This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 832981

Deliverable D6.2

Scenario Development and Specifications of the Evaluation Methodology

Due date of deliverable: 31/07/2020

Actual submission date: 12/03/2021

Charlotte Hall*[1], William Nicholson*[1], Holly Carter[1], Frank Long[2], Tony Godwin[3], Irina Marsh[3] Richard Amlot[1] and Dale Weston[1] (* joint first authors)

1: PHE 2: Hampshire Fire and Rescue Service and Imperial College London 3: CBRNE Ltd

© Copyright 2021 PROACTIVE Project (project funded by the European Commission). All rights reserved.

No part of this document may be copied, reproduced, disclosed or distributed by any means whatsoever, including electronic without the express permission of the International Union of Railways (UIC), Coordinator of PROACTIVE Project. The same applies for translation, adaptation or transformation, arrangement or reproduction by any method or procedure whatsoever.

The document reflects only the author's views and the Commission will not be liable of any use that may be made of the information contained therein. The use of the content provided is at the sole risk of the user.

Project details

Project acronym	PROACTIVE
Project full title	**PR**eparedness against CBRNE threats through c**O**mmon **A**pproaches between security pra**CTI**tioners and the **V**ulneranbl**E** civil society
Grant Agreement no.	832981
Call ID and Topic	H2020-SU-SEC-2018, Topic SU-FCT01-2018
Project Timeframe	01/05/2019 – 30/04/2022
Duration	36 Months
Coordinator	UIC – Grigore Havarneanu (havarneanu@uic.org)

Document details

Title	Scenario Development and Specifications of the Evaluation Methodology
Work Package	WP6
Date of the document	12/03/2021
Version of the document	05
Responsible Partner	PHE
Reviewing Partner	ETICAS, WMP, FFI, UIC
Status of the document	Final
Dissemination level	Public

Document history

Revision	Date	Description
01	09/07/2020	First Draft
02	14/07/2020	Reviewed by ETICAS and WMP
03	21/07/2020	Reviewed by FFI and UIC
04	28/07/2020	Reviewed and approved by PEO
05	12/03/2020	Update following mid-term periodic review

Consortium – List of partners

Partner no.	Short name	Name	Country
1	UIC	UNION INTERNATIONALE DES CHEMINS DE FER (COORDINATOR)	France
2	CBRNE	CBRNE LTD	UK
3	PPI	POPULATION PROTECTION INSTITUTE (MINISTRY OF THE INTERIOR OF THE CZECH REPUBLIC)	Czech Republic
4	DB	DEUTSCHE BAHN AG	Germany
6	UMU	UMEA UNIVERSITET	Sweden
7	DHPOL	DEUTSCHE HOCHSCHULE DER POLIZEI	Germany
8	RINISOFT	RINISOFT LTD	Bulgaria
9	WMP	WEST MIDLANDS POLICE AND CRIME COMMISSIONER	UK
10	ETICAS	ETICAS RESEARCH AND CONSULTING SL	Spain
11	SESU	STATE EMERGENCY SERVICE OF UKRAINE	Ukraine
12	PHE	DEPARTMENT OF HEALTH	UK
13	SPL	STATE POLICE OF LATVIA	Latvia
14	AGS	AN GARDA SÍOCHÁNA – NATIONAL POLICE FORCE IRELAND	Ireland
15	FFI	FORSVARETS FORSKNINGSINSTITUTT	Norway
16	NPH	KOMENDA GŁÓWNA POLICJI	Poland

Executive summary

This deliverable represents the output from two parallel activities conducted as part of PROACTIVE Task 6.2: scenario development and specification of the evaluation methodology.

First, through a process of iterative review and stakeholder engagement, we have developed an initial prototype scenario, complete with communication strategies representing both poor and optimised communication. Alongside this prototype scenario, we have also provided a series of recommendations for key components that should be included in future scenario development discussions with eNOTICE to ensure that: a) the requirements of PROACTIVE are met; b) the exercises represent a reasonable worst-case scenario, thus demonstrating maximum value for participating practitioners, and; c) opportunities to maximise learning from the exercises are taken. The prototype scenario and these recommendations are intended for use to inform ongoing discussions and scenario development between PROACTIVE and eNOTICE both in relation to the rescheduled Rieti exercise and future exercises.

Second, following a process of rapid evidence review and synthesis with subject matter expertise, we have developed an initial plan for exercise evaluation. For each exercise, we will:

- carry out a hot debrief (which would take place immediately after the exercise) to capture the views of emergency responders, exercise planners, and project partners;
- use pre- and post-exercise casualty feedback questionnaires to identify any changes in casualty volunteers' perceptions as a result of taking part in the exercise;
- observe exercise play to identify any challenges that arise (particularly in relation to the management of vulnerable groups);
- and carry out focus groups with casualty volunteers to facilitate an in-depth understanding of their experiences during the exercise and to identify any areas for improvement.

Depending on the exercise scenario, consideration will also be given to using physical quantitative measures (e.g. timing data, measures of efficacy) to evaluate the exercise. In addition to this initial specification, a detailed evaluation plan will be developed for each of the PROACTIVE exercises, and the evaluation methods used will be tailored to each specific exercise scenario.

Acronym Table

Acronym	Definition
CBRNe	Chemical, Biological, Radiological, Nuclear and explosive
D	Deliverable
eNOTICE	European Network Of CBRN TraIning CEnters
EU	European Union
LEA	Law Enforcement Agencies
N	Number
P	Participant
PM	Project Meeting
PSAB	Practitioner Stakeholder Advisory Board
SOP	Standard Operating Procedure
WN	William Nicholson
WP	Work Package

Table of Contents

1. INTRODUCTION

As per the PROACTIVE grant agreement, Task 6.2 was originally envisaged as encompassing two complementary activities of critical importance for exercise planning and evaluation.

