

Development of an Online Toolkit for Measuring Performance in Health Emergency Response Exercises

Foluso Agboola, MD, MPH, MPH;¹ Dorothy Bernard, MPH;¹ Elena Savoia, MD, MPH;^{1,2} Paul D. Biddinger, MD, FACEP^{1,3}

1. Division of Policy Translation and Leadership Development, Harvard T.H. Chan School of Public Health, Boston, Massachusetts USA
2. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts USA
3. Department of Emergency Medicine, Massachusetts General Hospital, Boston, Massachusetts USA

Correspondence:

Foluso Agboola, MD, MPH
Division of Policy Translation and Leadership Development
Harvard T.H. Chan School of Public Health
401 Park Drive
Boston, Massachusetts 02115 USA
E-mail: fagboola@hsph.harvard.edu

Conflicts of interest/funding: The authors acknowledge funding support from the US Centers for Disease Control and Prevention (CDC; Atlanta, Georgia USA) for supplement to grant number 5PO1TP000307-05 (Preparedness and Emergency Response Research Center – Linking Assessment to Measurement and Performance in Public Health Emergency Preparedness Exercises) and grant number CDC 1U90TP000417-05 (Preparedness and Emergency Response Learning Center). The content of this publication, as well as the views and discussions expressed in this report, are solely those of the authors and do not represent the views of any partner organizations, the CDC, or the US Department of Health and Human Services (Washington, DC USA), nor does mention of trade names, commercial practices, or organizations imply endorsement by the US Government. The authors declare no conflict of interest.

Keywords: disaster planning; emergency preparedness exercise; exercise evaluation; measures; performance improvement

Abstract

Introduction: Exercises that simulate emergency scenarios are accepted widely as an essential component of a robust Emergency Preparedness program. Unfortunately, the variability in the quality of the exercises conducted, and the lack of standardized processes to measure performance, has limited the value of exercises in measuring preparedness.

Methods: In order to help health organizations improve the quality and standardization of the performance data they collect during simulated emergencies, a model online exercise evaluation toolkit was developed using performance measures tested in over 60 Emergency Preparedness exercises. The exercise evaluation toolkit contains three major components: (1) a database of measures that can be used to assess performance during an emergency response exercise; (2) a standardized data collection tool (form); and (3) a program that populates the data collection tool with the measures that have been selected by the user from the database. The evaluation toolkit was pilot tested from January through September 2014 in collaboration with 14 partnering organizations representing 10 public health agencies and four health care agencies from eight states across the US. Exercise planners from the partnering organizations were asked to use the toolkit for their exercise evaluation process and were interviewed to provide feedback on the use of the toolkit, the generated evaluation tool, and the usefulness of the data being gathered for the development of the exercise after-action report.

Results: Ninety-three percent (93%) of exercise planners reported that they found the online database of performance measures appropriate for the creation of exercise evaluation forms, and they stated that they would use it again for future exercises. Seventy-two percent (72%) liked the exercise evaluation form that was generated from the toolkit, and 93% reported that the data collected by the use of the evaluation form were useful in gauging their organization's performance during the exercise. Seventy-nine percent (79%) of exercise planners preferred the evaluation form generated by the toolkit to other forms of evaluations.

Conclusion: Results of this project show that users found the newly developed toolkit to be user friendly and more relevant to measurement of specific public health and health care capabilities than other tools currently available. The developed toolkit may contribute to the further advancement of developing a valid approach to exercise performance measurement.

Agboola F, Bernard D, Savoia E, Biddinger PD. Development of an online toolkit for measuring performance in health emergency response exercises. *Prehosp Disaster Med.* 2015;30(5):503-508.

