International Journal of Disaster Risk Reduction 19 (2016) 413–446

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr

Scoping the field of disaster exercise evaluation - A literature overview and analysis

Ralf Josef Johanna Beerens[a,b,*], Henrik Tehler[b,c]

[a] Institute for Safety (IFV), PO Box 7010, 6801 HA Arnhem, The Netherlands
[b] Division of Risk Management and Societal Safety, Lund University, Lund, Sweden
[c] Lund University Centre for Risk Assessment and Management, LUCRAM, PO Box 118, SE-221 00 Lund, Sweden

ARTICLE INFO

Keywords:
Disaster
Crisis
Emergency
Exercise
Drill
Evaluation
Assessment
Analysis
Valuation
Review
Scoping study
Literature overview

ABSTRACT

The evaluation of emergency, disaster and crisis management exercises supports both individual and organisational learning, facilitates the development of response capabilities, and helps to determine whether the current level of preparedness is good enough. Nevertheless, despite its importance in the field of disaster risk management, there is a lack of a comprehensive overview of research in the area. The aim of the paper is to provide such an overview.

A scoping study identified the key contributions on the topic of disaster exercise evaluation, provides an overview of research in the area, and analyses opportunities for future work. The purpose, function and form of the evaluation provided the framework for the analysis, which was applied to the scoping study results.

The results indicate a lack of academic interest. Although exercises take place on a regular basis and are often used for research purposes, their evaluations are seldom the focus of attention per se. Moreover, contributions that do focus on evaluations are spread over several disciplines. Nevertheless, the results indicate that recent contributions are becoming more coherent as they build on each other (or at least refer to each other), even if they are produced within different disciplines.

Despite encouraging signs of a more cohesive scientific corpus on the evaluation of disaster exercises, there is still room for improvement. The scientific discourse would benefit from greater clarity regarding: (1) the purpose and context in which a specific evaluation method is designed to be used; (2) what the method needs to do (or produce) in order for it to fulfil the purpose; and (3) how the method achieves its goal and thereby fulfils its purpose. Moreover, in order to help researchers to build on each other's work and suggest improvements to evaluation methods, it is urgent that the supporting evidence (for example, empirical data or logical reasoning) for claims regarding the usefulness of a specific method is clearly presented. This is likely to lead to a more vigorous scientific discourse, which will result in increasingly relevant and robust arguments related to how to approach the problem of evaluating disaster exercises in practice.

1. Introduction

The data testifies to the high frequency of disasters, together with the associated human impacts. Disasters are a global issue, with substantial human and economic costs [1,2]. In preparing for such events societies take various measures to limit the potential consequences and cope more efficiently with the effects [3]. For example, they organize professional teams, such as firefighters, medical personnel, or other types of so-called first responders. Disaster management exercises, such as table-top or field exercises, play an important role in preparing these organisations to mitigate the effects of disasters. They can be used to test equipment and responses i.e. plans, procedures, skills and knowledge [4]. In other cases, they are used to facilitate learning by, for example, developing and embedding routine skills and procedures.

Evaluation is an important aspect of these exercises. Whatever the style or level of exercise, a well-constructed evaluation process is the key to providing evidence-based feedback on performance. It supports the direction of, and investment in, future learning and development. However, there is little research that provides a comprehensive overview of the evaluation of disaster management exercises. The aim of the present paper is to provide such an overview through the use of a scoping study approach.

The following section provides some general, theoretical background on evaluation. This information serves as a basis for the

* Corresponding author at: Institute for Safety (IFV), PO Box 7010, 6801 HA Arnhem, The Netherlands.
E-mail addresses: ralf.beerens@ifv.nl (R.J.J. Beerens), henrik.tehler@risk.lth.se (H. Tehler).

http://dx.doi.org/10.1016/j.ijdrr.2016.09.001
Received 9 May 2016; Received in revised form 1 September 2016; Accepted 1 September 2016
Available online 28 September 2016

framework that is used to analyse the results of the scoping study. The third section describes the methods used to structure data collection and the analysis of the literature. In the fourth section, we present the results and analysis. These results are discussed in Section five. The article ends with some conclusions regarding the current state of research in the area of the evaluation of disaster exercises.

2. Background

The field of evaluation research is rich in concepts, standards, principles, practical guidelines and approaches. One definition of "evaluation", provided in 1994 by the Joint Committee on Standards for Educational Evaluations [5] is, "*a systematic assessment of the worth or merit of an object*", which serves as a point of departure for other, more specific definitions. For example, the 'object' can be a specific type of disaster exercise. As it is unfeasible to describe all of the concepts, approaches and methods in the domain, some of the most relevant aspects are described briefly in Table 1. This table is used to categorise the methods and approaches for exercise evaluation found in the scoping study, and is the point of departure for the analysis of the study's results (Section 4).

3. Method

In order to provide a comprehensive overview of the scientific literature on the evaluation of emergency, disaster and crisis management exercises, evidence was collected using a scoping method. The study was conducted in July 2015 using the 'six step framework' [12,13] as a guideline. This section describes the framework and its application.

3.1. Scoping study

Grant and Booth [14] identified fourteen types of literature reviews and associated methodologies. They conclude that there is no internationally-agreed set of discrete, coherent and mutually-exclusive types. Instead each researcher must use the purpose of his/her study to judge whether a specific methodology is suitable or not. The aim of our review is not to develop or test theory [15], but to provide an overview of available research and suggest ways to improve it. Thus, the aim of scoping studies *"to map the literature on a particular topic or research area and to provide an opportunity to identify key concepts; gaps in the research; and types and sources of evidence to inform practice, policymaking, and research"* [16], fits our purpose. Specifically, we followed the 'six step framework' presented by Arksey and O'Malley [12]. The first four steps are described below, and final two in the results and discussion section.

A scoping study does not usually analyse the quality of research in a specific area. Nevertheless, as our ambition was to investigate opportunities for improvement, we also performed an in-depth assessment of the study's results. The point of departure for the analysis was the concepts and methods found in the field of evaluation research (see previous section).

3.1.1. Step 1: identifying the research question

The first step is to identify a broad and open research question. Here, the question is: *what is known in the scientific literature about the evaluation of disaster management exercises?* Before starting the systematic search (step 2 and onward) for material we conducted a set of quick-scan searches in academic journals in the field of disasters and emergencies. The aim was to develop an understanding of the material, where it might be found, and the terminology used to address the topic. This information provided input for the subsequent steps.

3.1.2. Step 2: identifying relevant studies/papers

The second step is to identify relevant studies. Arksey and O'Malley [12] recommend a broad search that is consistent with the overall research question. In order to systematically identify relevant work, a

Table 1
Relevant concepts, approaches and methods from the evaluation literature.

Aspect of evaluation	Methods, approaches or classifications	Brief explanation of categories (including references)
The way evaluation is performed	(1) Formal (2) Informal	*Formal evaluations* are systematic and rigorous, while *informal evaluations* are more ad-hoc [6]. Although this is an important distinction for research purposes, as informal evaluation is ongoing, the focus here is on formal evaluation.
Data Collection	Systematic and rigorous collection of information	It is important that findings are based on appropriate, credible and reliable information to be able to draw any conclusions about merit and worth [6]. The Joint Committee on Standards for Educational Evaluation [7] defined a range of standards. It is important that evaluators make selective use of both qualitative and quantitative data collection tools and strategies. It is therefore interesting to identify what information is considered relevant for a specific method, in what context and how its reliability and validity are guaranteed (e.g. what data collection methods are used and how). The use of multiple information sources and techniques, and how they are applied in the evaluation of disaster exercises is particularly important.
The functioning of the evaluation	(1) Measuring (2) Describing (3) Judging (4) Stakeholder centred	Guba and Lincoln [8] use the term 'generations' to denote the focus of a specific type of evaluation. The first generation of methods focus on *measuring*, the second on *describing*, the third on *judging*, and the fourth on (constructivist) *stakeholder involvement*. Closely related to the fourth generation is responsive or stakeholder-centred evaluation [9]
Why the evaluation is performed	(1) Improvement (2) Accountability (3) Dissemination (4) Enlightenment	Stufflebeam and Coryn [6] identify four main uses of evaluations: improvement, accountability, dissemination and enlightenment. This is also partially supported by Hertting and Vedung [10]
When the evaluation is performed	(1) Formative (2) Summative	*Formative* evaluations are used to provide information for developing a service, ensuring its quality, or improving it. They are prospective, proactive and undertaken during the development or operation of a program. *Summative* evaluations are retrospective assessments of completed projects, established programs, finished products, or services rendered. They provide an overall judgement of the effectiveness of the individual, team, organisation, or policy with respect to the expected outcomes [11].
Evaluating the evaluation	Meta-evaluation	it is important to ensure the overall integrity and credibility of various, single case, evaluations. This can be achieved by applying a meta-evaluation, which is defined as "the process of evaluating an evaluation" [6]. It involves isolating, obtaining, and applying descriptive and judgemental information about an evaluation's utility, feasibility, propriety, accuracy, and accountability for the purposes of guiding it and reporting its strengths and weaknesses [6]. Meta-evaluations can serve various purposes, but are often needed to assess and identify the need to adjust or amend the system.

distinction was made between: 'database selection' (i.e. where to search); and 'search query identification' (i.e. how to search).

3.1.2.1. Database selection. The Scopus, Academic Search and Web of Science databases, owned by Elsevier, EBSCO and Thomson Reuters were selected, as they are multi-disciplinary, cover a wide range of research fields, and are owned by various publishers. A fourth, PubMed, owned by the United States National Centre for Biotechnology Information, was added because the quick-scan results showed that some of the relevant literature is in the health and medical sciences field. Because the focus was limited to peer-reviewed scientific papers, no grey literature was searched.

3.1.2.2. Search query identification. The search string was based on a Boolean approach. The three distinctive keywords in the research question are: (a) disaster, (b) exercise and (c) evaluation. However, these three keywords alone were insufficient, as these words have many synonyms. Therefore, a list of synonyms was compiled by searching thesauruses [17], combined with the function provided by various databases (for example EBSCO) to identify related terms. The results of this search are listed in Appendix A.

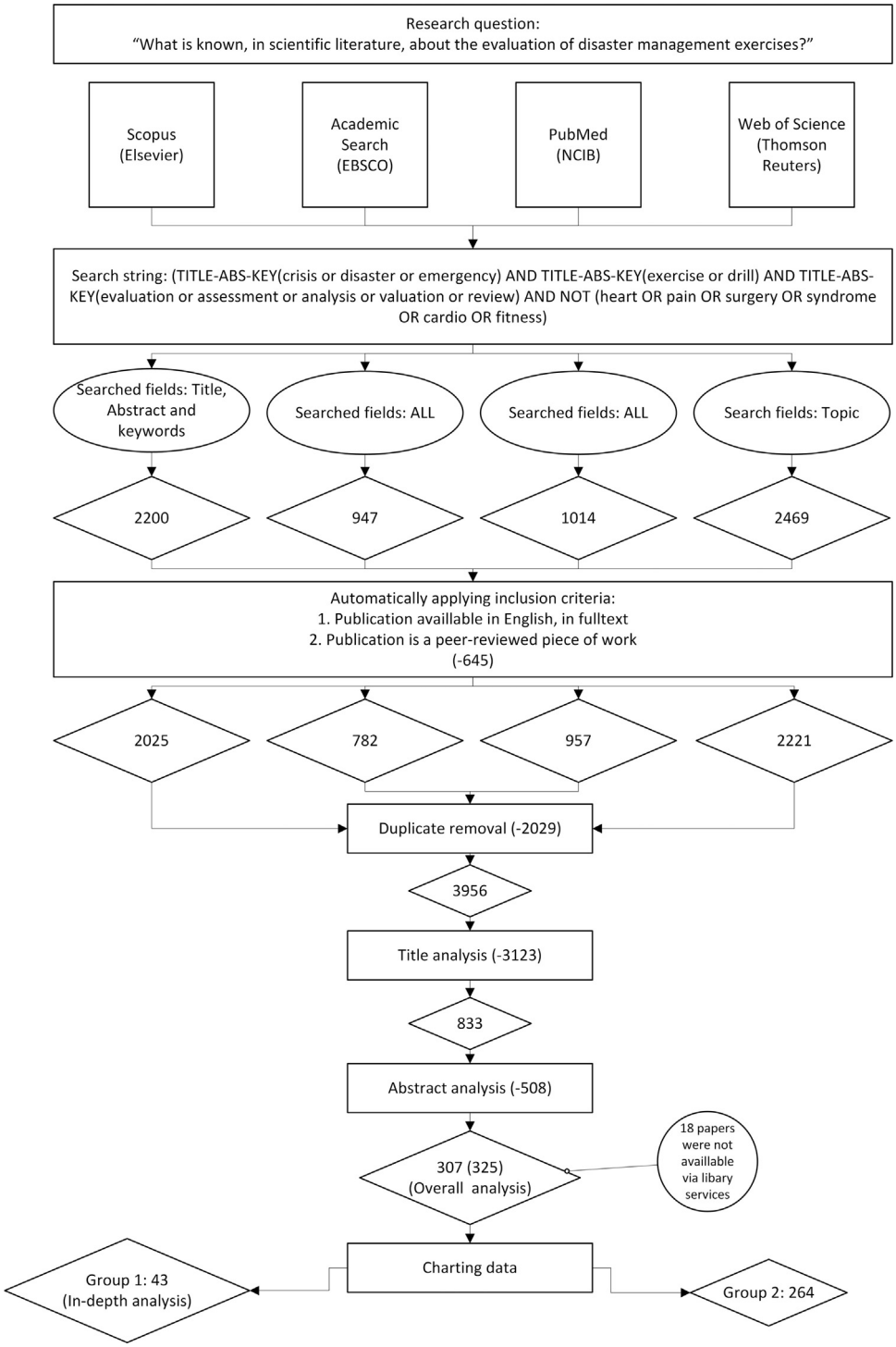

Fig. 1. Overview of search method (search date: 22 July 2015).

