for example, DHS, HSEEP, Volume II, 13.



Copyright © 2011 by the author(s). Homeland Security Affairs is an academic journal available free of charge to individuals and institutions. Because the purpose of this publication is the widest possible dissemination of knowledge, copies of this journal and the articles contained herein may be printed or downloaded and redistributed for personal, research or educational purposes free of charge and without permission. Any commercial use of Homeland Security Affairs or the articles published herein is expressly prohibited without the written consent of the copyright holder. The copyright of all articles published in Homeland Security Affairs rests with the author(s) of the article. Homeland Security Affairs is the online journal of the Naval Postgraduate School Center for Homeland Defense and Security (CHDS).

<http://www.hsaj.org>