Abbreviations:

ASPR: Assistant Secretary for Preparedness and Response
CDC: Centers for Disease Control and Prevention
EEG: exercise evaluation guides
HPP: hospital preparedness capability
HSEEP: Homeland Security Exercise Evaluation Program

PHEP: public health preparedness capability

Received: May 5, 2015

Accepted: July 12, 2015

Online publication: September 15, 2015

doi:10.1017/S1049023X15005117

Introduction

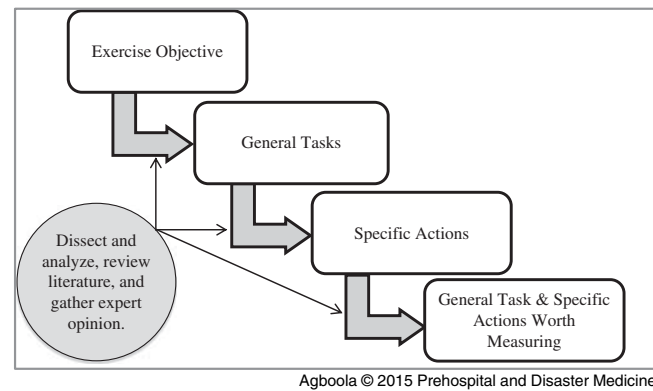
The routine conduct of exercises that simulate emergency scenarios is accepted widely as an essential component of a robust Emergency Preparedness program.¹⁻³ As proxies for actual emergencies, emergency response exercises can improve an organization's employees' fluency with existing plans and can provide opportunities to practice how different organizations and disciplines will work together to respond capably during emergency events.⁴⁻⁷ Also, and perhaps most importantly, exercises provide an opportunity to identify specific problems with an organization's emergency planning, training, and/or response that are in need of improvement before an actual disaster event occurs.¹

Unfortunately, the lack of commonly accepted, valid, and reliable measurement processes to use when quantifying the individual elements of performance in an exercise has limited many organizations' ability to feel certain that they have captured and documented the key successes and response challenges from their exercises accurately.⁸ This uncertainty, in turn, has limited these organizations' ability to gauge whether they are making improvements in their response over time, since response to different exercise scenarios is therefore difficult to compare. Moreover, the lack of standardized exercise evaluation methods makes it challenging for national and regulatory agencies responsible for overseeing public health Emergency Preparedness to aggregate and to compare results from local-level exercises as they attempt to identify what capabilities are in need of improvement.⁹ Therefore, the vast majority of data derived from the observation of public health emergency response exercises can neither provide reliable, objective support for most conclusions about the quality of the public health emergency response system, nor can they support systematic learning from one organization to another.¹⁰

To help health organizations improve the quality and standardization of the observational data they collect during simulated emergencies, a model exercise evaluation toolkit was developed.¹¹ The developed exercise evaluation toolkit contains three components: (1) a database of performance measures for emergency response exercises that have been vetted and tested in exercises; (2) a standardized data collection tool (form); and (3) a program that populates the data collection tool with the appropriate measures that have been selected by the user from the database for distinct exercises. Thus, the exercise evaluation toolkit allows anyone with a web browser to generate an exercise evaluation tool with performance measures that are relevant to a proposed exercise. The generated evaluation tool can then be used during exercises to gather data in a standard manner to support improved evaluation of performance. This report describes the process of developing and evaluating the usefulness of the toolkit in practice.

Methods

Over the past six years, a team of Emergency Preparedness experts with extensive backgrounds in design, implementation, and evaluation of exercises collaborated with several public health agencies, hospitals, and other health responders to evaluate a wide variety of capabilities, systems, and plans in more than 60 public health emergency response exercises. As part of this process, exercise evaluations plans and forms were designed, developed, and tested. This experience was the catalyst for creating the online toolkit, for which the methodology described below was used.



Agboola © 2015 Prehospital and Disaster Medicine

Figure 1. Approach to Exercise Evaluation.

Approach to Exercise Evaluation

Drawing on lessons learned from discussions with expert practitioners, from review of literature, and from review of available tools,¹² a process map was created to standardize the approach to exercise evaluation (Figure 1). Using the map, evaluation forms were created for each exercise. The evaluation forms were tailored to the specific exercises with performance measures aligned in a grid that allows documentation of performance by the use of the following four steps described below.