Nine synonyms were found for the word “disaster” ten for “exercise” and ten for “evaluation”. These terms were systematically combined, resulting in multiple search queries each consisting of three words (e.g. “disaster AND drill AND exercise”). Each combination was used to search the Scopus database (title+abstract+keywords) and the number of hits was noted for each query. The titles of the first 100 hits (most recent publications) were analysed, and synonyms that appeared to generate irrelevant hits were removed from the list. For example, if including a specific synonym (e.g. “training”), generated irrelevant hits, it was removed from the list. This reduced the number of synonyms to the following:

Variable A: Disaster
Synonyms: Crisis, Emergency
Variable B: Exercise
Synonym: Drill
Variable C: Evaluation
Synonyms: Assessment, Analysis, Valuation, Review

In addition to keywords, this analysis also identified various 'excluding keywords', i.e. words that produced a large number of irrelevant results. For example, the synonym ‘emergency’ identified studies in the medical domain related to cardiology research into fitness. Therefore, terms such as ‘heart’, ‘cardio’, ‘pain’, ‘surgery’, ‘syndrome’ and ‘fitness’ were introduced as excluding keywords.

Fig. 1 shows the search string and the initial number of hits (6630) from the various databases. The application of inclusion criteria 1 and 2 (see Fig. 1) reduced this to 5985 papers, and automated duplicate removal left a total of 3956 papers. These were analysed in detail in the next step.

3.1.3. Step 3: study selection

Many of the final sample of 3956 papers did not address the evaluation of disaster exercises and had to be removed by assessing the title against additional inclusion and exclusion criteria. These criteria are presented in Appendix B. Only papers that were clearly irrelevant were removed and borderline cases were retained for further analysis. This led to the removal of 3123 papers. The abstracts of each of the remaining 833 papers were then read and assessed against the inclusion and exclusion criteria. Once again, borderline cases were retained and the whole article was analysed. Around 508 papers were removed at this point. The list was manually scanned in order to verify whether there were any duplicates, which were removed. This left 325 papers for further analysis. The abstracts of 18 papers were not available[1] leaving 307.

3.1.4. Step 4: charting the data

The analysis was performed in two steps. An initial, overall analysis focuses on broad trends and characteristics. This corresponds to the fourth step in Arksey & O’Malley's framework. The second step involved a more detailed analysis that focused on the role of evaluation. This analysis used a template based on the key aspects found in the evaluation literature (see Section 2). We call this the 'in-depth analysis'.

3.1.4.1. Overall analysis. Each of the 307 abstracts was read and analysed in order to determine whether the topic was relevant. If this was inconclusive, the whole article was read (n=147 papers). The aim was to identify papers that explicitly addressed disaster exercise evaluation per se. We therefore grouped papers based on their overall focus; either (1) the exercise and its evaluation method, or (2) other aspects of disaster exercises. Group 1 contained 43 papers and Group 2 consisted of 264 articles. As the papers in Group 2 contained interesting information regarding the context in which exercises are performed and their evaluation, we included them in the overall analysis, but excluded them from the in-depth analysis.

3.1.4.2. In-depth analysis. The 43 papers in Group 1 were analysed based on the framework described in Section 2 (Appendix C). Furthermore, we performed a network citation analysis using a MATLAB algorithm and VosViewer software [18,19]. The aim was to determine the most influential studies (in terms of citations) and identify clusters of researchers who cited each other's work. Visualizing bibliometric networks (also referred to as ‘science mapping’) is a comprehensive approach to coding a specific research stream [20]. Complex bibliographic data can be analysed relatively easily by visualizing the core aspects [19].

The data were analysed in MATLAB following an approach that identifies group structures in networks [21]. The results were verified using VosViewer software. According to Eck and Waltman [19] bibliographic network visualizations are most useful when they are carefully interpreted and used in combination with an expert evaluation. Therefore, we also performed a manual analysis that aimed to check the data from the citation analysis and extract information related to the purpose, function and form of the evaluation.

Arksey and O’Malley's scoping study framework includes a sixth step. In this step, outcomes are discussed in a consultation exercise designed to validate the method, suggest additional references and provide new insights. This step was performed with an expert group[2] in the Netherlands. Their feedback has been incorporated into the discussion section.

4. Results and analysis

4.1. Overall analysis of all papers (groups 1 and 2)

Papers were categorised into general subject areas based on their title and discipline (e.g. computer science, engineering and medicine[3]). Table 2 presents an overview of this classification. Most papers were published in the area of medicine, followed by the social sciences. Most were journal articles, followed by conference proceedings (Table 3).

Fig. 2 shows the number of papers published annually, which suggests that there is a growing interest in research that focuses on disaster exercises and their evaluation.

The categorization provided a structured overview. Most papers fell into the category of ‘evaluation research’ [22] and used standard social science data collection methods including self-evaluation surveys [23–31], questionnaires [24,32–35], interviews [36–38] or literature reviews [39–42]. In a few cases, Delphi techniques were used to determine the importance of evaluation criteria [43], or the quality of a tool [44]. Several papers that discussed evaluation in an exercise context employed trained evaluators to collect data through observations and registration [31,43,45–56]. In a few cases these evaluators were experienced staff members, but in most cases they were third-party or external experts [57–62]. A few papers collected data using specific tools, such as video or software [63–66]. In a few cases (notably medicine and health), data

[1] These 18 papers could not be retrieved through interlibrary loan services (most were published in conference proceedings or books).

[2] The expert group consisted of seven persons from the Netherlands with a background in research/academia in the social sciences (public administration, organisational science) and medicine. Participants were familiar the key topics of this research (i.e. disaster management and/or evaluation). Five had carried out incident and/or exercise evaluations in the Netherlands. Two had carried out other types of evaluation (i.e. policy evaluations or evaluative research). All participants had a relevant academic background.

[3] As the papers were drawn from various databases they were manually categorized by one of the authors using automated Scopus subject areas as a starting point. The title (of the journal, proceedings, etc.) was read and used to categorize the article into a subject area.

Table 2
Documents by subject area.

Documents by subject area	Total (n=307)		Group 1 (n=43)	
Computer science	32	10.4%	1	2.3%
Earth and planetary sciences	4	1.3%	0	0.0%
Energy	1	0.3%	0	0.0%
Engineering	48	15.6%	4	9.3%
Environmental	25	8.1%	5	11.6%
Health	40	13.0%	6	14.0%
Medicine	87	28.3%	20	46.5%
Nursing	5	1.6%	0	0.0%
Social sciences	65	21.2%	7	16.3%
Total	**307**	**100%**	**43**	**100%**

Table 3
Publication type overview.

Publication type	Total (n=307)		Group 1 (n=43)	
Book	16	5.2%	2	4.7%
Book chapter	2	0.7%	1	2.3%
Journal article	230	74.9%	36	83.7%
Conference paper	59	19.2%	4	9.3%
Total	**307**	**100%**	**43**	**100%**

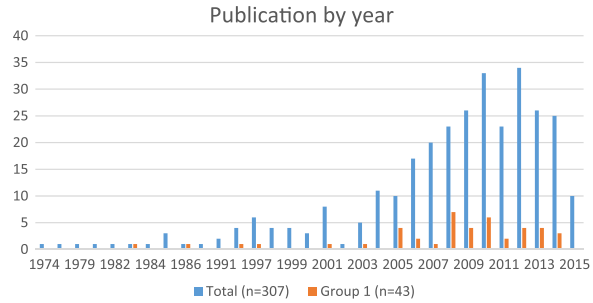

Fig. 2. Number of publications by year.

is mainly based on statistics and quantifiable indicators such as time or number of victims [46,55,67–71].

There was similar variety in terms of focus. Some studies looked at certain types of equipment or tools [72–78] (e.g. robots, sensors, victim cards), more comprehensive systems [79–86] (e.g. decision support, simulation, information or communication systems), or general skills [87–95] (e.g. information management, decision-making, improvisation or collaboration). Another distinction is research that uses evaluation related to a specific process, such as triage [96–98], decontamination [69,99,100], points of dispensing [46,101] and evacuation [102,103]. In some cases, models are tested for assessment [104], decision- and sense-making [105–107] types of management such as emergency management, command and control or resource management and awareness [108–112]. There is also research that focuses on artificial environments i.e. virtual environments [113–115] or behavioural aspects such as learning [45,60,116] and risk perception [117].

Most papers were descriptive, i.e. they described the design of an exercise, the scenario that was employed, the lessons that were learned and the method that was used. However, a few prescriptive papers described how an exercise should be set up [118].

Papers were also classified in terms of how they described and identified lessons to be learned. In particular, Borodzicz & van Haperen [40] discuss processes of individual and group learning in crisis simulations. Lessons are mainly derived from observations or collected after debriefings [119–121]. These papers rarely provide details of what information was collected or how it was obtained, simply making use of the judgement of observers or participants [122,123]. In various cases, lessons learned were combined with a description of the exercise and scenario [54,124–127]. A few papers addressed learning processes [128,129]. In most cases, conclusions were based on a pre- and post-exercise survey that assessed participants' impressions and viewpoints.

Several articles addressed the design of exercises [64,130–133]. In these papers, evaluation is often seen as an element or step in the process; rather than provide details of how it should be done, they focus on its purpose. Scenario development and description often play an important role. Typically, these papers discuss various types of exercises, such as table-top and field exercises, and their execution. A specific group of papers relates to the United States 'Homeland Security Exercise and Evaluation Program (HSEEP) Guidelines' [29,41,42,48,54,118,134,135]. This policy document is often used as a reference for setting up an exercise and its evaluation. Although it is mainly cited in North American papers, there is one example from another country [132].

A final category was papers that focused on the testing of a plan or procedure [68]. Most of these addressed preparedness [43,49,136,137]. Exercises are used in the context of preparedness programs and in this case, it is not the exercise itself that is evaluated. Instead, it is combined with other aspects such as training and the formulation of guidelines.

Papers that focused on longer-term responses, such as developmental or educational disaster preparedness programs, were excluded from the search (see Section 3), as were papers that collected data over a longer period of time from several exercises for benchmarking purposes [138].

4.2. In-depth analysis

Following the overall analysis, we conducted an in-depth analysis of the 43 papers in Group 1, which focused on the evaluation of disaster exercises. Like Group 2, most of this research is published in medical and social sciences journals (see Tables 2, 3), mainly published in North America and Europe. In the section below, we present the results of the citation analysis and the application of the framework (Section 2).

4.2.1. Citation analysis

The citation analysis was performed using MATLAB and data was verified using VosViewer software. This showed that 28 out of the 43 papers referred to others in the group (see Fig. 3). Analysis of the nodes (papers) helped to identify the most influential work in terms of citations. Although citations can be seen as an indication of the importance of a work, it should be noted that recently-published papers are less likely to be cited than older papers. The analysis found that the paper by Biddinger et al. [41] was referred to most often by the other authors in the group (eight citations). It was closely followed by Savoia et al. [27,29] (both five citations), [139] (cited eight other papers in the group), while Agboola et al. [61] and Ingrassia et al. [140] were all citing seven other papers.

A MATLAB analysis identified three clusters (Fig. 3). Within these clusters, nodes (papers) are more highly connected to each other than to the rest of the network (i.e. papers within a cluster refer more extensively to each other than to other papers). A fourth group was identified based on a content analysis and a visual inspection of the network. The fact that two authors have written papers that are connected through citations does not necessarily mean that they collaborated or built upon each other's work. Thus, belonging to the same cluster does not imply actual collaboration as it may merely be a consequence of the way in which the authors refer to each other's work.

The first cluster is most active, both in terms of the number of papers published and references to other papers in the sample. It contains work by researchers including Biddinger, Savoia and Agboola [27,29,30,41,61,135,139,141,142]. This main focus of this cluster is developing a framework for performance evaluation in the area of public health. The research focuses on all aspects of evaluation—from measurement criteria and tool development, to post-action review analysis and lessons learned.

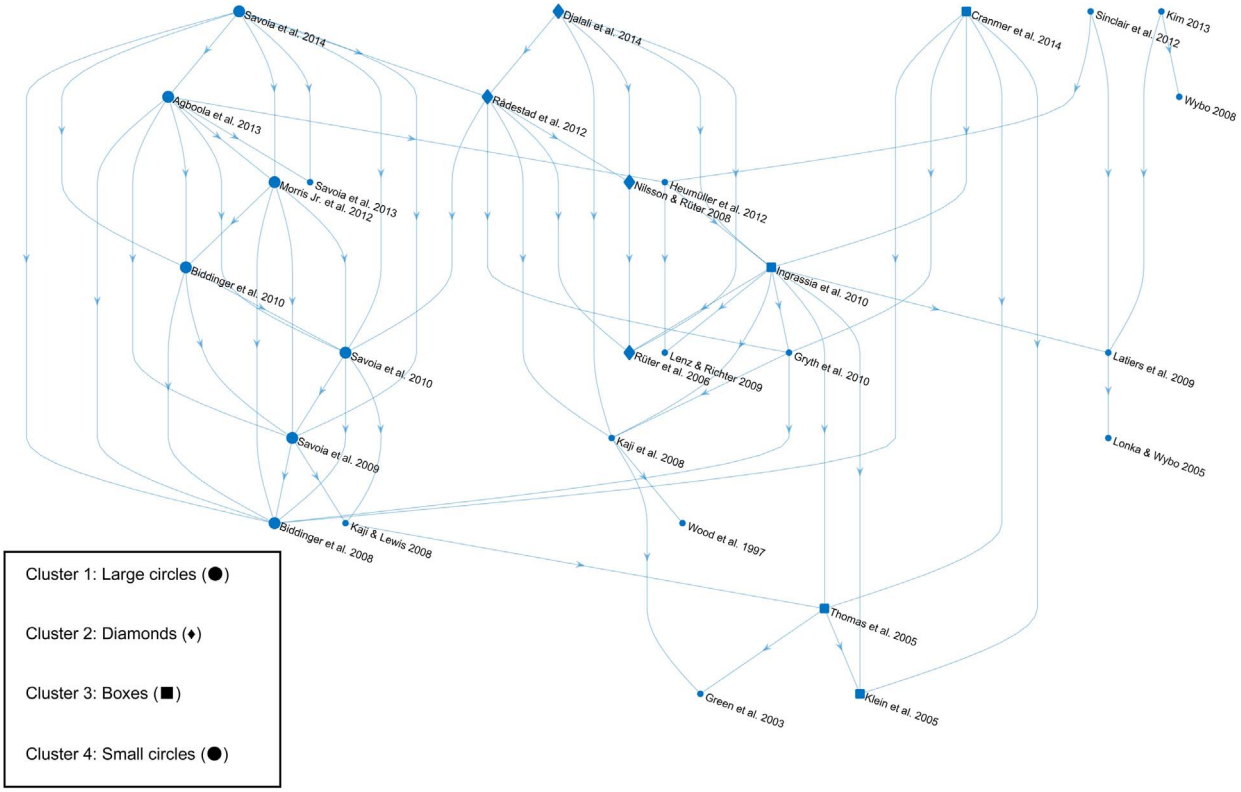

Fig. 3. MATLAB Citation Analysis Network overview.

