

ORIGINAL RESEARCH

The Challenge of Measuring Emergency Preparedness: Integrating Component Metrics to Build System-Level Measures for Strategic National Stockpile Operations

Brian A. Jackson, PhD, MA, and Kay Sullivan Faith, MS

ABSTRACT

Objective: Although significant progress has been made in measuring public health emergency preparedness, system-level performance measures are lacking. This report examines a potential approach to such measures for Strategic National Stockpile (SNS) operations.

Methods: We adapted an engineering analytic technique used to assess the reliability of technological systems—failure mode and effects analysis—to assess preparedness. That technique, which includes systematic mapping of the response system and identification of possible breakdowns that affect performance, provides a path to use data from existing SNS assessment tools to estimate likely future performance of the system overall.

Results: Systems models of SNS operations were constructed and failure mode analyses were performed for each component. Linking data from existing assessments, including the technical assistance review and functional drills, to reliability assessment was demonstrated using publicly available information. The use of failure mode and effects estimates to assess overall response system reliability was demonstrated with a simple simulation example.

Conclusions: Reliability analysis appears an attractive way to integrate information from the substantial investment in detailed assessments for stockpile delivery and dispensing to provide a view of likely future response performance. (*Disaster Med Public Health Preparedness*. 2013;7:96-104)

Key Words: emergency response, bioterrorism preparedness, dispensing, Strategic National Stockpile, performance measures

During the past decade, significant investments in public health emergency preparedness have been made in response to health threats from natural and human sources. To build the analytic and evaluation infrastructure needed to assess the value of those investments, set priorities, and determine local, regional, or national readiness for events of concern, complementary investments have been made in the development of measures and metrics for preparedness.

The Strategic National Stockpile (SNS) and the delivery of the pharmaceuticals that it contains are prominent components of the U.S. public health emergency preparedness effort. As part of the cities readiness initiative (CRI), a clear goal has been defined for part of SNS operations—the delivery of antibiotics to the entire population of a metropolitan area within 48 hours.¹

Substantial investments have been made to develop evaluation tools for SNS operations. The technical

assistance review (TAR), a part of the assistance and evaluation provided to states and localities through relevant federal programs, collects data on elements of SNS preparedness at high resolution.^{2,3} Standards for dispensing operations have been developed that identify specific requirements for planning at a higher level of detail than the overall CRI performance standard.⁴ Drills and exercises that test different parts of a dispensing system in a modular way have been designed and field tested, and are used nationwide.⁵

Although substantial progress has been made, fundamental measurement challenges remain. Even in intensively-assessed functions such as SNS operations, the ability to predict how a specific public health system is likely to perform during future incidents of different scales or complexities is still elusive, and efforts to develop composite performance measures are ongoing.⁶ The National Health Security Strategy, which focuses on building an integrated, national-level

capability, provides added impetus to develop integrated performance metrics for public health preparedness activities.

The need for composite, system-level measures that provide insight into likely future performance is illustrated by hypothetical SNS delivery and dispensing systems from 2 areas that are otherwise identical:

- When preparing for a CRI-type deployment, the first area plans to use the minimum number of points of dispensing (PODs) and building staff based on estimates of how many people could be assisted at each location. Because of resource constraints, the plan anticipates adapting existing resources to the task, such as communication and transportation systems, rather than obtaining resources dedicated to SNS operations. Similarly, training and exercises are kept to a minimum to avoid disrupting existing emergency operations or inconveniencing volunteers.
- The second area also calculates the minimum PODs and staff needed to meet the CRI standard, and adds an extra 25% “safety margin” to each, potentially serving more than the total population of their metropolitan area. Dedicated resources are established to speed communication and management of PODs and to operate the logistics system needed for SNS deployment. Regular exercises are included to keep plan requirements and training fresh for professional responders and the volunteers who would be called in for an actual event.

Both approaches are legitimate for planning for a contingency of high consequence but low probability. In spite of their differences, both areas might appear similarly prepared to receive and use the SNS. In the first case, however, meeting the CRI requirement during an actual incident would require the system to operate at peak efficiency, with nothing causing a slow down or interruption in operations. The fact that its preparedness program minimizes training and relies on systems that might not be well suited to SNS-related operations could increase the chance that something *will* go wrong and the response will not go as planned. In the second case, meeting the goal does not require that everything go right; building extra capacity into the system would mean that it could absorb some problems and still reach its intended outcome. In contrast to the first, the second system’s dedicated resources and the higher level of training might also reduce the chance that problems that hurt performance would occur.