Step 1: Define Exercise Objectives—Exercise objectives serve as the starting point for creating exercise evaluations. For example, in a medical surge tabletop exercise involving acute care hospitals, one of the five defined objectives was to “evaluate the strategies available for hospital staff to optimize the caching and use of potentially scarce critical resources in accordance to existing plans.” Measures that are specific and measurable are then developed to help assess if the defined exercise objectives are met.

Step 2: Develop General Tasks to Measure Each Exercise Objective—General tasks are the broader categories of response elements that are expected to be completed by exercise participants to satisfy a specific exercise objective. For each objective, on average, five to seven general tasks are included. An example of a general task for the objective described in Step 1 is: “Track real-time resource availability and usage.”

Step 3: Identify and Integrate Specific Actions—Each of the general tasks is measured by a combination of multiple specific actions that can be observed clearly in an exercise. For example, for the general task above, the following specific actions are expected: (a) uses a system for prioritizing, allocating, and tracking real-time resource availability and usage; (b) shares data in real-time with other external response partners; and (c) analyzes and forecasts resource availability and usage.

The general tasks and specific actions are organized into sections in the form, as shown in Figure 2. As mentioned above, the exercise evaluation tool uses a combination of checklists and rating scales to produce quantifiable representations of tasks and actions to assess exercise performance. The performance of the entity being observed on a particular task is judged by (i) aggregate review of the number of specific actions observed to have been completed on the checklists (ie, Yes/No); and (ii) a rating of how

Demonstrate the ability to optimize the caching and use of potentially scarce critical resources in accordance to existing plans.

Response Element	Checklist of Actions (check all boxes and fill in spaces as appropriate)	Specific Data for exercise (Additional information specific to the exercise)	Overall performance (in each specific response element)	
Management of internal material resources needed for the response	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No	Uses a system for resource identification, typing and inventorying (Specify: _____) Has cache of critical supplies that were easily accessible Uses and/or redeploys on-site material inventories to respond to the incident Assesses need for additional inventories within critical timeframe Identifies ways to leverage/conserv e existing inventories Follows plan or procedures to acquire and order resources Anticipates material needs until the incident is concluded	How did they coordinate resource management within their facility? _____ How are material needs communicated within entity: _____ _____	1 2 3 4 5 6 7 8 9 10
Track real time resource availability and usage	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No	Uses a system for prioritizing, allocating, and tracking real time resource availability and usage Shares data in real time with other external response partners Uses standard messaging Makes data electronically available at a single point of access Analyzes and forecast resource availability and usage	List how resource availability and usage is prioritized, tracked and analyzed: _____ _____ What critical resource(s) is/are likely to run out? _____ _____	1 2 3 4 5 6 7 8 9 10
Request for external material resources	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No	Has clear triggers for when to request material resources from outside entity Identifies a person with authority to request external resources Uses mutual aid agreements or other means to request additional resources through Maintains an internal record of all requests for assistance	What is the role of DPH or emergency management in support of external material resource needs? _____ _____	1 2 3 4 5 6 7 8 9 10
Management of external resources needed for the response	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> No	Tracks and incorporates external resources into the response Identifies specific responsibilities, lines of authority, and means of coordination of resource management plan Prioritizes the use of scare resources Identifies who makes decision to allocate scarce resources Uses a database or other means to track requested and assigned external material resources	How would scarce critical resources be prioritized and allocated? _____ _____	1 2 3 4 5 6 7 8 9 10

Evaluation is designed to assess the emergency response capabilities required to respond to a Hospital Surge exercise
 Each response element receives a subjective overall performance score of 1 to 10
 1 = Unsatisfactory performance. 5 = Good performance. 10 = Excellent performance

Agboola © 2015 Prehospital and Disaster Medicine

Figure 2. Sample Exercise Evaluation Form.

well the general task was performed (in the opinion of the expert evaluator) on a Likert scale ranging from 1 to 10. In addition, each section of the exercise evaluation tool also includes open-ended questions allowing evaluators to add contextual elements to their assessment, describing root causes of an unexpected success or improvisation, a response failure, or recommendations for improvement.