The second cluster includes work by researchers such as Rådestad and Rüter [55,70,143,144]. This group focuses on developing and implementing performance indicators for the evaluation of medical responses and, in particular, hospital preparedness. The smaller, third cluster [38,56,140,145] contains papers that address the evaluation of medical management during mass casualty incidents. These articles focus on data collection by developing tools for observation, or training observers in medical management evaluation.

Finally the fourth cluster includes researchers such as Sinclair, Latiers and Wybo [146–150] who approach the subject from a broader, methodological perspective. They present overviews based on other literature and research, for example accident investigation, in order to understand and assess disaster exercises in general and thus be able to evaluate them.

In addition to the automated citation analysis of Group 1 papers, we performed a manual snowballing analysis in order to check whether the references in the papers in Group 1 contained any other relevant information. No other references were identified that met the inclusion and exclusion criteria (Appendix B (Table B1)).

4.2.2. Form of the evaluation (see appendix C, column A)

Forty-two papers took a formal approach. There was one exception [141], where it was not possible to determine the form of the evaluation as the paper focused on how post-action reviews should be structured, completed, and aggregated in order to improve quality. This paper discussed evaluation in general, rather than a particular method, unlike the other papers in the group.

*4.2.3. Method (*see appendix C, column B*)*

In 15 papers, observation was the only data collection method. In another nine papers, observations were combined with questionnaires. In five, they were combined with debriefings, and in three others with interviews. Only one paper [27] used questionnaires as the primary data collection method. In this paper, the questionnaire was validated as a self-assessment performance measurement tool, and was filled out by both participants and external experts. In all other cases a combination of methods was used. Nineteen papers took a qualitative approach to information collection. In 15 papers, there was a combination of quantitative and qualitative methods and four employed quantitative methods using scored indicators and measurements of time in order to evaluate medical performance [55,70,140,143]. In four papers it was not possible to determine the data collection method.

Twenty-eight papers described a method and identified the information to be collected. These papers presented general methods or process descriptions (such as the official HSEEP guidelines). The earliest paper (Crankshaw and Fenton [151]) introduced general evaluation criteria to be used for measuring performance. They also proposed a weighted method for calculating and comparing the results. Later papers developed a particular evaluation method. For example Lonka and Wybo [150] and Wybo [152] introduce the 'REXAO' method, which aims to develop organisational learning from accidents and crises, while Cranmer [38] use a competency-based tool. Kaji et al. [24] and Kaji and Lewis [142] introduce the module-based AHRQ tool, which can be used to evaluate processes during a hospital disaster drill. Duarte et al. [153] describe the 'ASC Process Model', which describes the evaluation process, while Savoia et al. [139] present a framework for performance measurement. Finally, Heumüller et al. [154] propose an action-based research framework for exercise development and evaluation.

The reason for using a particular method varied—from using a standardized method to be able to compare and share data—to more specific, contextual justifications (see Appendix C, column C). Examples apply to, for example, oil spill processes [155], hospital management [24,29,55] and exercises at schools [134]. In most cases, the chosen method aimed to improve validity and reliability. In some cases, methods were related to the clusters; for example [55,112] use a similar method to measure performance in hospital management. However, it was not always clear whether the method could be applied outside the specific context.

In six papers no specific method was implemented or developed. Steward and Wan [156] linked exercise evaluation to simulation

modelling. However, it was not possible to determine which method for evaluation that was employed. A further eight papers could not be classified as the information was lacking. This may be related to the nature of the work, which did not aim to use or develop a particular method or framework, instead it might be focused on, for example, identify criteria [157] or provide lessons learned [158].

4.2.4. *Context (*see appendix C, column D*)*

Twenty-one papers used a technological scenario (i.e. a human-made disaster), seven a natural disaster scenario, and one paper used both types. The remainder either did not explicitly mention a scenario, or a scenario was not applicable. Most papers discussed a mass casualty incident (MCI). In particular, Johnston et al. [159] used a school environment and Green et al. [160] described exercises in developing countries. Twenty-four papers focused on the local, regional or state level. In two cases these levels were combined with the national level, and in three cases the scenario is at a national level. In the remainder, it was not possible to determine the level, or no information about the scale of the exercise was provided. More than 50% of papers did not refer to previous exercises or incidents, implying that they discuss individual cases.

4.2.5. *Function (*see appendix C, column E*)*

In 17 papers, the function of the evaluation was explicitly related to measurement. Of these, most used evaluation to measure performance. Three papers focused on an evaluation method aimed at describing, and two on the process of judging. However, there is a thin line between measurement and judging, because in some cases measurement is a form of judging. In the remaining papers it was not possible to determine the function of the evaluation. In eight papers, stakeholders were involved in the setup and preparation of the evaluation, while in the remainder this was either not the case or it was not possible to know.

4.2.6. *Purpose (*see appendix C, column F*)*

Nine papers make an explicit link with learning/development (also referred to as 'evaluand'). In six, the evaluation is used for accountability purposes, while in seven others it is a combination of both learning and accountability. Seven had a specific research purpose (e.g. they validated an evaluation method). Nine papers took a holistic approach, focused on the systems' level. Eight papers evaluated parts of a crisis response system, while the focus could not be determined in the remaining cases.

4.2.7. *Timing (*see appendix C, column G*)*

In most cases it was difficult to determine the timing of the evaluation. In seven papers the focus was on summative evaluations, in three it was on formative evaluations and two papers used both types of evaluations.

5. Discussion

5.1. *Methodological concerns in evaluation design*

Methodological concerns such as threats to reliability and validity are a traditional aspect of research. Many of the reviewed papers contain general discussions of both reliability and validity with respect to a particular data collection method, e.g. semi-structured interviews. Although such discussions are valuable and necessary, here we are more interested in investigating whether the papers present motivations for why the *suggested evaluation method* is suitable in the context in which it is supposed to be used. Such discussions, however, are far more uncommon. This is surprising, as when suggesting a new method (or combination of methods), or the development of an existing one, the key point is *value or utility*, i.e. does the method work (or work better than existing methods)?

One possible explanation for this might be that it is a perspective that is not traditionally found in the natural and social sciences. Instead, the focus is on describing, understanding and explaining the external world, including its social components. The utility of an artefact (e.g. a method for evaluating disaster exercises) is, on the other hand, the main focus of evaluation research (see Section 2) or more broadly design science [161,162]—the "sciences of the artificial", as Herbert Simon put it [163]. These differences in perspective might explain why the reviewed papers pay relatively little attention to the utility of their methods. Nevertheless, we believe that an important step towards a more cohesive and active scientific discourse in the present area lies in using ideas and concepts from evaluation research and design science.

5.2. *Design aspects of disaster exercise evaluation*

Although there is much to say about how to conduct research that aims to provide solutions to real-world problems (see for example the references in Section 2, and [164–168]), here we focus on a few aspects that are important when *reporting* such research.

First of all, the *purpose* of the method, and the context in which it is going to be used needs to be clearly described. The purpose of a method relates to the question "why does it exist, or why should it be built, used or implemented?". As our results show, there are several possible purposes, including learning, development and accountability. Unless the purpose is clearly described, other researchers will not be able to judge whether a proposed method is an improvement over other methods. In addition, methods that are used for learning purposes might not be useful for accountability purposes (and vice versa) as they have different aims and thus produce different results.

Secondly, there needs to be a description of its *function*; in other words, *what* the method does in order to achieve its purpose. Functions are related to the question "what does the method do, when used in context, to achieve its purpose?" We call this the *output* of the method. It describes the *mechanism(s)* by which a method produces the outcome, and achieves its purpose. For example, if the purpose of an evaluation is to learn how to improve performance, it may need to produce an account of how successful the exercise was (i.e. was the outcome good or bad with respect to the scenario and the available resources). It might also need to provide advice on how to improve performance. Output can then be used by participants when thinking about how to improve.

Thirdly, it is important to describe the *form*, or *how* the method produces the necessary output and how that in turn fulfils its purpose. Form is related to the question "how does the method produce the output necessary to achieve its purpose?". The results of the scoping study indicate that papers often include descriptions of how the evaluation method should be used (which is useful), but they seldom provide arguments that support why the suggested method produces outputs that are better at achieving the overall purpose than other methods.

Purpose, function and form thus represent different levels of abstraction that can describe a method (see for example [169,170]). The important point here is that in order for other researchers to judge the usefulness of a new method (i.e. the extent to which it fulfils a specific purpose), they need to see arguments that link form, function and purpose. Many of the 43 core papers identified here include descriptions of form; several of them also discuss purpose; but most lack descriptions of what the method does to achieve its purpose (i.e. functions). Moreover, they also lack a holistic description that clearly links purpose, function and form. For example, claims that a specific method should be used in a specific context (e.g. type of exercise) need to be supported by arguments that show what the application of the method (form) in the context does (function) so that the purpose is fulfilled. Such a description would greatly enhance the ability of other researchers to build upon the results.

Despite this general lack of focus on the links between purpose, function and form, some meta-evaluations came close to such discussions (for example [41]). Improving descriptions and arguments about links would not only improve the ability of other researchers to build on the results, but also motivate practitioners to use a specific evaluation method.

5.3. Impact of disaster exercise evaluation

Few studies discussed in any detail the value of exercises for first responders, or the organisation being prepared for a disaster response. This is surprising, as a very important reason for conducting a disaster exercise in the first place is to strengthen the preparedness of an organisation. A known example of a direct, but accidental, link between exercises (preparedness) and disasters is between 'exercise hurricane Pam' and the real hurricane Katrina. However, it is difficult, if not impossible to establish such direct links since disasters are rare events, and there is even less data gathered on disaster response performance linked to exercises. Our results confirm that little is known about the impact of a disaster exercise on operational preparedness, and how this can best be evaluated. The rare examples we found relate to learning [40], in particular the impact of exercises on participants, but not on the organisation [128,129]. Of course there are many examples of exercises that test and develop new artefacts or methods, but there is no overall method, framework or standard that can express the general outcome of an exercise and its contribution to improving preparedness. Furthermore, the researchers and professionals involved in disaster exercises may consider it obvious that its evaluation, including lessons learnt, always contributes to increased preparedness. This assumption has, however, not received much attention in the literature we reviewed and it would be interesting to see a more thorough investigation of it.

Most of the papers included in the in-depth analysis focussed on rather narrow contexts (e.g. mass casualty incidents and public health). This is understandable, since the context in which the method is developed (and intended to be used) is likely to coincide with the authors' expertise. Nevertheless, it would be useful for the scientific community, and professionals who evaluate disaster exercises to test such methods in a broader context. This would be in line with the argument given above, i.e. to focus on the general, context-independent purpose of disaster evaluation, and try to produce knowledge about how to achieve such purposes in different contexts. This is likely to improve the ability of researchers and professionals to utilize knowledge from different fields. It would probably also help the scientific community to produce *actionable knowledge* [165,171].

Such knowledge is found in so-called *design propositions*, which take the form "If you want to achieve A [purpose] in context B, then do something like C [form]" [166,172]. Propositions regarding the form of the evaluation (C) could be tested in different contexts (B) to investigate whether they are valid, i.e. whether the action actually achieves the purpose (A). The accumulation of empirical evidence would help to establish the scope of the context (B) in which the proposition is valid, which could gradually be refined and expanded. This is likely to increase the relevance of research, and possibly exercises per se. Critics of exercise evaluations often argue that they are only for show and that they have little benefit beyond a symbolic action. However, if researchers focused more on the relationship between purpose and form (as suggested above) more robust arguments regarding, for example, the effect of exercises on preparedness, might be developed. This is likely to reinforce the value of exercises in disaster preparedness activities.

5.4. Limitations

Our focus on *exercise* evaluation represents a significant limitation. Besides exercises that take place in the preparedness phase, substantial research efforts are being made to carry out post-incident and disaster evaluations, and during the response phase. The focus on exercises excluded some key papers relevant to evaluating disaster response exercises; for example, papers related to planning such as Lindell & Perry [173] or Alexander [174]. In addition, other papers have a broader scope and do not focus on exercises, but do provide criteria for evaluating disaster management; these include Quarantelli [175], Borell and Eriksson [176], Abrahamsson et al. [177], Ritchie and MacDonald [178] or Alexander [179]. Also excluded were general studies of evaluating preparedness (such as Nelson et al. [180]) or meta-analyses of real incidents following action reviews [181]. This type of research could potentially be very useful when developing disaster exercise evaluation practices. The choice to exclude these papers was based on the need to reduce the amount of information and provide a more thorough and in-depth analysis of research that explicitly focuses on exercise evaluation. In addition, the decision to exclude grey literature represents a significant limitation as many substantial contributions to the field take the form of informal documents such as reports and articles.