The central difference between the 2 hypothetical systems is their *response reliability*.⁷ Both might be prepared to meet the CRI performance goal in principle, but the *probability* is much higher that the second system will be able to perform as planned in a future incident. Even though the CRI goal specifies a time and performance target for dispensing operations, without addressing the likelihood that plans will achieve that target, it provides only a partial picture of preparedness.

Response systems can be made more reliable by adopting hedging strategies such as providing extra capacity or building systems that are less likely to break down, as illustrated in the examples above. Therefore, a response system that is almost certain to be able to deliver antibiotics to everyone in its area of responsibility (achieving the CRI goal of near 100% reliability) may look very different than one that has a 70%, 80% or even a 90% chance of doing so. It will likely be more costly as well, because redundancy, more frequent training, and other measures that can make response systems more reliable impose an associated “shadow price” to build and maintain them. Which of these is the “right amount” of reliability is a policy choice, but one with significant resource and other implications.

Making that policy choice requires metrics that provide insight into the reliability of different systems, and enable cost-benefit tradeoffs among policy options that either strengthen or weaken response system reliability. In this report, we describe research focused on building such metrics, including the adaptation of engineering reliability analysis techniques to analyzing SNS delivery and dispensing systems, the demonstration of how those techniques could be used to make reliability estimates via numerical simulation of a simple example system, and the examination of how existing assessment tools for SNS preparedness could be integrated into reliability analysis.

METHODS

Basically, assessing response system reliability is achieved by adding the concept of probability to preparedness measurement; that is, reliability represents the probability that a system will achieve a level of performance within a desired time. Any event that potentially can halt response activities entirely, reduce the efficiency or effectiveness of those activities by delaying them, decrease the capacity of the response system, or decrease the ability of response personnel or assets to achieve their missions when deployed will reduce reliability. The more likely such an event is, and the larger its effect on response performance, the greater its reliability reduction. The CRI goal provides a target for a response system, but in reality a system will have different reliability levels for different-sized incidents, with more reliable operations at smaller events that are well below the system’s maximum capacity compared to incidents closer to the upper limits of the system’s performance.

The simplest way to assess the reliability of a response system for a particular incident is to consider the scale of the incident versus the expected maximum capacity of the system. If a great deal of extra capacity above the level needed to meet the requirements of the incident (eg, delivering antibiotics to a small fraction of the area) exists, that extra amount would allow the system to absorb many faults or events that hurt performance and still perform at the required level.

Under such circumstances, reliability would be expected to be high, as the probability of enough things going wrong to exhaust the extra capacity would presumably be low. However, using this method, demanding scenarios that require the entire capacity of an emergency response system would be expected to have a low probability of going well. The CRI goal—treating everyone in a protected area—is one such scenario. A more detailed approach, one that can target scarce resources to high-impact areas of improvement, is needed to plan for a situation that pushes the limits of a response system.

Deeper insight into system reliability can be gained by breaking the functions of the system (ie, the national, regional, and local delivery and dispensing activities needed to move material from the SNS and provide it to the population) into their component parts and asking specific questions about the types of problems that might arise and their likely effect on response outcomes. Such an analysis is what an engineer would call a “failure mode and effects analysis,” which identifies individual events that could affect performance and, based on data about the response system, estimates their probability and consequences. The technique and its use for systems from technical systems to technical-human systems such as nuclear power plant operations are described in numerous literature sources.^{8,9} With a sufficiently detailed analysis, the probability that the entire system can achieve specific performance levels can be explored by combining the likelihood and effects of the many individual failure modes that could occur.

Using this technique to assess SNS delivery and dispensing operations, we constructed a conceptual model of such a response, based on publicly available policy documents and previous analyses. Applying the failure mode, effects and criticality analysis technique, we identified potential problems by systematically looking at how personnel problems, technical breakdowns, management and coordination faults, and other causes could affect performance in each part of the model. This analysis produced a series of *failure trees*, indicating how each type of failure affects each part of the SNS dispensing system. Failure trees were produced for separable functional pieces of the system (eg, incident management, dispensing operations at a POD) or separable tasks (eg, requesting SNS delivery).

Building failure trees for a system is the first step in assessing its reliability. Although taking inventory of potential failure modes can be useful in planning, making reliability *estimates* requires systematically estimating the likelihood of each failure mode occurring and its consequences to response operations. With such estimates, overall system reliability—and therefore preparedness—can be explored using basic simulation techniques. Overall performance at an individual response operation will be determined by what failure modes occur (or what combination of failure modes occur simultaneously) and the resulting total dispensing performance.