Step 4: Testing Process—Leveraging on the opportunity to develop evaluation forms for multiple exercises, the evaluation forms and measures have been tested and refined in an iterative cycle over the past six years. The form and many of the measures that have been developed have been tested for their reliability, usability, and validity by independent evaluators during multiple exercises, and the results have been reported previously.¹³

Development of the Online Exercise Evaluation Toolkit

An online database of exercise evaluation measures was developed that allows anyone with a web browser to generate their own customized exercise evaluation form, similar to the one shown in Figure 2. The database includes the performance measures that were developed and tested over the past six years and currently contains about 160 general tasks and 500 related, specific actions. Every measure in the exercise evaluation database has been mapped to, and tagged with, at least one Assistant Secretary for Preparedness and Response (ASPR; US Department of Health and Human Services; Washington, DC USA) hospital preparedness capability (HPP) or one Centers for Disease Control

and Prevention (CDC; Atlanta, Georgia USA) public health preparedness capability (PHEP), as well as to a phase of incident management (Incident Recognition, Notification, Activation/Mobilization, Response, and Recovery). Therefore, users can use these tags to identify and select the most relevant performance measures based on the capabilities being tested during the exercise. Once a group of measures is identified and selected by the user, it can be formatted and downloaded by the toolkit software into an evaluation form, similar to the one shown in Figure 2.

Pilot Testing of the Online Exercise Evaluation Toolkit

The online evaluation toolkit was pilot tested from January through September 2014 in collaboration with 14 partnering organizations representing 10 public health agencies and four health care agencies from eight states across the US, including: Arizona, Connecticut, Iowa, Massachusetts, Rhode Island, Virginia, Vermont, and Washington. Exercise planners from all of the partnering organizations were asked to become familiar with the exercise evaluation toolkit, and to use it to assist with the evaluation of a health emergency response exercise in the study period.

Data Collection and Analysis

Structured interviews were conducted with all (14) exercise planners representing the 14 organizations participating in the project after they had become familiar with, and used, the exercise evaluation toolkit. Interviewees provided informed consent, and the information collected was not linked to the identity of the

interviewed individuals. Interviews were conducted with the use of a structured interview guide that was developed and recorded.

All interviews were transcribed verbatim and uploaded into NVivo qualitative data analysis software (QSR International Pty Ltd.; Doncaster, Victoria, Australia; Version 10; 2012). Two reviewers performed independent content analysis and coding of the interview transcripts. The two reviewers independently identified themes that emerged from the transcripts and agreement on such categories was achieved by discussion.

Results

Interviewees

Ten (71%) interviewees had prior training in the Homeland Security Exercise Evaluation Program (HSEEP; Washington, DC USA). Twelve (85%) had written an after-action report, 10 (71%) had experience in creating an exercise evaluation, and on average, interviewees had planned seven exercises over the past three years. Overall, the exercise planners used the developed exercise evaluation toolkit to evaluate a total of 23 exercises: one workshop, 10 tabletop exercises, eight functional, and four full-scale exercises.

Interview Findings

Interviewees' responses were aggregated to describe the: (1) feasibility of the online evaluation toolkit; (2) usefulness of the toolkit's generated evaluation form; and (3) comparison with other evaluation forms that may be used in exercises to evaluate performance. A detailed description of each theme and illustrative quotes from the interviewees are provided below.

Feasibility of the Online Evaluation Toolkit—The vast majority of interviewees (93%) reported that they found the online database of performance measures helpful in generating evaluation forms for their exercise and stated that they would use it again for future exercises. Eleven interviewees (79%) commented that the variety of databases of performance measures available was sufficient to build their evaluation forms. Alignment of the performance measures with the current federally defined health emergency response capabilities (the ASPR's HPP/the CDC's PHEP) was cited by many interviewees as a valuable feature of the toolkit, with an interviewee stating: "These metrics were already created for us and aligned to the capabilities that we are looking to test, so it really made the work easier." Other major benefits cited by interviewees included the ability to customize the evaluation form after it was generated by the website, the tool's scripted prompts that facilitated exercise participation, and applicability of performance metrics to organizational goals.