Categorizing and coding the literature (steps 3 and 4 of the scoping study) required judgements to be made, and specific criteria were applied to select the literature and chart the results. This process could impact the study's reliability and validity. Threats to reliability were managed by documenting the procedure, and inclusion and exclusion criteria together with other important choices in the search protocol. The method is described in detail and the study can be easily replicated. Judgements about inclusion and exclusion criteria and the categorization of the results could threaten reliability as interpretations might vary. In order to manage this, a selection of articles was reviewed by another researcher using the same criteria. Any differences were discussed and the final result was agreed between them. In addition, the use of software tools such as Mendeley, MATLAB and VosViewer supported the mapping of the research field and minimized biases associated with manual analysis (such as duplicate removal or qualitative identification of relevant papers). The snowballing method, which is based on reference lists, and peer consultations with librarians in the selection of the search-string, helped to manage threats to reliability. Finally, the scoping study and the results were discussed with a group of experts who provided feedback. Although it is highly unlikely that another researcher repeating our study would end up with exactly the same results (i.e. exactly the same papers assigned to the two groups), we believe that the overall conclusions would be similar.

6. Conclusions and future work

Although evaluation in the context of disaster exercises has considerable practical importance, the results presented here demonstrate that there is a lack of literature that focuses specifically on this topic. Although the field has received increasing attention from researchers in recent times, most contributions seem to focus on individual cases from various (research) perspectives, and there appears to be a lack of interest in building a solid knowledge base. A first step in this direction would be to conduct empirical research that, for example, identifies and compares current exercise evaluation programs or meta-evaluations of particular programs. Despite the overall lack of coherence, there are exceptions that demonstrate a more systematic effort to build on existing work.

The scoping approach used here resulted in the identification of 43 papers that specifically deal with the evaluation of disaster exercises. Although they are published in several subject areas and scientific disciplines, most take a medical perspective. A citation analysis and closer investigation of the content of these publications found three sub-groups of papers in the contexts of public health, hospital preparedness and medical management. A small, fourth group that considered evaluation from a methodological perspective in a disaster context was also identified. These results indicate that contributions are starting to build on each other (or at least refer to each other), even though they are the product of different disciplines.

There are signs of the creation of a more cohesive corpus of scientific information on the evaluation of disaster exercises. However, the results presented here also show that there is room for improvement. Greater clarity, in terms of how specific evaluation methods are linked to their overall purpose is needed. More precisely, attention should be paid to: (a) The purpose(s) and context in which a specific evaluation method is designed to be used; (b) What the method needs to achieve (or produce) in order for it to fulfil its purpose(s); and (c) How the method achieves this and thereby fulfils its purpose. If researchers are to be able to build on

each other's work, it is important to explicitly describe the connections between these three aspects. Moreover, there needs to be a clear description of the available evidence, for example in terms of empirical data or logical reasoning, in support of claims regarding the usefulness of a specific method. Improvements in these areas are likely to lead to a livelier scientific debate and more relevant, robust arguments regarding the evaluation of disaster exercises in practice. In addition, it would be interesting to monitor how research influences the practical design of exercise evaluations that are implemented by emergency responders and other stakeholders.

Given that significant financial and material resources are dedicated to disaster exercises, developing the scientific foundations for their evaluation is clearly important. Researchers should be able to provide advice that is grounded in scientific knowledge, and which has been tested in the field and found to be effective. Such a contribution may help to justify investments, by explaining why it is useful to conduct exercises in the first place. Only then will it become more obvious that disaster exercises are not only for show, and that their evaluation is not simply an exercise in paper pushing.

Acknowledgements

The Institute for Safety (IFV) of the Netherlands and the Swedish Civil Contingencies Agency (MSB) are gratefully acknowledged for having funded the research upon which this article is based.

Appendix A

See:Table A1.

Table A1
Synonyms.

Synonym A	Synonym B	Synonym C
Disaster	**Exercise**	**Evaluation**
Crisis	Drill	Assessment
Emergency	Training	Analysis
Incident	Operation	Valuation
Catastrophe	Performance	Review
Accident	Lesson	Judgement
Contingency	Test	Estimation
Intervention	Activity	Examination
Safety	Learning	Audit
Hazard	Practice	Appraisal
	Simulation	Rating

Appendix B

see:Table B1.

Table B1
Inclusion and exclusion criteria and their rationale.

		Rational/Comments
Inclusion criteria		
I1	Full text publication available in English.	In order to be understood and analysed the full article should be available in English.
I2	Journal is peer-reviewed.	Only articles published in peer-reviewed journals were included. No technical reports, Master's theses, policy documents or other grey literature.
I3	The paper should contain all three elements of the research question i.e. it addresses (operational) exercises in a disaster context/scenario that are evaluated or used as the basis for a response *evaluation*.	If the article only contains one or two of the main elements it does not provide a full picture; therefore, articles should address all three aspects (or synonyms).
I4	Papers that propose a tool/framework that can be applied during exercises to evaluate the disaster response.	If a paper proposes a tool/framework or something similar that is designed to be used for evaluation (without referring to a particular exercise) it was included.
I5	Papers that discuss lessons learned (in combination with evaluation).	To provide insight into how lessons learned are collected. It is expected that this includes some form of evaluation.
Exclusion criteria		
E1	Answer is no to I1, I2, I3	
E2	The document should not solely focus on an incident or preparedness program but specifically deal with a simulated 'real-life' exercise environment.	Some articles focus on the evaluation of incidents and disasters; however, here the focus is on responses in a controlled environment (exercises). The evaluation of preparedness is acceptable if it is based on an exercise evaluation (i.e. excluding preparedness surveys including exercises)
E3	The article should not solely address evaluation as a task, role or function within a response.	Evaluation is not a part (function/task) of the exercise e.g. assessment of victims.
E4	The evaluation of an exercise should not only be mentioned as a task (under exercise design or preparedness approaches) but should be described in detail or applied e.g. gathering lessons learned from processes.	The focus of the research is on disaster exercise evaluation, and in particular how it is performed. It was insufficient for evaluation to simply be mentioned as a task to be executed within the exercise design process. The paper had to provide further details about its execution.
E5	The focus should be on short-term disaster management response *exercises* rather than long-term development/training or education/preparedness programs, courses, plans or projects. Exercises can be evaluated as part of a program; nevertheless, the focus should be on the exercise and not on the program, course, project or training.	The context of the paper was operational exercises that addressed preparations for the acute disaster response phase. Although recovery follows this phase, this is a longer-term issue that is not usually dealt with by first responders, and therefore outside the scope of this research. It is understood that exercises are used to evaluate preparedness; however, the article was only included if it dealt with the evaluation process rather than preparedness.

Appendix C

see:Table C1.

Table C1
In-depth analysis overview.

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Provides information on the (top) authors publishing about disaster exercise evaluation.	Provides general information on the topic of the paper	Provides information on the publications per year in order to possibly identify a trend	(1) North America, (2) South-America, (3) Europe, (4) Africa, (5) Asia, (6) Australia	(1) Formal, (2) Informal, (3) Both, (4) Unknown	(1) Questionnaires, (2) Interviews, (3) Observations (4) Debriefs and combinations of 1, 2 and 3.	(1) Quantitative, (2) Qualitative, (3) Both, (4) Not explicitly mentioned.	(1) Yes, (2) No. If (1) it will be specified.	(1) Yes, (2) No or (3) Unknown. If (1) it will be specified.	(1) Natural disaster: (a) Flood, (b) earthquake, (c) disease, (d).; (2) Technological Disaster: (a) aircraft incident, (b) chemical incident, (c) nuclear incident, (d) ...	(1) local/ regional/ state, (2) national, (3) international or combinations of 1, 2 and 3	(1) Yes with previous exercises, (2) Yes, with previous incidents, (3) Yes, with both, (4) No, (5) Unknown	(1) Measuring, (2) Describing, (3) Judging and combinations of 1, 2 and 3 (4) Unknown /not possible to determine	(1) Yes, (2) No. If (1) it will be specified. (3). Unknown	(1) Learning/ development purpose, (2) Accountability, (3) Both, (4) Unkonwn (or Only presentation of method)	(1) Whole, (2) Parts, (3) Unknown.	(1) Formative evaluation, (2) Summative evaluation, (3) Both, (4) Unknown.
			Provides information on what country or continent disaster manage-	Provides information if the evaluation was performed formally, infor-	Provides information on how evaluation data systematically was collec-	Provides information on the type of information that is collected and what	Provides information on if there are specific methods that can be used to evaluate exercises i.e. they reference to a specific method for evaluation	Provides information whether the author gives arguments for the selection of the evaluation method to be used in exercises? (why is it useful)	Provides information on the context of the evaluation and/or exercise.	Provides information on the scale of the exercise i. e. local/ regional, national or international.	Provides information on the origin and relationship of the scenario with previous experiences.	Provides information on what the evaluation intends to pursue i.e. measuring, describing or judging (generations of evaluation).	Provides information on the stakeholder involvement in evaluation (4th gen. Evaluation).	Provides information on the purpose of the evaluation/ method.	Was the whole disaster system evaluated or only parts?	Provides information on when the exercise was executed i.e. during a development program or at the end.

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
			ment evaluation is receiving scientific attention and that is published.	mally or both.	ted.	indicators/ criteria were used.	(not general method) described elsewhere or provide a description of the method used.									
Crankshaw, Donald, H.; Fenton, Roland, D.	Evaluation criteria for Emergency Plan Drills and Exercises	1983	1	1	3, 1	1	1	1	N/A	N/A	N/A	1	2	2	1	4
Goodwin, C.	Disaster drills.	1986	1	1	3, 4	1	1	1	3	1	5	1	1	2	1	2
Villoria, Carlos; Gomez, Elia; Tavel, Nelson García	National Drill of the Venezuelan Oil Spill Contingency Plan	1995	2	1	3, 1	3	1	1	2	2	5	1	3	3	1	2

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Wood, C. J. B.; Foster, H. D.; Hardy, N. E.	Crisis Simulation and Health Care Systems	1997	1	1	1, 3, 4	1	1	3	1	2	1	1	1	3	1	2
Tyre, Te	Wake-up call: a bioterrorism exercise.	2001	1	1	1, 3	1	2	2	2	1.2	5	4	3	2	3	2
Green, Gary B.; Modi, Surbhi; Lunney, Kevin; Thomas, Tamara L.	Generic evaluation methods for disaster drills in developing countries	2003	2	1	3, 4	3	1	1	2	1	4	1	3	Research	2	4
Carrel, Laurent F.	"Epidemic in Switzerland": Description of a strategic leadership exercise	2005	3	1	1, 3	1	2	2	1	2	2	4	3	2	3	2

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	by the Swiss government															
Lonka, Harriet; Wybo, Jean Luc	Sharing of experiences: a method to improve usefulness of emergency exercises	2005	3	1	2.4	1	1	1	2	1	5	4	1	1	3	4
Thomas, Tamara L; Hsu, Edbert B; Kim, Hong K; Colli, Sara; Arana, Guillermo; Green, Gary B	The incident command system in disasters: evaluation methods for a hospital-based exercise.	2005	2	1	3.4	3	2	1	2	1	4	1	3	1	2	1

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Klein, Kelly R; Brandenburg, Dale C; Atas, Jenny G; Maher, Ann	The use of trained observers as an evaluation tool for a multi-hospital bioterrorism exercise.	2005	1	1	1, 3	3	1	1	2	1.2	5	2	3	1	1	4
Rüter, Anders; Nilsson, Heléne; Vilkström, Tore	Performance indicators as quality control for testing and evaluating hospital management groups: a pilot study.	2006	3	1	3	2	1	1	2	1	5	1	3	4	2	4
Gebbie, Kristine M;	Role of exerci-	2006	1	1	N/A	3	N/A	1	N/A	N/A	N/A	1	1	N/A	N/A	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Valas, Joan; Merrill, Jacqueline; Morse, Stephen	ses and drills in the evaluation of public health in emergency response.															
Steward, Duane; Wan, T. T H	The role of simulation and modelling in disaster management	2007	1	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	4	3	4	N/A	4
Kaji, Amy H.; Langford, Vinette; Lewis, Roger J.	Assessing Hospital Disaster Preparedness: A Comparison of an On-Site Survey, Directly	2008	1	1	1.3	3	1	1	2	1	5	1	3	1	2	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	Observed Drill Performance, and Video Analysis of Teamwork															
Kaji, Amy H.; Lewis, Roger J.	Assessment of the Reliability of the Johns Hopkins/ Agency for Healthcare Research and Quality Hospital Disaster Drill Evaluation Tool	2008	1	1	3	3	1	1	2	1	5	1.2	3	Research	1	N/A
Allen, Korrie;	Con-	2008	1	1	1.3	3	1	1	2	1	2	1.2	3	4	2	2

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Lorek, Edward; Mensia-Joseph, Nita	ducting a school-based mock drill: lessons learned from one community.															
Nilsson, Heléne; Rüter, Anders	Management of resources at major incidents and disasters in relation to patient outcome: a pilot study of an educational model.	2008	3	1	3	2	2	3	2	1	4	1	3	4	2	4
Biddinger,	On	2008	1	1	3.4	1	1	N/A	N/A	N/A	1	4	1	N/A	N/A	N/A