It should be noted that these estimates could be done at varied levels of detail and realism. The simplest approach is to treat failure modes as being independent; that is, one failure mode does not affect the probability of a different failure mode occurring, or its consequences if it does (beyond those if both randomly occur simultaneously). Many failure modes in a response operation are reasonably independent from one another, while many others are not. For example, breakdowns in management or messaging that create public confusion about dispensing practices or availability of supplies could result in behavior by patients that affect dispensing rate, security at POD sites, and so on. Adding such interactions in a detailed and quantitative way would make this type of assessment more realistic, but would also become more labor and time intensive. Our proof of concept takes a middle path, where failures occurring in one part of dispensing operations (eg, dysfunction in incident management) are themselves potential causes of failures elsewhere (eg, POD operations). This cross-linking of different elements of the system model allows failure modes and their effects to cascade through the system at the level of each functional component of the model rather than at the level of each individual failure mode.

For a system with a set of identified failure modes with associated probabilities and consequences, overall system performance can be estimated using a Monte Carlo simulation. Performance is simulated as a simple flow of dispensing over time, with random draws to determine whether different failure modes occur and, if so, when they happen. The effects of different failures are reflected in dispensing rates and, for each simulated response operation, the total number of doses that could be dispensed. By conducting many individual simulation runs of the same system, the probability of performance reaching specific numbers of doses dispensed (ie, the reliability of performance at those levels) can be estimated.

Because our work explored how these techniques could be applied generally, and not on a specific area's or region's preparedness efforts, we did not assign values to all of the elements in our failure trees. To demonstrate this process, we used a simple example of a dispensing system with a small set of failure modes with hypothetical probabilities and consequences for performance. Also, we implemented the simulation in Microsoft Excel[®], using its internal random number generation capabilities. Details of the simulation, the illustrative failure modes, and the resulting response reliability curves are described in the “Results” section.

Customized data collection and evaluation tools could be developed for conducting this type of analysis for an area's delivery and dispensing system. However, extensive efforts in evaluation and metrics development have already been devoted to SNS preparedness and the CRI program in particular. As a result, we examined how the results of existing evaluation tools could be used to provide the basis for reliability analysis. Doing so would take advantage of existing

data as a foundation for future analyses and limit additional evaluation burden where significant resources at multiple levels of government are already devoted to assessment. In examining how the results of existing evaluation tools could be used to inform an assessment of an SNS plan's response reliability, we considered (1) the technical assistance review process carried out cooperatively by the Centers for Disease Control and Prevention, states, and localities, and (2) drill-based exercises for SNS preparedness assessment. Our analysis included linking the substance of those tools to the content of the reliability assessment and failure trees and examining the (admittedly limited) publicly available information on the results of those assessments with respect to reliability lessons.

RESULTS
Reliability Analysis

Based on existing SNS deployment and dispensing policies, standards, and previously cited analyses, we constructed a detailed system model for SNS operations. Figure 1 shows the portion of that model that relates to POD operations (the full model is included in online supplementary materials). Examining each component of the model and systematically identifying the different failure modes that could affect operations, we constructed a set of failure trees mapping both

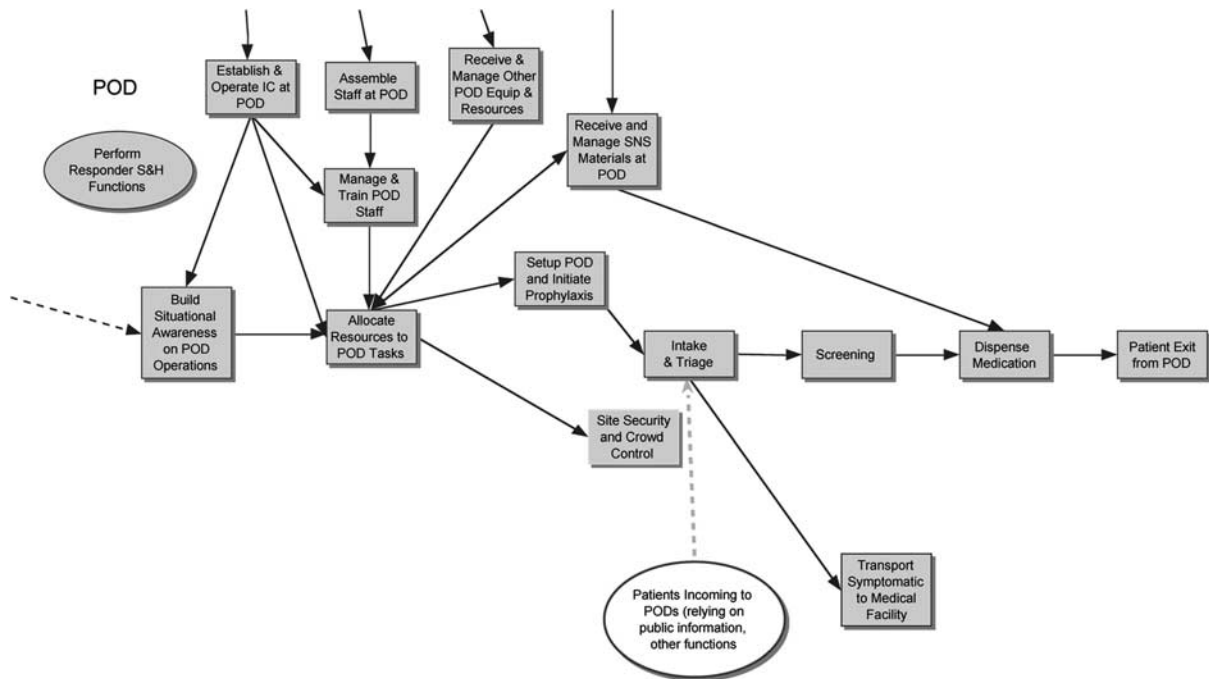
the faults that would affect individual parts of the model and the interactions among different functions, where failures in one part of the model would affect performance elsewhere.