Approximately one-half of the interviewees reported they found the online toolkit useful for purposes other than creating an exercise evaluation form. For example, approximately one-half of the interviewees reported the toolkit helped with the overall process of exercise planning (beyond performance measurement) by presenting an exercise planner with a diverse range of pre-defined objectives with related measures. One interviewee said, "... it was helpful in pointing out what I needed to pay attention to," while another stated, "It gave me a lot of great ideas." Another unexpected finding was the suggestion by a few interviewees that the toolkit had the potential to assist with developing a customized evaluation form in real-time to evaluate their organization's response to a real-world incident.

Usefulness of the Generated Evaluation Tool (Form)—The majority of interviewees (71%) liked the exercise evaluation form generated from the toolkit, and 93% reported that the data collected with the evaluation tool were useful in gauging their organization's performance during exercise participation. When exercise planners were asked to elaborate on this point, many of them mentioned that the data gathered with the evaluation tool were helpful in identifying strengths and weaknesses and also helped generate benchmarks for future exercises: "...it gave us a good base of where we should be, and if we're not there, it gave us goals to reach for."

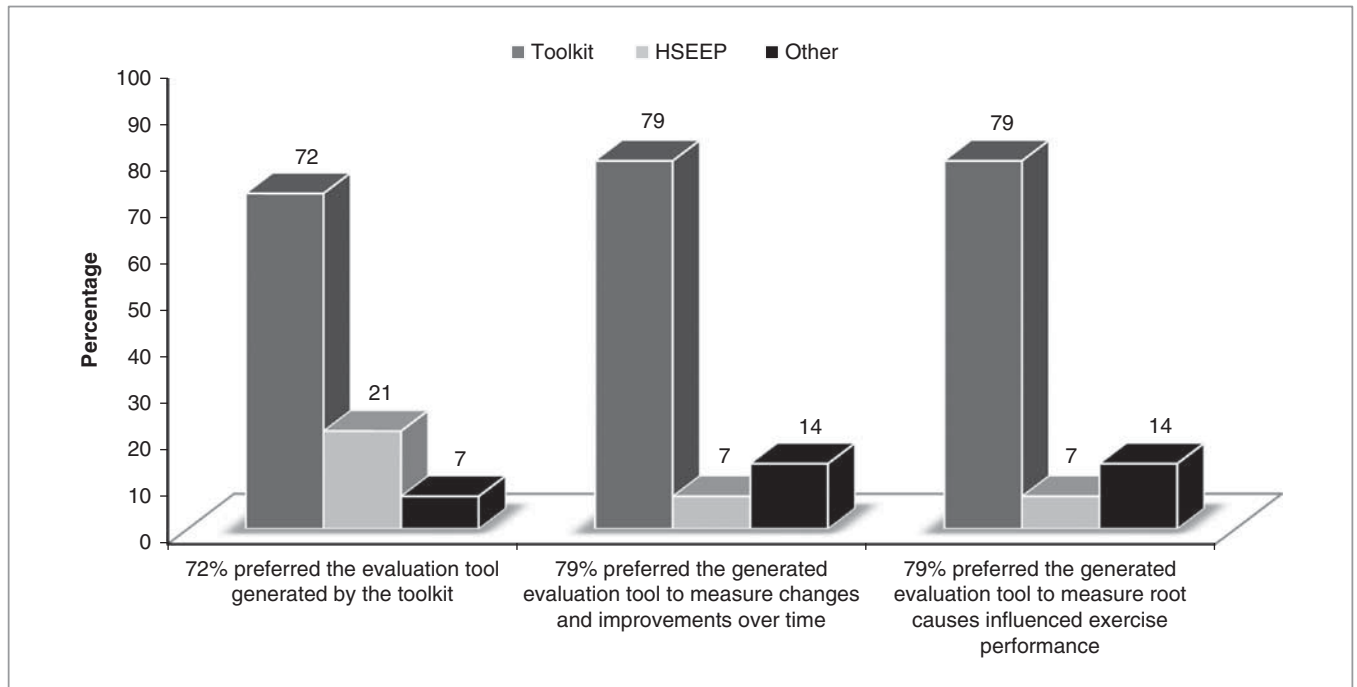
Four interviewees stated that their evaluators found the combination of checklists, open-ended questions, and rating scale in the evaluation tool to be useful to capture the key elements of the organization's performance during the exercise, "...we got so much positive feedback from the evaluators." Three interviewees reported their evaluators found the checklists served as an excellent guideline and reminder of critical performance measures, which they thought were helpful particularly for less-experienced evaluation personnel. While interviewees generally appreciated having open-ended questions to gather qualitative data, one interviewee cautioned that, "...you could get some pretty erroneous data with the open-ended questions if somebody doesn't really understand the exercise," suggesting that evaluators' experience may affect answers to open-ended response questions.