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Paul D.; Cadigan, Rebecca Orfaly; Auerbach, Bruce S.; Burstein, Jonathan L.; Savoia, Elena; Stoto, Michael a.; Koh, Howard K.	Linkages: Using Exercises to Identify Systems-Level Preparedness Challenges															
Edwards, Janine C.; Kang, Jungeun; Silenas, Rasa	Promoting regional disaster preparedness among rural hospitals	2008	1	1	3	3	2	2	1	1	2	4	3	3	1	4
Wybo, Jean-Luc	The Role of Simulation Exercises in the Assessment of Robustness	2008	3	1	2.3	1	1	1	2	N/A	N/A	2	3	1	3	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	and Resilience of Private or Public Organisations															
Savoia, Elena; Testa, Marcia a.; Biddinger, Paul D.; Cadigan, Rebecca O.; Koh, Howard; Campbell, Paul; Stoto, Michael a.	Assessing Public Health Capabilities During Emergency Preparedness Tabletop Exercises: Reliability and Validity of a Measurement Tool	2009	1	1	1	1	1	1	1	1	5	3	1	Research	3	N/A
Lenz, Matthias; Richter, Tanja	Disaster response to the	2009	3	1	3	1	1	1	2	1	5	1	3	1	1	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	release of biohazardous agent: instrument development and evaluation of a firefighter's exercise.															
Latiers, Melanie; Jacques, Jean Marie	Emergency and crisis exercises: methodology for understanding safety dimensions	2009	3	1	3	1	1	1	2	N/A	N/A	N/A	N/A	N/A	1	N/A
Quaranta, J. D.; Childers, H. M.;	Lessons learned from tabletop	2009	1	1	3.4	1	2	2	3	1	2	4	3	4	3	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Myles, P.	reviews of emergency action plans for high hazard dams in West Virginia, USA															
Gryth, Dan; Rüter, Anders	Evaluation of medical command and control using performance indicators in a full-scale, major aircraft accident exercise.	2010	3	1	3	3	1	2	2	1	5	1	2	1	3	1
Ingrassia, Pier Luigi; Prato,	Evaluation of medical	2010	3	1	3	2	1	1	2	1	4	1	1	4	2	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Federico; Geddo, Alessandro; Colombo, Davide; Tengattini, Marco; Calligaro, Sara; La Mura, Fabrizio; Franc, Jeffrey Michael; Della Corte, Francesco	management during a mass casualty incident exercise: an objective assessment tool to enhance direct observation.															
Savoia, Elena; Biddinger, Paul D; Burstein, Jon; Stoto, Michael a	Interagency communication and operations capabilities during a hospital functional exercise: reliability and	2010	1	1	1, 3, 4	1	1	1	1	1	1	1	3	3	3	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	validity of a measurement tool.															
High, Erika H; Lovelace, Kay a; Gansneder, Bruce M; Strack, Robert W; Callahan, Barbara; Benson, Phillip	Promoting community preparedness: lessons learned from the implementation of a chemical disaster tabletop exercise.	2010	1	1	1, 3, 4	1	2	2	2	1	3	4	2	1	3	4
Biddinger, Paul D.; Savoia, Elena; Massin-Short, Sarah B.; Preston, Jessica; Stoto,	Public Health Emergency Preparedness Exercises: Lessons Learned	2010	1	1	1, 3, 4	3	N/A	N/A	N/A	4	5	N/A	N/A	N/A	N/A	N/A

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Michael a. Nilsson, Heléne; Vikström, Tore; Rüter, Anders	Quality control in disaster medicine training– initial regional medical command and control as an example.	2010	3	1	3	3	1	1	N/A	4	5	1.3	3	3	3	4
Hirt, R.; Lörtscher, Y	Planning, implementation and evaluation of an emergency exercise at the Paul Scherrer Institute	2011	3	1	3	1	2	2	2	4	1	3	3	2	3	2

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Johnston, David; Tarrant, Ruth; Tipler, Karlene; Coomer, Maureen; Pedersen, Sandy; Garside, Ruth	Preparing schools for future earthquakes in New Zealand: Lessons from an evaluation of a Wellington school exercise	2011	6	1	3	1	1	2	1	1	5	2	3	3	3	4
Sinclair, Helen	Assessing emergency management training and exercises	2012	1.6	1	1, 2, 3, 4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
R/aa destad, Monica; Nilsson, Heléne; Castrén, Maaret;	Combining performance and out-	2012	3	1	3	3	1	1	2	1	1	1.3	3	4	2	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Svensson, Leif; Rüter, Anders; Gryth, Dan	come indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises															
Morris, J. Glenn; Greenspan, Allison; Howell, Kelly; Gargano, Lisa M.; Mitchell, Joanne; Jones, Jeffrey L.; Potter, Morris; Isakov,	Southeastern Centre for Emerging Biologic Threats Tabletop Exercise: Foodborne Toxo-	2012	1	1	1, 3	1	2	2	1	1	5	1.3	3	1	3	4

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Alexander; Woods, Christopher; Hughes, James M.	plasmosis Outbreak on College Campuses															
Heumüller, Erich; Richter, Sebastian; Lechner, Ulrike	Towards a framework for command post exercises	2012	3	1	1, 3	1	1	1	3	1	1	1	3	Research	3	4
Savoia, Elena; Preston, Jessica; Biddinger, Paul D	A consensus process on the use of exercises and after action reports to assess and improve public health emer-	2013	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	gency preparedness and response.															
Duarte, Alayne; Duarte, Costa; Gomes, José Orlando; Victor, Paulo; Carvalho, Rodrigues De	ASC Model : a process model for the evaluation of simulated field exercises in the emergency domain	2013	2	1	3	N/A	1	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Agboola, Foluso; McCarthy, Tara; Biddinger, Paul D	Impact of emergency preparedness exercise on performance.	2013	1	1	3	3	1	1	2	1	4	1.3	3	Research	3	1
Kim, Hakkyong	Improving simula-	2013	5	1	2.3	N/A	N/A	N/A	1, 2	1	N/A	N/A	N/A	Research	N/A	N/A

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
	tion exercises in Korea for disaster preparedness															
Savoia, Elena; Agboola, Foluso; Biddinger, Paul	A Conceptual Framework to Measure Systems' Performance during Emergency Preparedness Exercises	2014	1	1	N/A	3	1	1	N/A	N/A	N/A	1.3	1	3	3	3
Cranmer, Hilarie; Chan, Jennifer L; Kayden, Stephanie; Musani, Altaf; Gasquet,	Development of an evaluation framework suitable for	2014	4	1	2.3	1	1	1	N/A	N/A	N/A	1	3	2	3	3

(continued on next page)

Table C1 *(continued)*

				A			B	C	D			E		F		G
General information				Form of the evaluation (how)?			Evaluation method	Justification/ Explanation	Context			Function of the evaluation (what?)		Purpose of the evaluation (why?)		Timing of the evaluation
Author(s)	**Title**	**Year**	**Continent**	**How is the evaluation performed?**	**What data collection methods were used?**	**What information was collected?**	**Was there a specific method presented for running the evaluation and identifying the information to be collected?**	**Is the method justified/ argumented?**	**What was the context/ scenario of the exercise?**	**What was the 'scale' of the exercise?**	**Was there an explicit relationship with previous exercises and/or incidents?**	**What does the evaluation do?**	**Were various stakeholders involved in the setup of the evaluation?**	**Why was the exercise evaluated?**	**Was it a holistic evaluation?**	**When was the evaluation executed with the exercise program?**
Philippe E; Walker, Peter; Burkle, Frederick M; Johnson, Kirsten	assessing humanitarian workforce competencies during crisis simulation exercises.															
Djalali, Ahmadreza; Carenzo, Luca; Ragazzoni, Luca; Azzaretto, Massimo; Petrino, Roberta; Della Corte, Francesco; Ingrassia, Pier Luigi	Does Hospital Disaster Preparedness Predict Response Performance During a Full-scale Exercise? A Pilot Study.	2014	3	1	3	2	1	1	2	1	5	1	3	Research	3	N/A

References

[1] D. Guha-Sapir, P. Hoyois, R. Below, Annual Disaster Statistical Review 2014: The numbers and trends, Louvain, Brussels, Belgium, 2015. ⟨http://www.cred.be/sites/default/files/ADSR_2014.pdf⟩.

[2] UNISDR, Making Development Sustainable: The Future of Disaster Risk Management. Global Assessment Report on Disaster Risk Reduction, 2015.

[3] A. Kirschenbaum, Chaos Organization and Disaster Management, Marcel Dekker, New York, 2004. http://dx.doi.org/10.1111/j.1467-9299.2006.00628_2.x.

[4] D.E. Alexander, Principles of Emergency Planning and Management, Oxford University Press, USA, New York, 2002.

[5] Joint Committee on Standards for Educational Evaluation, The Program Evaluation Standards, second ed., Corwin Press, Thousand Oaks, 1994.

[6] D. Stufflebeam, C. Coryn, Evaluation Theory, Models, and Applications, second ed., Jossey-Bass, San Francisco, 2014.

[7] Joint Committee on Standards for Educational Evaluation, The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users, Third ed., SAGE Publications Inc., Thousand Oaks, 2011.

[8] E.G. Guba, Y.S. Lincoln, Fourth Generation Evaluation, SAGE Publications Inc., Newbury Park, 1989.

[9] R.E.E. Stake, Evaluating the Arts in Education: A Responsive Approach, Merrill, Columbus, 1974.

[10] N. Hertting, E. Vedung, Purposes and criteria in network governance evaluation: how far does standard evaluation vocabulary takes us?, Evaluation 18 (2012) 27–46. http://dx.doi.org/10.1177/1356389011431021.

[11] M. Scriven, The methodology of evaluation, in: R.W. Tyler, R. Gagne, M. Scriven (Eds.), Perspectives of Curriculum Evaluation, Rand McNally, Skokie, 1967, pp. 39–83.

[12] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, J. Soc. Res. Methodol. (2005) 19–32 ⟨http://www.tandfonline.com/doi/abs/10.1080/1364557032000119616⟩ (accessed February 6, 2014).

[13] D. Levac, H. Colquhoun, K.K.O. Brien, Scoping studies: advancing the methodology, Implement. Sci. 5 (2010) 69. http://dx.doi.org/10.1186/1748-5908-5-69.

[14] M.J. Grant, A. Booth, A typology of reviews: an analysis of 14 review types and associated methodologies, Health Inf. Libr. J. 26 (2009) 91–108. http://dx.doi.org/10.1111/j.1471-1842.2009.00848.x.

[15] R.F. Baumeister, M.R. Leary, Writing narrative literature reviews, Rev. Gen. Psychol. 1 (1997) 311–320.

[16] H.M.L. Daudt, C. van Mossel, S.J. Scott, Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework, BMC Med. Res. Methodol. 13 (2013) 48. http://dx.doi.org/10.1186/1471-2288-13-48.

[17] Thesaurus.com, Roget's 21st Century Thesaurus, (n.d.). ⟨http://www.thesaurus.com⟩, (accessed 11.04.16).

[18] N.J. van Eck, L. Waltman, Software survey: vosviewer, a computer program for bibliometric mapping, Scientometrics 84 (2010) 523–538. http://dx.doi.org/10.1007/s11192-009-0146-3.

[19] N.J. Van Eck, L. Waltman, Visualizing bibliometric networks, in: Y. Ding, R. Rousseau, D. Wolfram (Eds.), , Measuring Scholarly Impact: Methods and Practice, Springer, 2014, pp. 285–320. http://dx.doi.org/10.1007/978-3-319-10377-8_13.

[20] J. Lecy, G. State, Structured literature reviews using constrained snowball sampling and citation network analysis, SSRN (2012) 1–15.

[21] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (2005) 814–818. http://dx.doi.org/10.1038/nature03607.

[22] R.R. Powell, Evaluation research: an overview, Libr. Trends 55 (2006) 102–120. http://dx.doi.org/10.1353/lib.2006.0050.

[23] C.J. Goodhue, R.V. Burke, S. Chambers, R.R. Ferrer, J.S. Upperman, Disaster Olympix: a unique nursing emergency preparedness exercise, J. Trauma Nurs. 17 (2010) 5–10. http://dx.doi.org/10.1097/JTN.0b013e3181d915f0.

[24] A.H. Kaji, V. Langford, R.J. Lewis, Assessing hospital disaster preparedness: a comparison of an on-site survey, directly observed drill performance, and video analysis of teamwork, Ann. Emerg. Med. 52 (2008) 195–201. http://dx.doi.org/10.1016/j.annemergmed.2007.10.026.

[25] O.M. Araz, M. Jehn, Improving public health emergency preparedness through enhanced decision-making environments: a simulation and survey based evaluation, Technol. Forecast. Soc. Change 80 (2013) 1775–1781. http://dx.doi.org/10.1016/j.techfore.2012.09.018.

[26] J.C. Hunter, J.E. Yang, M. Petrie, T.J. Aragón, Integrating a framework for conducting public health systems research into statewide operations-based exercises to improve emergency preparedness, BMC Public Health 12 (2012) 680. http://dx.doi.org/10.1186/1471-2458-12-680.

[27] E. Savoia, M. a Testa, P.D. Biddinger, R.O. Cadigan, H. Koh, P. Campbell, M.A. Stoto, Assessing public health capabilities during emergency preparedness tabletop exercises: reliability and validity of a measurement tool, Public Health Rep. 124 (2009) 138–148 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-67650066804 & partnerID=tZOtx3y1⟩.

[28] C.J.B. Wood, H.D. Foster, N.E. Hardy, Crisis simulation and health care systems, Simul. Gaming 28 (1997) 198–216. http://dx.doi.org/10.1177/1046878197282004.

[29] E. Savoia, P.D. Biddinger, J. Burstein, M.A. Stoto, Inter-agency communication and operations capabilities during a hospital functional exercise: reliability and validity of a measurement tool, Prehosp. Disaster Med. 25 (2010) 52–58. http://dx.doi.org/10.1017/S1049023X00007664.

[30] J.G. Morris, A. Greenspan, K. Howell, L.M. Gargano, J. Mitchell, J.L. Jones, M. Potter, A. Isakov, C. Woods, J.M. Hughes, Southeastern center for emerging biologic threats tabletop exercise: foodborne toxoplasmosis outbreak on college campuses, Biosecur. Bioterror. Biodef. Strateg. Pract. Sci. 10 (2012) 89–97. http://dx.doi.org/10.1089/bsp.2011.0040.