Figure 2 shows 1 failure tree for POD operations (from a total set of 23 failure trees). Failures stemming from planning problems are grouped together on the left; personnel failures are in the middle; equipment-based failures appear on the right; and those in the "other" category appear at the bottom. The diagram demonstrates the interconnections that exist in response operations. The triangles with arrows show the links between this failure tree and others that indicate the causes of that failure in detail. These interconnections approximate the effect that a failure in one part of the system can have on the performance of other parts, but do not capture in full the individual failures affecting the probability or consequences of other individual failure modes.

Identifying failure modes is the foundation for assessing the reliability of a response system, but making the system-level assessment requires (1) assessing the likelihoods and consequences of those failure modes and (2) integrating those individual estimates into an overall reliability for the system for different levels of dispensing performance. Since our work focused on demonstrating this technique in general (vs examining a specific jurisdiction's preparedness), using

FIGURE 1

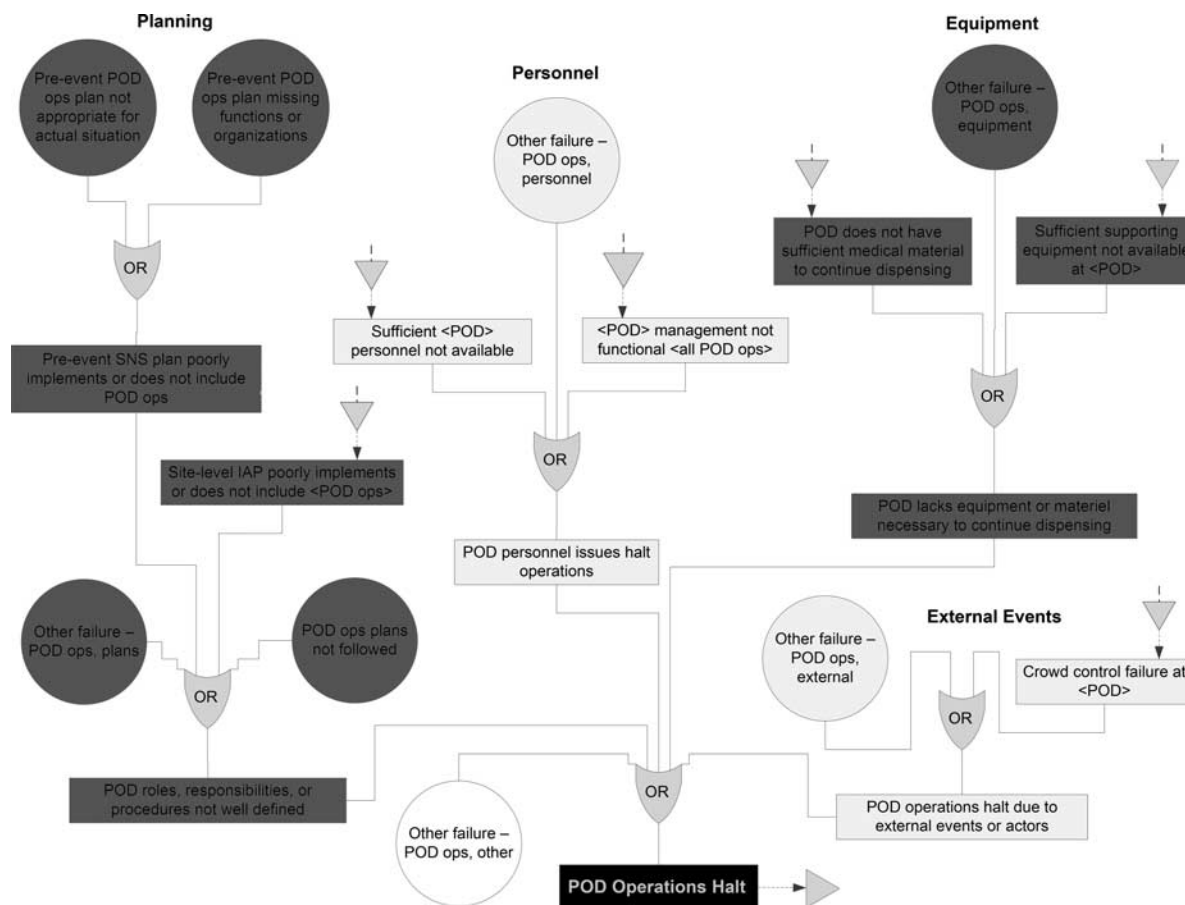
Conceptual Model of Point of Dispensing (POD) Operations for Reliability Analysis.