Seventy-two percent (72%) of interviewees described the toolkit as simple, easy to use, and/or intuitive. Interviewees also reported finding the data gathered by the use of the tool helpful in generating after-action reports, again, because of the alignment of the performance measures with target capabilities, "I think that's one of the biggest strengths of the tool."

Comparison with Other Methods of Evaluation—Most evaluators preferred the evaluation form generated from the toolkit to the HSEEP exercise evaluation guides (EEGs) or other evaluation methods (Figure 3). Reasons for this preference included ease of use and the relevance of the collected data to, and alignment with, public health and health care capabilities. One interviewee commented that, "... as compared to the HSEEP method, it's a lot more direct and applicable."

Discussion

While it is common for public health and health emergency managers to plan and conduct emergency response exercises to test their plans and systems, it is uncommon for them to collect standardized and structured data from those exercises to evaluate their response in a standardized manner so to compare their response with prior efforts, and to compare their response to the results of their peers.¹⁴ There are many reasons for this, and many barriers to effective exercise evaluation that have been described previously in the literature, including a lack of expertise, a lack of sufficient time, a lack of funding, and poorly defined exercise objectives.⁶ Therefore, the aim of this project was to develop an exercise evaluation toolkit that could assist exercise planners in their evaluation of Emergency Preparedness exercises. In an attempt to address the barrier of a lack of evaluation expertise, the toolkit was designed with a database of previously vetted and tested measures of performance, as well as an exercise evaluation tool that prompts the user to gather observations in a standardized manner and to grade the essential elements of performance for a given response objective. The database of performance measures was organized by



Agboola © 2015 Prehospital and Disaster Medicine

Figure 3. Comparison with Other Methods of Evaluations.

Abbreviation: HSEEP, Homeland Security Exercise Evaluation Program.

capability to make identification of relevant measures easier to the end user. The website has been structured to populate the evaluation tool automatically with performance measures selected to facilitate the exercise evaluation design process for the end user. The tool is free to users and does not have a financial barrier.

The results of this project show that users found the newly developed toolkit to be user friendly and more relevant to measurement of specific public health and health care capabilities than other tools and EEGs that are available. The toolkit has been revised to address the shortcomings cited by users, including redesigning the generated evaluation form and creating an easy deselect function that allows users to remove measures that are repeated and/or not needed. Furthermore, new features have been added to increase the practicality of the tool. Users can now create and save generated evaluation forms online, send generated evaluation forms to multiple evaluators via email, allow evaluators to upload and store their exercise evaluation data into an online database, and generate basic reports based on exercise evaluation data gathered during an exercise.