[31] R. Hirt, Y. Lörtscher, Planning, implementation and evaluation of an emergency exercise at the Paul Scherrer Institute, Kerntechnik, 76, 456–459. ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-84855769706 & partnerID=tZOtx3y1⟩, 2011 (accessed 06.10.14).

[32] E. Wahl, B. Willumsen, L. Jensvoll, I.H. Finstad, T.M. Berglund, The learning effect of a foodborne emergency exercise, Br. Food J. 117 (2015) 1981–1994. http://dx.doi.org/10.1108/BFJ-10-2014-0340.

[33] G.D. Gilmore, W.R. Schwan, M.K. McLaughlin, An assessment of emergency preparedness in western Wisconsin, WMJ 106 (2007) 71–77 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-34247139643 & partnerID=tZOtx3y1⟩.

[34] C. Villoria, E. Gomez, N.G. Tavel, National Drill of the Venezuelan Oil Spill Contingency Plan, in: International Oil Spill Conference, pp. 513–522, ⟨http://dx.doi.org/10.7901/2169-3358-1995-1-513⟩, 1995.

[35] T.E. Tyre, Wake-up call: a bioterrorism exercise, Mil. Med. (2001) 90–91 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0035664183 & partnerID=tZOtx3y1⟩.

[36] J.M. Berlin, E.D. Carlström, The dominance of mechanistic behaviour: a critical study of emergency exercises, Int. J. Emerg. Manag. 9 (2013) 327. http://dx.doi.org/10.1504/IJEM.2013.059878.

[37] W. Muhren, B. Van De Walle, Sensemaking and information management in humanitarian disaster response: observations from the TRIPLEX exercise, in: Proceedings of the 6th International. Conference on Information Systems for Crisis Response and Management, ISCRAM, pp. 1–8. ⟨file:///C:/Users/Φρόσω/Downloads/3deec523b45c0c81ec.pdf⟩, 2009.

[38] H. Cranmer, J.L. Chan, S. Kayden, A. Musani, P.E. Gasquet, P. Walker, F.M. Burkle, K. Johnson, Development of an evaluation framework suitable for assessing humanitarian workforce competencies during crisis simulation exercises, Prehosp. Disaster Med. 29 (2014) 69–74. http://dx.doi.org/10.1017/S1049023X13009217.

[39] E.B. Hsu, M.W. Jenckes, C.L. Catlett, K.A. Robinson, C. Feuerstein, S.E. Cosgrove, G.B. Green, E.B. Bass, Effectiveness of hospital staff mass-casualty incident training methods: a systematic literature review, Prehosp. Disaster Med. Off. J. Natl. Assoc. EMS Physicians World Assoc. Emerg. Disaster Med. Assoc. Acute Care Found 19 (2004) 191–199. http://dx.doi.org/10.1017/S1049023X00001771.

[40] E.E. Borodzicz, K. van Harperen, K. van Haperen, Individual and group learning in crisis simulations, J. Contingencies Cris. Manag 10 (2002) 139–147. http://dx.doi.org/10.1111/1468-5973.00190.

[41] P.D. Biddinger, R.O. Cadigan, B.S. Auerbach, J.L. Burstein, E. Savoia, M. a Stoto, H.K. Koh, L. Jonathan, E. Savoia, M. a Stoto, H.K. Koh, On linkages: using exercises to identify systems-level preparedness challenges, Public Health Rep. 123 (2008) 96–101.

[42] E.H. High, K. a Lovelace, B.M. Gansneder, R.W. Strack, B. Callahan, P. Benson, Promoting community preparedness: lessons learned from the implementation of a chemical disaster tabletop exercise, Health Promot. Pract. 11 (2010) 310–319. http://dx.doi.org/10.1177/1524839908325063.

[43] B. Adini, A. Goldberg, R. Cohen, Y. Bar-Dayan, Relationship between equipment and infrastructure for pandemic influenza and performance in an avian flu drill, Emerg. Med. J. 26 (2009) 786–790. http://dx.doi.org/10.1136/emj.2008.066746.

[44] G. Tami, A. Bruria, E. Fabiana, C. Tami, A. Tali, A.-D.D. Limor, An after-action review tool for EDs: learning from mass casualty incidents, Am. J. Emerg. Med. 31 (2013) 798–802. http://dx.doi.org/10.1016/j.ajem.2013.01.025.

[45] B. Adini, A. Goldberg, R. Cohen, Y. Bar-Dayan, Impact of pandemic flu training on ability of medical personnel to recognize an index case of avian influenza, Eur. J. Public Health 22 (2012) 169–173. http://dx.doi.org/10.1093/eurpub/ckr030.

[46] E. Ablah, E. Scanlon, K. Konda, A. Tinius, K.M. Gebbie, A large-scale points-of-dispensing exercise for first responders and first receivers in Nassau County, New York, Biosecur. Bioterror. 8 (2010) 25–35. http://dx.doi.org/10.1089/bsp.2009.0014.

[47] M.W. Jenckes, C.L. Catlett, E.B. Hsu, K. Kohri, G.B. Green, K.A. Robinson, E.B. Bass, S.E. Cosgrove, Development of evaluation modules for use in hospital disaster drills, Am. J. Disaster Med. 2 (2007) 87–95 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-40349094875 & partnerID=tZOtx3y1⟩.

[48] A. Kilianski, A.T.O. Rourke, C.L. Carlson, S.M. Parikh, F. Shipman-amuwo, A. O'Rourke, The planning, execution, and evaluation of a mass prophylaxis full-scale exercise in cook County, IL, BiosecurBioterror. 12 (2014) 106–116. http://dx.doi.org/10.1089/bsp.2013.0089.

[49] L.S. Hites, M.M. Sass, L.D. Ambrosio, L.M. Brown, A.M. Wendelboe, K.E. Peters, R.K. Sobelson, The preparedness and emergency response learning centers: advancing standardized, Eval. Public Health Prep. Response Train. 35294 (2014) 17–23. http://dx.doi.org/10.1097/PHH.0000000000000066.

[50] R. Nathawad, P.M. Roblin, D. Pruitt, B. Arquilla, Addressing the gaps in preparation for quarantine, Prehosp. Disaster Med. 28 (2013) 132–138. http://dx.doi.org/10.1017/S1049023X1200180X.

[51] K. Lerner, Using a simulation cell for exercise realism, J. Emerg. Manag. 11 (2013) 338–344.

[52] E.M. Wilson, A new concept in emergency operations exercises: peer evaluation, J. Emerg. Nurs. 21 (1995) 435–436. http://dx.doi.org/10.1016/S0099-1767(05)80118-7.

[53] B. Adini, A. Goldberg, D. Laor, R. Cohen, Y. Bar-Dayan, Factors that may influence

the preparation of standards of procedures for dealing with mass-casualty incidents, Prehosp. Disaster Med. Off. J. Natl. Assoc. EMS Physicians World Assoc. Emerg. Disaster Med. Assoc. Acute Care Found 22 (2007) 175–180. http://dx.doi.org/10.1017/S1049023X00004611.
[54] K. Hardie, T. Kitchen, Operation crash and surge: lessons learned from a region-wide exercise, J. Bus. Contin. Emer. Plan. 7 (2014) 302–311 ⟨http://www.ncbi.nlm.nih.gov/pubmed/24854731⟩.
[55] A. Rüter, H. Nilsson, T. Vilkström, Performance indicators as quality control for testing and evaluating hospital management groups: a pilot study, Prehosp. Disaster Med. 21 (2006) 423–426. http://dx.doi.org/10.1017/S1049023X00004131.
[56] K.R. Klein, D.C. Brandenburg, J.G. Atas, A. Maher, The use of trained observers as an evaluation tool for a multi-hospital bioterrorism exercise, Prehosp. Disaster Med. Off. J. Natl. Assoc. EMS Physicians World Assoc. Emerg. Disaster Med. Assoc. Acute Care Found 20 (2005) 159–163. http://dx.doi.org/10.1017/S1049023X00002387.
[57] D. a Klima, S.H. Seiler, J.B. Peterson, a B. Christmas, J.M. Green, G. Fleming, M.H. Thomason, R.F. Sing, Full-scale regional exercises, J. Trauma Acute Care Surg. 73 (2012) 592–598. http://dx.doi.org/10.1097/TA.0b013e318265cbb2.
[58] E.K. Lee, S. Maheshwary, J. Mason, W. Glisson, Decision support system for mass dispensing of medications for infectious disease outbreaks and bioterrorist attacks, Ann. Oper. Res. 148 (2006) 25–53. http://dx.doi.org/10.1007/s10479-006-0087-7.
[59] J.C. Edwards, J. Kang, R. Silenas, Promoting regional disaster preparedness among rural hospitals, J. Rural Heal 24 (2008) 321–325. http://dx.doi.org/10.1111/j.1748-0361.2008.00176.x.
[60] G.N. Rutty, J.E. Rutty, Did the participants of the mass fatality exercise operation torch learn anything?, Forensic Sci. Med. Pathol. 8 (2012) 88–93. http://dx.doi.org/10.1007/s12024-010-9218-1.
[61] F. Agboola, T. McCarthy, P.D. Biddinger, Impact of emergency preparedness exercise on performance, J. Public Heal. Manag. Pract. 19 (Suppl. 2) (2013) S77–S83. http://dx.doi.org/10.1097/PHH.0b013e31828ecd84.
[62] M. Rådestad, H. Nilsson, M. Castrén, L. Svensson, A. Rüter, D. Gryth, Combining performance and outcome indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises, Scand. J. Trauma. Resusc. Emerg. Med. 20 (2012) 58. http://dx.doi.org/10.1186/1757-7241-20-58.
[63] S.D. Smith, R. Smith, J. Albanese, E. Forte, J.L. Paturas, W. Halstead, A. Tomassoni, Disaster and exercise performance information collection tool: capturing observations in four minutes or less, J. Bus. Contin. Emer. Plan. 6 (2013) 151–163 ⟨http://www.ncbi.nlm.nih.gov/pubmed/23315250⟩.
[64] J.B. Mitchell, Exercise Builder International - A tool for developing exercises, in: 3rd Int. Jt. Top. Meet. Emerg. Prep. Response Robot. Remote Syst. 2011, EPRRSD, 13th Robot. Remote Syst. Hazard. Environ., 2011: pp. 267–277, ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-84855720509 & partnerID=40 & md5=5e6a2d155ee9a8b997669df38df9d293⟩.
[65] H.K. Shin, J.Y. Kim, M.K. Kim, S.H. Kim, Development of nuclear emergency exercise programme (NEEP) in Korea, Radiat. Prot. Dosim. 146 (2011) 46–49. http://dx.doi.org/10.1093/rpd/ncr104.
[66] M. Lenz, T. Richter, Disaster Response to the Release of Biohazardous Agents : Instrument Development and Evaluation of a Firefighter's Exercise, 2, 2009.
[67] J. Caum, S. Alles, Ready or not: analysis of a no-notice mass vaccination field response in Philadelphia, Biosecur. Bioterror. Biodef. Strateg. Pract. Sci. 11 (2013) 262–270. http://dx.doi.org/10.1089/bsp.2013.0064.
[68] P.F. Lau, C.C. Lau, A disaster drill in Hong Kong, Accid. Emerg. Nurs. 5 (1997) 34–38. http://dx.doi.org/10.1016/S0965-2302(97)90061-8.
[69] Joyce Hood, Jessica Fernandes-Flack, Michael D. Larrañaga, Effectiveness of hospital-based decontamination during a simulated mass casualty exposure, J. Occup. Environ. Hyg. 8 (12) (2011) D131–D138. http://dx.doi.org/10.1080/15459624.2011.626267.
[70] H. Nilsson, A. Rüter, Management of resources at major incidents and disasters in relation to patient outcome: a pilot study of an educational model, Eur. J. Emerg. Med. 15 (2008) 162–165. http://dx.doi.org/10.1097/MEJ.0b013e3282f4d14b.
[71] D. Gryth, M. Radestad, H.H. Nilsson, O. Nerf, L. Svensson, M. Castren, A. Riiter, M. Rådestad, M. Castrén, A. Rüter, H.H. Nilsson, O. Nerf, L. Svensson, M. Castrén, A. Rüter, Evaluation of medical command and control using performance indicators in a full-scale, major aircraft accident exercise, Prehosp. Disaster Med. 25 (2010) 118–123. http://dx.doi.org/10.1017/S1049023X00007834.
[72] P.M. Paris, R.D. Stewart, G.H. Pelton, G. Porter, A. Sanzo, Triage success in disasters: dynamic victim-tracking cards, Am. J. Emerg. Med. 3 (1985) 323–326. http://dx.doi.org/10.1016/0735-6757(85)90057-9.
[73] D. Barton, G.G. Bodiwala, Assessment of a triage label system during a major incident exercise, Prehosp. Disaster Med. Off. J. Natl. Assoc. EMS Physicians World Assoc. Emerg. Disaster Med. Assoc. Acute Care Found 6 (1991) 473–476 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0026236234 & partnerID=tZOtx3y1⟩.
[74] J. Muskat, M. Lampinen, R. Imai, The use of GIS technology at the California SONS 2004 extravaganza, in: 2005 International Oil Spill Conference (IOSC 2005), pp. 10837–10840. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-33646054131 & partnerID=tZOtx3y1⟩, 2005.
[75] M. Korner, L.L. Geyer, S. Wirth, C.-D. Meisel, M.F. Reiser, U. Linsenmaier, M. Körner, L.L. Geyer, S. Wirth, C.-D. Meisel, M.F. Reiser, U. Linsenmaier, Analysis of responses of radiology personnel to a simulated mass casualty incident after the implementation of an automated alarm system in hospital emergency planning, Emerg. Radiol. 18 (2011) 119–126. http://dx.doi.org/10.1007/s10140-010-0922-7.
[76] M. Leingartner, J. Maurer, G. Steinbauer, A. Ferrein, Evaluation of sensors and mapping approaches for disasters in tunnels, in: 2013 IEEE International. Symp. Safety, Secur. Rescue Robot. SSRR 2013, IEEE, pp. 1–7. ⟨http://dx.doi.org/10.1109/SSRR.2013.6719349⟩, 2013.
[77] E.R. Messina, A.S. Jacoff, Measuring the performance of urban search and rescue robots, in: 2007 IEEE Conference. Technol. Homel. Secur., IEEE, pp. 28–33. ⟨http://dx.doi.org/10.1109/THS.2007.370015⟩, 2007.
[78] J.L. Burke, R.R. Murphy, M.D. Coovert, D.L. Riddle, Moonlight in Miami: field study of human-robot interaction in the context of an urban search and rescue disaster response, Train. Exerc. Hum.-Comput. Inter. 19 (2004) 85–116. http://dx.doi.org/10.1080/07370024.2004.9667341.
[79] Y. Okada, Y. Haruki, Y. Ogushi, Disaster drills and continuous medical education using satellite-based internet, Methods Inf. Med. 39 (2000) 343–347 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0034524119 & partnerID=tZOtx3y1⟩.
[80] Y. Javed, T. Norris, D. Johnston, Design, implementation and evaluation of web-based integrated incident Resource Management System for decision support in emergency operation centres, in: International. Conference. Enterp. Inf. Syst. Web Technol. 2009, EISWT 2009, 2009: pp. 209–216. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84878122458 & partnerID=tZOtx3y1⟩.
[81] C.S. Sandlin, R.C. Johnson, L. Swaim, D.L. Ashley, Laboratory information management system for emergency response: validation and quality assurance of analytical methodologies, J. Assoc. Lab. Autom. 14 (2009) 126–132. http://dx.doi.org/10.1016/j.jala.2009.02.001.
[82] D. a Edwards, H. Cuthbertson, D. Peterson, Developing and testing an emergency notification system for a county emergency management agency, J. Homel. Secur. Emerg. Manag. 8 (2011). http://dx.doi.org/10.2202/1547-7355.1911.
[83] A. Al-akkad, L. Ramirez, A. Boden, D. Randall, A. Zimmermann, Help beacons: design and evaluation of an Ad-Hoc lightweight S.O.S. system for smartphones, in: CHI'14 Proceedings. SIGCHI Conference. Hum. Factors Comput. Syst., 1485–1494. ⟨http://dx.doi.org/10.1145/2556288.2557002⟩, 2014.
[84] H.B. Andersen, H. Garde, V. Andersen, M.M.S. An, electronic message management system for emergency response, IEEE Trans. Eng. Manag. 45 (1998) 132–140. http://dx.doi.org/10.1109/17.669758.
[85] Y.C. Liu, P.H. Yin, J. Zhang, G. Ren, The research and development of simulation exercise system of marine oil spill crisis response based on GIS, 2003. Go to ISI: ⟨//WOS:000186892000075⟩.
[86] B. Gillett, B. Peckler, R. Sinert, C. Onkst, S. Nabors, S. Issley, C. Maguire, S. Galwankarm, B. Arquilla, Simulation in a disaster drill: comparison of high-fidelity simulators versus trained actors, Acad. Emerg. Med. 15 (2008) 1144–1151. http://dx.doi.org/10.1111/j.1553-2712.2008.00198.x.
[87] B. Osatuyi, D. Ding, M. Chumer, Shareable technology mash-up for situation awareness in catastrophic events, in: Proceedings. Annu. Hawaii International. Conference. Syst. Sci, pp. 684–692. ⟨http://dx.doi.org/10.1109/HICSS.2012.526⟩, 2011.
[88] Z. Genc, F. Heidari, M.A. Oey, S. Van Splunter, F.M.T. Brazier, Intelligent systems for crisis management, Lect. Notes Geoinf. Cartogr. (2013) 349–355. http://dx.doi.org/10.1007/978-3-642-33218-0.
[89] B. Ryan, A. Matheson, Significance of communication in emergency management, Aust. J. Emerg. Manag. 25 (2010) 54–57 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-77954133790 & partnerID=tZOtx3y1⟩.
[90] J. Hernantes, J.M. Torres, A. Lauge, J.M. Sarriegi, I. Starc, E. Zupancic, D. Trcek, Using GMB methodology on a large crisis model, in: Crisis, Information Systems for Crisis Response and Management, ISCRAM, pp. 1–5, ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-84905668369 & partnerID=tZOtx3y1⟩, 2010.
[91] J.O. Gomes, M.R.S. Borges, G.J. Huber, P.V.R. Carvalho, Analysis of the resilience of team performance during a nuclear emergency response exercise, Appl. Ergon. 45 (2014) 780–788. http://dx.doi.org/10.1016/j.apergo.2013.10.009.
[92] S. Van Der Haar, M. Jehn, K. a Segers, Measuring the effectiveness of emergency management teams: scale development and validation, Int. J. Emerg. Manag. 9 (2013) 258–275. http://dx.doi.org/10.1504/IJEM.2013.058547.
[93] L. Alison, N. Power, C. van den Heuvel, M. Humann, M. Palasinksi, J. Crego, Decision inertia: deciding between least worst outcomes in emergency responses to disasters, J. Occup. Organ. Psychol. 88 (2015) 295–321. http://dx.doi.org/10.1111/joop.12108.
[94] J.G. Hodge, T. Lant, J. Arias, M. Jehn, Building evidence for legal decision making in real time: legal triage in public health emergencies, Disaster Med. Public Health Prep. 5 (Suppl. 2) (2011) S242–S251. http://dx.doi.org/10.1001/dmp.2011.57.
[95] Z.E. Franco, N. Zumel, D. Ph, K. Blau, Evaluating the impact of improvisation on the incident command system: a modified single case study using the DDD simulator, in: Proceedings of the 6th International Conference on Information Systems for Crisis Response and Management, ISCRAM. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84905584174 & partnerID=tZOtx3y1⟩, 2009.
[96] P.B. Andreatta, E. Maslowski, S. Petty, W. Shim, M. Marsh, T. Hall, S. Stern, J. Frankel, Virtual reality triage training provides a viable solution for disaster-preparedness, Acad. Emerg. Med. 17 (2010) 870–876. http://dx.doi.org/10.1111/j.1553-2712.2010.00728.x.
[97] J.D. Schenker, S. Goldstein, J. Braun, A. Werner, F. Buccellato, G. Asaeda, D.J. Prezant, Triage accuracy at a multiple casualty incident disaster drill: the emergency medical service, fire department of New York City experience, J. Burn Care Res. 27 (2006) 570–575. http://dx.doi.org/10.1097/01.BCR.0000235450.12988.27.
[98] D.M. Navin, W.J. Sacco, R. Waddell, Operational comparison of the simple triage and rapid treatment method and the sacco triage method in mass casualty exercises, J. Trauma Inj. Infect. Crit. Care 69 (2010) 215–225. http://dx.doi.org/