Abbreviations: S&H, safety and health; IC, incident command.

FIGURE 2

Failure Tree Showing Potential Faults Causing Halt of Operations at a Point of Dispensing (POD).



Abbreviation: ops, operations; IAP, incident action plan.

the full system model was neither practical nor meaningful. Therefore, we used a simpler system, with a small set of failure modes, to demonstrate how the overall system reliability values could be estimated.

In that example system, 20 PODs attempt to distribute antibiotics at a rate that should allow prophylactic treatment of approximately 100 000 people within 24 hours. Because people arrive at random intervals and take different amounts of time to pass through the POD queues, dispensing performance can be modeled as a Poisson process. In each 1-hour step, the simulation calculates how many people can be treated by the system. By repeating the simulation many times, we could estimate the probability of treating some number of people in 24 hours. For this analysis, we set the average arrival and service rate such that, with no failures, the system has a 95% chance of treating 95 000 or more people within 24 hours and a 50% chance of treating 100 000 or more.

To describe a response system's reliability characteristics, an exceedance curve is used. It shows the likelihood that the system will be able to successfully deliver treatment to increasing numbers of patients, starting at zero and ending at the maximum planned capacity of the system. For each number of patients (on the x-axis), the y value is the probability that the system will be able to successfully deliver *at least* that number of treatments. As a result, every response reliability curve will begin at 100% for zero patients, because any response system, no matter how challenged, will be able to deliver at least zero treatments. As patient numbers increase, the probability will decrease, as less extra capacity will be available to absorb the effects of any failures. Different types of failure modes affect the shape of a reliability curve differently. In general, faults that could stop response operations entirely push the curve downward from the top, while faults that cause smaller capacity reductions or delays push the curve inward from the right.⁷

FIGURE 3

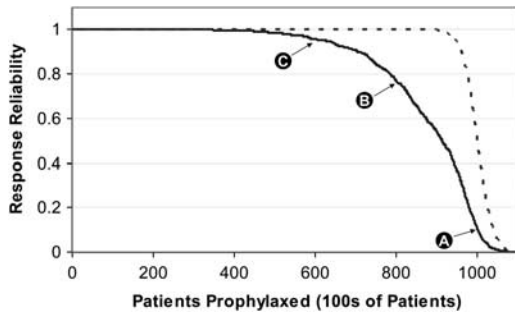
A Response Reliability Analysis for a Strategic National Stockpile (SNS) Example.

Figure 3 presents two response reliability curves for our example system, showing the likelihood of treating a given number of people in the 24-hour limit. The dotted line shows performance with no failures, demonstrating perfect reliability (100% likelihood of performance) until the system approaches its theoretical maximum capacity. Added to this baseline simulation are 4 different failure modes that could reduce dispensing performance:

- Initial delivery of the SNS materials to PODs takes longer than planned. This failure mode would occur at the beginning of the operation and reduce the time available for dispensing. The probability of this failure occurring was set at 10%, and its impact on reducing performance time was set at 25% (6 hours of idle time).
- Problems arising at the regional distribution site make the resupply of PODs less efficient. This failure could occur anytime during the response and would reduce system capacity from that point onward. We set this failure's probability at 30% and its consequences as a 20% reduction in treatment rate.
- Security breakdown shuts down 1 or more PODs. This failure could occur at any time and would reduce system capacity based on the number of PODs affected. Our example includes 2 versions: a 10% chance of 1 POD shutting down (5% capacity reduction) and a 0.5% chance of 2 PODs shutting down (10% capacity reduction). Both failures could theoretically occur in a single simulation run, producing a 0.05% chance of a 15% capacity reduction.
- Staff is not available as planned, reducing the efficiency of dispensing operations. This failure could occur at the beginning of the simulation, and was modeled at 3 levels: high chance (75%) that a few staff would be unavailable (5% capacity reduction), medium chance (35%) that some staff would be unavailable (20% capacity reduction), and small chance (5%) that many staff would be unavailable (40% capacity reduction). More than 1 personnel failure could occur in a single simulation run.