Limitations

This work presents similar limitations to any evaluation study that uses exercises to measure performance. Exercise evaluation outcomes generally are influenced by many factors, such as exercise design, exercise evaluators, exercise facilitators, and level of

participation. Therefore, the ability to test the reliability and validity of the developed performance measures is limited. It is important to be aware of such limitations when using the exercise evaluation toolkit. Progress in this research area, which encourages the use of a standard set of measures, will lead to a better understanding of how to assess a public health system performance prior to an emergency.

Conclusion

While future work remains to be done to examine the optimal evaluation of performance in emergency response exercises, the development and use of a standard set of exercise measures and a common data collection tool will help to improve observation, documentation, and aggregation of data about the capabilities of the nation's public health and health care Emergency Preparedness systems. The developed toolkit may contribute to the further advancement of developing valid exercise performance measurement tools and criteria.

Acknowledgments

The authors thank Max Su, Sean Lunde, Tara McCarthy, and Jessica Preston for their role in implementing this project, and all the interviewed exercise planners, and their agencies, for their insight and participation in this project.

References

1. Federal Emergency Management Agency. National Preparedness Cycle. FEMA Web site. <https://www.fema.gov/national-preparedness-cycle>. Accessed April 22, 2015.
2. World Health Organization. Hospital Safety Index: Guide for Evaluators. WHO; 2008. <http://www.paho.org/english/dd/pec/SafeHosEvaluatorGuideEng.pdf>. Accessed April 22, 2015.
3. Djalali A, Hosseinijab V, Hasani A, et al. A fundamental, national, disaster management plan: an education based model. *Prehosp Disaster Med.* 2009;24(6):565-569.
4. Dausey DJ, Moore M. Using exercises to improve public health preparedness in Asia, the Middle East, and Africa. *BMC Res Notes.* 2014;7:474.
5. Agboola F, McCarthy T, Biddinger PD. Impact of emergency preparedness exercises on performance. *J Public Health Manag Pract.* 2013;19(Suppl 5):S77-S83.
6. Biddinger PD, Savoia E, Massin-Short S, Preston J, Stoto MA. Public health emergency preparedness exercises: lessons learned. *Public Health Rep.* 2010;125(Suppl 5):100-106.
7. Fowkes V, Blossom HJ, Sandrock C, Mitchell B, Brandstein K. Exercises in emergency preparedness for health professionals in community clinics. *J Community Health.* 2010;35(5):512-518.

8. Kollek D, Cwinn AA. Hospital emergency readiness overview study. *Prehosp Disaster Med.* 2011;26(3):159-165.
9. Nelson C, Lurie N, Wasserman J. Assessing public health emergency preparedness: concepts, tools, and challenges. *Annu Rev Public Health.* 2007;28:1-18.
10. Savoia E, Preston J, Biddinger PD. A consensus process on the use of exercises and after action reports to assess and improve public health emergency preparedness and response. *Prehosp Disaster Med.* 2013;28(3):305-308.
11. Harvard School of Public Health Preparedness and Emergency Response Research Center. Emergency Preparedness Exercise Evaluation Tool. <http://phasevtechnologies.com/studies/lamps/index.php>. Accessed April 22, 2015.
12. Parker AM, Nelson C, Shelton SR, et al. *Measuring Crisis Decision Making for Public Health Emergencies*. Santa Monica, California USA: Rand Corporation; 2009. http://www.rand.org/pubs/technical_reports/TR712. Accessed April 22, 2015.
13. Agboola F, Dauksewicz B, McCarthy T, Biddinger PD. *Evaluation of Public Health Emergency Preparedness Exercises: New Research and Tools for Exercise Planners*. Orlando, Florida USA. 2013 Preparedness, Emergency Response, and Recovery Consortium and Exposition. <http://www.perrc.org/sites/default/files/PERRC%20Session%202009.pdf>. Accessed April 22, 2015.
14. Lazar EJ, Cagliuso NV Sr, Gebbie KM. Are we ready and how do we know? The urgent need for performance metrics in hospital emergency management. *Disaster Med Public Health Prep.* 2009;3(1):57-60.