10.1097/TA.0b013e3181d74ea4.
[99] J.R. Egan, R. Amlôt, Modelling mass casualty decontamination systems informed by field exercise data, Int. J. Environ. Res. Public Health 9 (2012) 3685–3710. http://dx.doi.org/10.3390/ijerph9103685.
[100] H. Carter, J. Drury, G.J. Rubin, R. Williams, R. Amlôt, Public experiences of mass casualty decontamination, Biosecur. Bioterror. 10 (2012) 280–289. http://dx.doi.org/10.1089/bsp.2012.0013.
[101] A. Stergachis, C.M. Wetmore, M. Pennylegion, R.D. Beaton, B.T. Karras, D. Webb, D. Young, M. Loehr, Evaluation of a mass dispensing exercise in a cities readiness initiative setting, Am. J. Heal. Pharm. 64 (2007) 285–293. http://dx.doi.org/10.2146/060289.
[102] A. Khorram-Manesh, C. Angthong, A. Pangma, S. Sulannakarn, R. Burivong, R. Jarayabhand, P. Ortenwall, Hospital evacuation; learning from the past? Flooding of Bangkok 2011, Br. J. Med. Med. Res. 4 (2014) 395–415.
[103] D. Kaspar, An evaluation of the nuclear alarm exercise performed at the Neckarwestheim Nuclear power plant, Knowl. Creat. Diffus. Util. 14 (1997) 20–54 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0019650976&partnerID=tZOtx3y1⟩.
[104] S. Belardo, W.D. Danko, W.A. Wallace, Decision aid for red cross emergency managers, Int. J. Policy Inf. 6 (1982) 11–23 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0020241437&partnerID=tZOtx3y1⟩.
[105] D.M. Bravata, K.M. McDonald, H. Szeto, W.M. Smith, C. Rydzak, D.K. Owens, A conceptual framework for evaluating information technologies and decision support systems for bioterrorism preparedness and response, Med. Decis. Mak. 24 (2004) 192–206. http://dx.doi.org/10.1177/0272989X04263254.
[106] J. Lundberg, E. Törnqvist, S. Nadjm-Tehrani, Resilience in sensemaking and control of emergency response, Int. J. Emerg. Manag. 8 (2012) 99–122.
[107] A. Djalali, M. Castren, V. Hosseinijenab, M. Khatib, G. Ohlen, L. Kurland, Hospital incident command system (HICS) performance in Iran; decision making during disasters, Scand. J. Trauma. Resusc. Emerg. Med. 20 (2012) 14. http://dx.doi.org/10.1186/1757-7241-20-14.
[108] T. Henning, K.T. Perkins, G. Stankovich, Comprehensive Resource Management, in: 2005 International Oil Spill Conference (IOSC 2005), pp. 871–875, ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-33646042512&partnerID=tZOtx3y1⟩, 2005.
[109] L. Lundberg, A. Jonsson, T. Vikström, A. Rüters, Performance indicators for prehospital command and control developed for civilian use tested in a military training setting, a pilot study, J. R. Army Med. Corps. 154 (2008) 236–238.
[110] L. Jian, G. Na, L. Chuangui, J. Longzhe, Study on emergency management system in chemical industry regions. in: International Symposium on Safety Science and Technology, 2008.
[111] J. Trnka, J. Jenvald, A real-time role-playing exercise as a methodology to support command and control research, in: Proceedings of the 3rd International Conference on Information Systems for Crisis Response and Management (ISCRAM 2006), Royal Flemish Academy of Belgium, pp. 559–568. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84905474719&partnerID=tZOtx3y1⟩, 2006.
[112] A. Rüter, T. Vikstrom, Improved staff procedure skills lead to improved management skills: an observational study in an educational setting, Prehosp. Disaster Med. Off. J. Natl. Assoc. EMS Physicians World Assoc. Emerg. Disaster Med. Assoc. Acute Care Found. 24 (2009) 376–379. http://dx.doi.org/10.1017/S1049023X00007172.
[113] C. Franklin, Review of the disaster game, J. Homel. Secur. Emerg. Manag 7 (2010) 59. http://dx.doi.org/10.2202/1547-7355.1786.
[114] Y.F. Chen, D. Watson, E. Miles, Investigating the effectiveness of using immersive virtual world exercises to teach disaster management and development: A comparative study of tools developed by Coventry University, in: Proceedings. 5th International. Disaster Risk Conference. Integr. Risk Manag. - Role Sci. Technol. Pract. IDRC Davos 2014, Global Risk Forum (GRF), 2014: pp. 151–154. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84924937578&partnerID=tZOtx3y1⟩.
[115] J. Rudinsky, E.T. Hvannberg, Communication Interface for Virtual Training of Crisis Management, in: Proceedings of the 10th International ISCRAM Conference. ISCRAM 2013, Karlsruher Institut fur Technologie (KIT), pp. 125–134. ⟨http://www.iscramlive.org/ISCRAM2013/files/136.pdf⟩, 2013.
[116] J.M. Berlin, E.D. Carlström, Collaboration exercises—the lack of collaborative benefits, Int. J. Disaster Risk Sci. 5 (2014) 192–205. http://dx.doi.org/10.1007/s13753-014-0025-2.
[117] M. Armstrong, Rehearsing for the plague: citizens, security, and simulation, Can. Rev. Am. Stud. 42 (2012) 105–120. http://dx.doi.org/10.3138/cras.42.1.105.
[118] L.C. McCormick, L. Hites, J.F. Wakelee, A.C. Rucks, P.M. Ginter, Planning and executing complex large-scale exercises, J. Public Heal. Manag. Pract. 20 (2014) S37–S43. http://dx.doi.org/10.1097/PHH.0000000000000068.
[119] K. Sullivan Faith, B. a Jackson, H. Willis, Text analysis of after action reports to support improved emergency response planning, J. Homel. Secur. Emerg. Manag. 8 (2011) 57. http://dx.doi.org/10.2202/1547-7355.1900.
[120] C.M. Singleton, S. Debastiani, D. Rose, E.B. Kahn, An analysis of root cause identification and continuous quality improvement in public health H1N1 after-action reports, J. Public Heal. Manag. Pract. 20 (2013) 197–204. http://dx.doi.org/10.1097/PHH.0b013e31829ddd21.
[121] K. Lerner, K. Bertram, Common issues in emergency public information as identified in CSEPP exercises, J. Homel. Secur. Emerg. Manag. 11 (2014) 155–168. http://dx.doi.org/10.1515/jhsem-2013-0059.
[122] T.V. Inglesby, R. Grossman, T. O'Toole, A plague on your city: observations from TOPOFF, Clin. Infect. Dis. 32 (2001) 436–445. http://dx.doi.org/10.1097/00019048-200101000-00026.
[123] L.F. Carrel, "Epidemic in Switzerland": description of a strategic leadership exercise by the Swiss government, J. Contingencies Cris. Manag. 13 (2005) 170–175. http://dx.doi.org/10.1111/j.1468-5973.2005.00473.x.
[124] R.J.J. Beerens, B. Kolen, I. Helsloot, EU FloodEx 2009: an analysis of testing international assistance during a worst credible flood scenario in the North Sea area, in: WIT Transactions on Ecology and the Environment, pp. 241–255, ⟨http://dx.doi.org/10.2495/FRIAR100211⟩, 2010.
[125] T. Hudecek, J. Juťanek, J. Pejcoch, Blackout 2014 exercise – Prague, the capital of the Czech Republic, J. Disaster Res. 10 (2015) 270–275 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-84926299990&partnerID=tZOtx3y1⟩.
[126] K. Andress, A postevent smallpox mass vaccination clinic exercise, Disaster Manag. Response 1 (2003) 54–58. http://dx.doi.org/10.1016/S1540-2487(03)00004-X.
[127] G.H. Kramer, D. Beaton, A. Tenney, B. Hauck, K. Capello, D. Quayle, Anatomy of and lessons learned from a Canadian National-Scale Radiological Exercise: ExIT-08, Ultraschall Der Med. 40 (2011) 154–157. http://dx.doi.org/10.1055/s-2004-815600.
[128] D.M. Peterson, R.W. Perry, The impacts of disaster exercises on participants, Disaster Prev. Manag. 8 (1999) 241–255. http://dx.doi.org/10.1108/09653569910283879.
[129] R.W. Perry, Disaster exercise outcomes for professional emergency personnel and citizen volunteers, J. Contingencies Cris. Manag. 12 (2004) 64–75. http://dx.doi.org/10.1111/j.0966-0879.2004.00436.x.
[130] G.H. Larry, Emergency preparedness: tabletop exercise improves readiness, Prof. Saf. 52 (2007) 48 ⟨http://proquest.umi.com/pqdweb?did=1479626221&Fmt=7&clientId=29974&RQT=309&VName=PQD⟩.
[131] B. Matzig, T.O. Solisch, Independent assessor's report on the experience with preparation, managing, observing and evaluation of emergency exercises without prior notice in nuclear power plants, Kerntechnik 76 (2011) 413–417 ⟨http://www.hanser-elibrary.com/doi/abs/10.3139/124.110196⟩ (accessed 06.10.14).
[132] T. Callan, So, you want to run an exercise?, Aust. J. Emerg. Manag. 24 (2009) 59–62 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-70349124668&partnerID=tZOtx3y1⟩.
[133] C. Goodwin, Disaster drills, JACEP 7 (1986) 20–33. http://dx.doi.org/10.1016/S0361-1124(78)80169-5.
[134] K. Allen, E. Lorek, N. Mensia-Joseph, Conducting a school-based mock drill: lessons learned from one community, Biosecur. Bioterror. 6 (2008) 191–201. http://dx.doi.org/10.1089/bsp.2007.0065.
[135] P.D. Biddinger, E. Savoia, S.B. Massin-Short, J. Preston, M. a Stoto, Public health emergency preparedness exercises: lessons learned, Public Health Rep. 125 (2010) 100–106 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-78649674123&partnerID=tZOtx3y1⟩.
[136] D. Balch, C. Taylor, D. Rosenthal, C. Bausch, D. Warner, R. Morris, Shadow Bowl 2003: a collaborative exercise in community readiness, agency cooperation, and medical response, Telemed. J. e-Health 10 (2004) 330–342. http://dx.doi.org/10.1089/tmj.2004.10.330.
[137] D.J. Dausey, M. Moore, Using exercises to improve public health preparedness in Asia, the Middle East and Africa, BMC Res. Notes 7 (2014) 474. http://dx.doi.org/10.1186/1756-0500-7-474.
[138] Anon, Benchmarking emergency preparedness, Chem. Eng. (2000) 18–19 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-0034612490&partnerID=tZOtx3y1⟩.
[139] E. Savoia, F. Agboola, P.D. Biddinger, A conceptual framework to measure systems' performance during emergency preparedness exercises, Int. J. Environ. Res. Public Health 11 (2014) 9712–9722. http://dx.doi.org/10.3390/ijerph110909712.
[140] P.L. Ingrassia, F. Prato, A. Geddo, D. Colombo, M. Tengattini, S. Calligaro, F. La Mura, J.M. Franc, F. Della Corte, Evaluation of medical management during a mass casualty incident exercise: an objective assessment tool to enhance direct observation, J. Emerg. Med. 39 (2010) 629–636. http://dx.doi.org/10.1016/j.jemermed.2009.03.029.
[141] E. Savoia, J. Preston, P.D. Biddinger, 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. 28 (2013) 305–308. http://dx.doi.org/10.1017/S1049023X13000289.
[142] A.H. Kaji, R.J. Lewis, Assessment of the reliability of the johns hopkins/agency for healthcare research and quality hospital disaster drill evaluation tool, Ann. Emerg. Med. 52 (2008) 204–210. http://dx.doi.org/10.1016/j.annemergmed.2007.10.026.
[143] A. Djalali, L. Carenzo, L. Ragazzoni, M. Azzaretto, R. Petrino, F. Della Corte, P.L. Ingrassia, Does hospital disaster preparedness predict response performance during a full-scale exercise? A pilot study, Prehosp. Disaster Med. 29 (2014) 441–447. http://dx.doi.org/10.1017/S1049023X1400082X.
[144] M. Rådestad, H. Nilsson, M. Castrén, L. Svensson, A. Rüter, D. Gryth, Combining performance and outcome indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises, Scand. J. Trauma. Resusc. Emerg. Med. 20 (2012) 58. http://dx.doi.org/10.1186/1757-7241-20-58.
[145] T.L. Thomas, E.B. Hsu, H.K. Kim, S. Colli, G. Arana, G.B. Green, The incident command system in disasters: evaluation methods for a hospital-based exercise, Prehosp. Disaster Med. 20 (2003) 14–23. http://dx.doi.org/10.1017/S1049023X00002090.
[146] H. Sinclair, E.E. Doyle, D.M. Johnston, D. Paton, Assessing emergency management training and exercises, Disaster Prev. Manag. 21 (2012) 507–521. http://dx.doi.org/10.1108/09653561211256198.
[147] H. Kim, Improving simulation exercises in Korea for disaster preparedness,