During each simulation, random draws are done to determine which failure modes occur and when (for those that could occur at any point in the response). When a failure occurs, it reduces system performance by reducing the system treatment rate. In this sample calculation, we treated each failure mode as independent of one another; although in an actual response one failure's occurrence could influence others. For example, supply failures (eg, failure mode 1) could affect the chance of security breakdowns occurring if the public responds negatively (mode 3). The failure probabilities were set high for illustrative purposes, but including such interactions would elevate those probabilities of the subset of related failures higher still.

In the simulations in which more than 1 failure occurred, performance "stepped down" with each additional failure. The total number of patients served was then calculated by summing all of the treatments dispensed over the full simulation, and a histogram of those results was used to calculate the probability that this system would be able to deliver at different levels of performance. The resulting response reliability curve (solid line in Figure 3) presents the probability that the system's performance will meet or exceed specific numbers of patients treated. It represents the net system performance for the 4 failure modes that could occur, and a total probability that it will be able to perform at or above each number of treated patients.

Although its performance matches the perfect reliability system for small incidents, the cumulative effect of the failure modes reduces its reliability significantly for larger scale dispensing operations. Three points (A, B, and C) illustrate examples of how reliability changes with incident size, decreasing the probability of delivering the identified levels of treatment from near 95% at point C to approximately 10% at point A.

Using Data From Existing Evaluation Tools

To analyze the reliability of an area's SNS and CRI planning, the types of system model and failure trees described are the starting point. To complete the analysis, estimates of the probability and consequences of individual failure modes need to be developed. There are varied ways this could be done. Expert or practitioner input during planning or evaluation processes is one strategy in which the models and failure trees are used to build consensus about the likelihood and consequences of different failure modes. The use of after-action reports from response operations or exercises also have been explored as possible sources of data.^{7,10} However, to take advantage of existing investments in SNS evaluation, we examined 2 evaluation tools, the technical assistance review (TAR) questionnaire and standardized functional drills to assess their potential contribution to reliability analysis.

TABLE

Percent of Basic Failure Modes in Domain Paired with at Least 1 Technical Assistance Review Question

Domain	%
Planning and organization	48
Equipment and technology	19
Personnel and training	12
Other	6
External environment or actors	0

The Technical Assistance Review

The TAR questionnaire addresses the full range of SNS preparedness issues. The approximately 90 questions are scored quantitatively, and some probe preparedness efforts at very detailed levels. We analyzed the content of the TAR using the individual failure modes developed in our analysis and found a significant fraction of some types of potential failure modes are addressed by the assessment. For example, almost half of the failure modes related to planning and organization identified by our analysis are covered by the TAR (Table). Some questions demonstrate that a potential failure mode has been eliminated for the system (eg, by demonstrating that a plan is in place), while others suggest that the likelihood of some failures has been reduced by the planning efforts. Other types of failure modes, including equipment and technology failures, personnel and training problems, and externally triggered response breakdowns are covered much less comprehensively in the TAR (between 0 and approximately 20%; Table). The results of our analysis demonstrated that the data collected in the TAR could be a strong foundation for reliability assessment, and improvements in the coverage and design of the questions could make the data collected more useful.

Assessing what previously collected TAR data can show about the reliability of deployment and dispensing would require examining scores for individual TAR components for specific jurisdictions to assess which failure modes included in the TAR had or had not been addressed in its plans. Because of the data's sensitivity, TAR scores are not publicly available at that level of resolution, although such an analysis could be done by jurisdictions using their own data.

Aggregated TAR data was made public in 2011. Average TAR scores were released for states and metropolitan statistical areas (MSAs), and component averages were made available for state-level TAR scores (ie, scores for TAR questions grouped by functional area, which are similar to but do not directly parallel our breakdown by failure tree). The aggregate numbers showed increases in average TAR scores at the state level from 87 to 94 (of 100) between 2007-2008 and 2008-2009.¹¹ Average scores for MSAs covered by the CRI also improved. More recently, additional data were published

on 2009-2010 TAR results for functional area averages and standard deviations around those means and the range of results for state and MSA levels. In some functional areas, those ranges were quite large (eg, the scores for security function at the MSA level ranged from 10-100 of a possible 100 points), suggesting significant underlying differences that would affect response reliability.^{3(pp9,11)}