Disaster Prev. Manag. 22 (2013) 38–47. http://dx.doi.org/10.1108/09653561311301961.
[148] J.-L. Wybo, The role of simulation exercises in the assessment of robustness and resilience of private or public organizations, in: H.J. Pasman, I.A. Kirillov (Eds.), , Resilience of Cities to Terrorist and other Threats (2008) 491–507. http://dx.doi.org/10.1007/978-1-4020-8489-8_23.
[149] M. Latiers, J.M. Jacques, Emergency and crisis exercises: methodology for understanding safety dimensions, Int. J. Emerg. Manag. 6 (2009) 73. http://dx.doi.org/10.1504/IJEM.2009.025174.
[150] H. Lonka, J.L. Wybo, Sharing of experiences: a method to improve usefulness of emergency exercises, Int. J. Emerg. Manag. 2 (2005) 189. http://dx.doi.org/10.1504/IJEM.2005.007359.
[151] H. Crankshaw Donald, D. Fenton Roland, Evaluation criteria for emergency plan drills and exercises, Trans. Am. Nucl. Soc. (1983).
[152] J. Wybo, M. De Paris, P.O. Box, The role of simulation exercises in the assessment of robustness and resilience of private or public organisations, 2008.
[153] A. Duarte, C. Duarte, J.O. Gomes, P. Victor, R. De Carvalho, ASC Model: a process model for the evaluation of simulated field exercises in the emergency domain, in:Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2013), Karlsruher Institut fur Technologie (KIT), pp. 551–555, ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84905671426 & partnerID=tZOtx3y1⟩, 2013.
[154] E. Heumüller, S. Richter, U. Lechner, Towards a framework for command post exercises, in: Proceedings of the 9th International. ISCRAM Conference. – Vancouver, Canada, April 2012, Simon Fraser University, pp. 1–6. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-84905586566 & partnerID=tZOtx3y1⟩, 2012.
[155] C. Villoria, R. Pérez, Oil spill decision support system for the Venezuelan oil industry, in: 2005 International Oil Spill Conference (IOSC 2005), pp. 5546–5548. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-33646062583 & partnerID=tZOtx3y1⟩, 2005.
[156] D. Steward, T.T.H. Wan, The role of simulation and modeling in disaster management, J. Med. Syst. 31 (2007) 125–130. http://dx.doi.org/10.1007/s10916-006-9047-7.
[157] K.M. Gebbie, J. Valas, J. Merrill, S. Morse, M.P.H. Rn, Role of exercises and drills in the evaluation of public health in emergency response, Prehosp. Disaster Med. 21 (2006) 173–182. http://dx.doi.org/10.1017/S1049023X00003642.
[158] J.D. Quaranta, H.M. Childers, P. Myles, Lessons learned from tabletop reviews of emergency action plans for high hazard dams in West Virginia, USA, in: Proceedings of the 12th International Conference Tailings Mine Waste'08, pp. 325–335. ⟨http://www.scopus.com/inward/record.url?Eid=2-s2.0-79952289343 & partnerID=tZOtx3y1⟩, 2009.
[159] D. Johnston, R. Tarrant, K. Tipler, M. Coomer, S. Pedersen, R. Garside, Preparing schools for future earthquakes in New Zealand: lessons from an evaluation of a Wellington school exercise, Aust. J. Emerg. Manag. 26 (2011) 24–30 ⟨http://www.scopus.com/inward/record.url?eid=2-s2.0-79952063365 & partnerID=tZOtx3y1⟩.
[160] G.B. Green, S. Modi, K. Lunney, T.L. Thomas, Generic evaluation methods for disaster drills in developing countries, Ann. Emerg. Med. 41 (2003) 689–699. http://dx.doi.org/10.1067/mem.2003.147.
[161] S.T. March, G.F. Smith, Design and natural science research on information technology, Decis. Support Syst. 15 (1995) 251–266. http://dx.doi.org/10.1016/0167-9236(94)00041-2.
[162] J.E. van Aken, Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules, J. Manag. Stud. 41 (2004) 219–246. http://dx.doi.org/10.1111/j.1467-6486.2004.00430.x.
[163] H.A. Simon, The Sciences of the Artificial, MIT Press, 1996 ⟨http://www.google.com/books?id=k5Sr0nFw7psC⟩.
[164] S. Gregor, D. Jones, The anatomy of a design theory, J. Assoc. Inf. Syst. 8 (2007) 312–335 ⟨http://www.layrib.com/DL/gregor_07_design_theory.pdf⟩.
[165] D. Denyer, D. Tranfield, Using qualitative research synthesis to build an actionable knowledge base, Manag. Decis. 44 (2006) 213–227. http://dx.doi.org/10.1108/00251740610650201.
[166] D. Denyer, D. Tranfield, J.E. van Aken, Developing design propositions through research synthesis, Organ. Stud. 29 (2008) 393–413. http://dx.doi.org/10.1177/0170840607088020.
[167] J.E. van Aken, A.G.L. Romme, Reinventing the future: adding design science to the repertoire of organization and management studies, Organ. Manag. J. 6 (2009) 2–12. http://dx.doi.org/10.1057/omj.2009.1.
[168] S.A. Carlsson, Developing knowledge through IS design science research, Scand. J. Inf. Syst. 19 (2007) 75–86.
[169] B. Brehmer, Understanding the functions of C2 Is the key to progress, Int. C2 J. 1 (2007) 211–232.
[170] A. Cedergren, H. Tehler, Studying risk governance using a design perspective, Saf. Sci. 68 (2014) 89–98. http://dx.doi.org/10.1016/j.ssci.2014.03.006.
[171] C. Argyris, Actionable knowledge: design causality in the service of consequential theory, J. Appl. Behav. Sci. 32 (1996) 390–406. http://dx.doi.org/10.1177/0021886396324004.
[172] J.E. van Aken, Management research as a design science: articulating the research products of mode 2 knowledge production in management, Br. J. Manag 16 (2005) 19–36. http://dx.doi.org/10.1111/j.1467-8551.2005.00437.x.
[173] M.K. Lindell, R.W. Perry, Evaluation criteria for emergency response plans in radiological transportation, J. Hazard. Mater. 3 (1980) 335–348. http://dx.doi.org/10.1016/0304-3894(80)80005-7.
[174] D. Alexander, Towards the development of a standard in emergency planning, Disaster Prev. Manag. 14 (2005) 158–175.
[175] E.L. Quarantelli, Ten criteria for evaluating the management of community disasters, Disasters 21 (1997) 39–56. http://dx.doi.org/10.1111/1467-7717.00043.
[176] J. Borell, K. Eriksson, Improving emergency response capability: an approach for strengthening learning from emergency response evaluations, Int. J. Emerg. Manag. 5 (2008) 324. http://dx.doi.org/10.1504/IJEM.2008.025101.
[177] M. Abrahamsson, H. Hassel, H. Tehler, Towards a system-oriented framework for analysing and evaluating emergency response, J. Contingencies Cris. Manag. 18 (2010) 14–25. http://dx.doi.org/10.1111/j.1468-5973.2009.00601.x.
[178] R. Heath, Looking for answers: suggestions for improving how we evaluate crisis management, Saf. Sci. 30 (1998) 151–163. http://dx.doi.org/10.1016/S0925-7535(98)00043-5.
[179] D.E. Alexander, Evaluating of civil protection programmes, with a case study from Mexico, Disaster Prev. Manag. 24 (2015) 263–283 ⟨https://login.ezproxy.net.ucf.edu/login?auth=shibb & url=http://search.ebscohost.com/login.aspx?direct=true & db=edb & AN=101529285 & site=eds-live & scope=site⟩.
[180] C. Nelson, N. Lurie, J. Wasserman, Assessing public health emergency preparedness: concepts, tools, and challenges, Annu. Rev. Public Health 28 (2007) 1–18. http://dx.doi.org/10.1146/annurev.publhealth.28.021406.144054.
[181] E. Savoia, F. Agboola, P.D. Biddinger, Use of after action reports (AARs) to promote organizational and systems learning in emergency preparedness, Int. J. Environ. Res. Public Health 9 (2012) 2949–2963. http://dx.doi.org/10.3390/ijerph9082949.