Standardized Functional Drills

To supplement tools such as the TAR, a set of drills has been developed that focus on individual functions related to SNS deployment and dispensing, with a standardized, quantitative set of measurements to report performance in each exercise.⁵ The drills include staff and site availability, inventory management activities, site setup, and dispensing operations. In principle, measurements of system performance in exercises—to the extent that they reflect the system's totality, its possible failure modes, and the conditions it would face in an actual incident—could provide a *direct* approach for assessing system reliability.¹² The quantitative data captured in these drills, including time-to-completion measures for tasks and direct performance measures such as POD throughput, could be used to assess if extra capacity exists in parts of the system. The data could also help to infer the effects of failures observed in the drill or, as drills are repeated over time, variance in performance values for different parts of the system from both observed and unobserved problems. For this evaluation tool, system reliability and the interdependencies among the failure trees of different system components provide a blueprint for combining performance measures for individual portions of an SNS response into an overall system-level assessment.

As was the case with the TAR, only a limited amount data from past drills is available publicly. Data are available on the number of drills that have been carried out for different functions; numbers of participants in some types of drills; and averages, ranges, and standard deviations for some of the quantitative measures from the drills.^{3,13} For example, in inventory management, times involved in creating “pick lists” of supplies varied from less than 1 minute to more than 2 hours, suggesting that some systems would have significant reliability issues meeting the CRI's 24-hour performance target.

Available information also suggests that some drills may not have been sufficiently realistic to have tested the system to the point where reliability problems would become apparent. For example, many drills testing the ability to call personnel involved relatively few people, and many also notified participants that drills were going to occur.¹³ Similar limits on actual dispensing drills, some with very few patients, have made it difficult to extrapolate likely performance in a future large-scale dispensing operation.³ Cross-analysis of detailed TAR responses (or even aggregate TAR averages) with drill performance data could provide an approach to more closely link TAR responses, eg, limited insight into failure mode

probabilities and insight into performance. Future plans include more extensive full-scale exercises, with more realistic conditions to identify and assess failure modes, to provide enhanced insight into dispensing system reliability.^{12,13}

DISCUSSION

The goal of this work was to demonstrate the utility of applying new techniques for the analysis of system performance, specifically failure mode and effects analysis from engineering, to help construct integrated, system-level performance measures for public health preparedness. Our findings showed that it was possible to construct detailed systems models for SNS operations and corresponding failure trees, and lay the groundwork for applying the technique to specific preparedness systems. The system model and failure trees were constructed to support metrics development, but the diagrams could also be used in plan development and testing. Systematic examination of possible failure modes also could be applied internally to identify potential problems to support continuous improvement and ongoing preparedness activities.

Our findings also demonstrated how simulation can be used to convert quantitative estimates of failure mode probabilities and consequences to an overall system reliability estimate and its potential value for decision making. Reliability curves provide a snapshot of performance across different incident sizes, answering the fundamental question of preparedness measurement: the likelihood that response operations to a future incidents will go as planned.

Regarding the CRI goal, our findings also demonstrated how important it is where that point goal falls on a system's reliability curve, as compared to the theoretical limits of response and dispensing performance. Taking the analysis presented in Figure 3 as an example, the likelihood of a jurisdiction of successfully achieving that goal would differ considerably if its CRI target fell at Point C on the curve rather than Point A. By providing a way to integrate the effects on response performance of a variety of different failure modes with different probabilities and consequences, this type of analysis makes it possible to construct an intuitive, integrated measure of system-level performance.

Assuming failure modes have been systematically identified and their likelihoods, consequences, and potential interactions have been estimated to a realistic level, it should also be noted that the area under the reliability curve is itself a composite measure for preparedness across the full range of dispensing scenarios. Systems with many high probability failure modes will have reliability curves that are depressed downward or inward. The type and amount of the consequences of those failure modes will determine whether the reliability reductions occur predominantly for large and demanding incidents (to the right of the curves) or across all response scenarios. Comparing the area of a system's *theoretical* reliability curve with its *estimated* one provides

a way to measure how much better it could perform in an ideal world, and how specific investments in preparedness could be used as the basis to compare their cost effectiveness.⁷

Existing evaluation tools for SNS preparedness, specifically the TAR questionnaire and evaluation drills, could make a significant contribution to reliability analysis. However, basing assessment *only* on the results of those tools may not provide all the information needed. The fact that only composite scores have been publically released limit the ability of using TAR data to generalize what the recent increases in scores demonstrate with respect to system reliability. Similarly, the limited publically available data from operational drills and exercises make it difficult to draw detailed conclusions about the reliability of existing SNS plans and systems at the state level or below. However, individual jurisdictions that have access to disaggregated data from their own assessments and from drills and other internal planning and information would be better positioned to calculate reliability estimates for their SNS deployment and dispensing systems.

CONCLUSIONS

Recent efforts to evaluate public health preparedness have generated new insights into how communities have organized to respond to public health incidents. Substantial progress has been made, but an important conceptual element for preparedness assessment—the reliability of response systems performance—has not been included. Addressing this shortfall requires new perspectives on measurement that aggregate existing measures and information, perhaps with additional data sources, to provide insight into the likely future performance of a public health preparedness system. This report presents an approach for doing so at varying levels of detail, focusing on intensively evaluated area of SNS preparedness and using the CRI goal as an example.

Assessment of preparations for other types of incidents requiring responses beyond delivery and dispensing could also be the focus of reliability analysis, eg, the probability of providing mass care to different populations or delivering different levels of laboratory testing capacity to support epidemic assessment and response. Although integrating probability into preparedness assessment will require more data and insight than are available from existing evaluation tools and activities, we believe that our reliability analysis for SNS operations shows that the information they capture provides a foundation for examining this issue.

About the Authors

RAND Corporation, Arlington, Virginia

Address correspondence and reprint requests to Brian A. Jackson, RAND Corporation, 1200 S Hayes St, Arlington, VA 22202 (e-mail: bjackson@rand.org).

Supported by The US Department of Health and Human Services provided financial support. The funder had no role in the design or conduct of the study; collection, management, analysis, or interpretation of data; or preparation, review or approval of the manuscript.

Acknowledgments

A number of RAND colleagues including Jeffery Wasserman PhD, Christopher Nelson PhD, and Henry Willis PhD provided comments during the course of the research and critical review of the manuscript. We would also like to acknowledge the comments of an anonymous reviewer that were valuable in improving the paper. The results of RAND research do not necessarily represent the views or policies of RAND or any of its research sponsors.

Supplementary materials

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/dmp.2013.14>

Received for publication February 28, 2012; accepted August 7, 2012.

REFERENCES

1. US Centers for Disease Control and Prevention. *Cities Readiness Initiative*. Atlanta, GA: Centers for Disease Control and Prevention; May 20, 2010. <http://www.bt.cdc.gov/cril/>. Accessed August 16, 2010.
2. Willis HH, Nelson C, Shelton SR, et al. *Initial Evaluation of the Cities Readiness Initiative*. Santa Monica, CA: RAND Corp; 2009.
3. Nelson C, Parker AM, Shelton SR, et al. *Analysis of the Cities Readiness Initiative*. Santa Monica, CA: RAND Corp; 2012.
4. US Department of Health and Human Services. *Point of Dispensing (POD) Standards*; April 2008. <http://health.mo.gov/emergencies/sns/pdf/12-PODStandards.pdf>. Accessed January 31, 2013.
5. Nelson C, Chan EW, Fan C, et al. *New Tools for Assessing State and Local Capabilities for Countermeasure Delivery*. Santa Monica, CA: RAND Corp; 2009.
6. Loveys D. Medical countermeasure distribution and dispensing composite measure. Paper presented at the 2011 Public Health Emergency Preparedness Director's Conference; July 26-28, 2011; Atlanta, GA. <http://www.ndhealth.gov/epr/php/Resources/Med%20Countermeasure%20Slideshow-Composite%20Score.pdf>. Accessed January 27, 2011.
7. Jackson BA, Sullivan Faith K, Willis HH. *Evaluating the Reliability of Emergency Response Systems for Large-Scale Incident Operations*. Santa Monica, CA: RAND Corp; 2010.
8. Ebeling CE. *An Introduction to Reliability and Maintainability Engineering*. New York, NY: McGraw Hill; 1997.
9. Modarres M, Kaminskiy M, Krivtsov V. *Reliability Engineering and Risk Analysis: A Practical Guide*. New York, NY: Marcel Dekker; 1999.
10. Sullivan FK, Jackson BA, Willis HH. Text analysis of after action reports to support improved emergency response planning. *J Homeland Security Emergency Manage*. doi: 10.2202/1547-7355.1900.
11. US Centers for Disease Control and Prevention. *Public Health Preparedness: 2011 State-by-State Update on Laboratory Capabilities and Response Readiness Planning*; September 20, 2011. Atlanta, GA: US Centers for Disease Control and Prevention.
12. Jackson BA, McKay S. Preparedness exercises 2.0: alternative approaches to exercise design that could make them more useful for evaluating—and strengthening—preparedness. *HSAJ*. May 2011; 7: 1–19.
13. Jones JR. Strategic National Stockpile (SNS)-related medical countermeasures drills and exercises. Paper presented at the 2011 Public Health Emergency Preparedness Director's Conference; July 26-28, 2011; Atlanta, GA.