First, the intention was for Task 6.2 to use an iterative process of scenario development (encompassing three distinct steps) to specify the scenarios that will be deployed as part of the exercises carried out within Tasks 6.3, 6.4, and 6.5. These three initial steps involved a process of scenario drafting and review, incorporating feedback and input from LEA partners and the PSAB, to result in the development of several well-developed scenarios for use in the exercises. Following the successful funding of the PROACTIVE project, the process of exercise planning has developed into a productive and successful collaboration between PROACTIVE and eNOTICE (European Network Of CBRN TraIning CEnters, 2020), which is an EU funded project that aims to enhance preparedness, resilience and response to CBRN incidents through stakeholder and practitioner interactions. The successful development of this collaboration has also meant that the responsibility for developing scenarios for the exercises is not the sole purview of the PROACTIVE project team, but is instead the result of a partnership between PROACTIVE, eNOTICE, and exercise organisers. Furthermore, the impact of the COVID-19 pandemic on emergency response organisations across the EU has meant that the Rieti exercise and associated planning has been extended into 2022. There are, therefore, further cross-project discussions and decisions required regarding the set-up of the Rieti exercise, and an associated opportunity to further refine the scenario and injects in collaboration with eNOTICE over the coming months.

When considered together, these two impacts have necessitated a slight change to the specific focus of the scenario development component of Task 6.2. Specifically, this deliverable will now present the process of iterative scenario development and review (incorporating input from practitioners and stakeholders, including the PROACTIVE PSAB), resulting in the specification of one broad template scenario, tailored for the Rieti exercise in the first instance, alongside a list of key factors that are important considerations for future scenario development. This template scenario and associated recommendations have been developed to reflect the needs of the PROACTIVE project (based on PSAB input and recommendations resulting from PROACTIVE Work Package 1) and are intended for use in collaboration with eNOTICE to inform the development of specific scenarios and related injects for each exercise. In this way, the scenarios for each exercise will be able to reflect both: a) the requirements of PROACTIVE and eNOTICE, and; b) the specific context of each future exercise.

Second, the task aimed to identify the most appropriate method for evaluating exercise outcomes, drawing on both existing research and the deliverable authors' previous experiences of exercise evaluation. To this end, this deliverable includes the outputs from a rapid evidence review of published literature, synthesised alongside the authors' expert opinion based on considerable experience evaluating exercises. While the deliverable provides an overview of the broad evaluation methodology that will be used to evaluate the joint PROACTIVE/ eNOTICE exercises, the specific evaluation of each exercise will be further developed in a bespoke fashion to reflect the precise requirements of each exercise (e.g., in terms of participants, context, content, etc.).

In the sections that follow, we first recap the aims, objectives and methodology for the exercises (detailed in WP6.1), before moving to detail the scenario development process and outcomes ,

followed by the rapid review of evaluation methodologies and subsequent synthesis with expert views.

2. PROACTIVE/ ENOTICE EXERCISE AIMS, OBJECTIVES, ETHICS, AND METHODOLOGY

As detailed in D6.1 (The PROACTIVE Methodology for the Field Exercises), PROACTIVE are working in conjunction with H2020 Project eNOTICE to co-ordinate, plan, and execute three field exercises across different countries. Although the ongoing COVID-19 pandemic requires flexibility in the planning and conduct of these exercises, the following strategic objective was agreed between PROACTIVE and eNOTICE in order to ensure coordination between the two projects:

> *"In partnership with eNOTICE evaluate the effectiveness of responses to a CBRNe incident focusing on harmonising of procedures and tools that support the needs of civil society, including those citizens that are vulnerable"*

Furthermore, the following tactical objectives have been identified by PROACTIVE in order to outline the key learnings that should come from the programme of exercises (Table 1). While they are still under review and subject to modification as the project advances, they clearly demonstrate how the exercises and scenarios are taking into account outputs from other work packages.

Table 1 Tactical Objectives for PROACTIVE eNOTICE exercises as of January 2020

No.	Objective
1	To benchmark current practices against the recommendations from WP1, D3.1
2	To test the effectiveness of the PROACTIVE App in supporting the needs of Civil Society
3	Test if appropriate consideration is given to delivering policy and procedures to assist those with mobility issues (e.g. animals and mobility aids) during CBRNE incidents
4	To further understand the needs of different vulnerable groups during CBRNe incidents
5	To test that messages are pitched at an appropriate level in terms of language and complexity

6	To test the effectiveness of pre-planning and pre-incident information during emergency communication with the public
7	To test the provision and suitability of public health messaging in an emergency
8	To gain an understanding of the additional requirements created through involving Civil Society in CBRNe exercises
9	To test if pre-planned information provided by authorities has been deployed in a consistent way and has been understood
10	To gain an understanding of factors that may increase public compliance during CBRNe incidents

These objectives are aligned not only to the PROACTIVE project Description of Action, and the requirements of eNOTICE, but also the outputs from Work Package 1 (specifically, the identified gaps/recommendations for best-practice detailed in D1.3), Work Package 2 (specifically, the requirements identified by members of the Practitioner Stakeholder Advisory Board), and Work Package 3 (specifically, the needs identified by the Civil Society Advisory Board). They are also designed to provide a platform for examining the utility and functionality of the PROACTIVE tools developed across Work Packages 4 and 5.

Although relatively high level, these objectives are being operationalised for each exercise through the use of the IIMARCH framework. Further detail is presented in D6.1, however, in brief, the IIMARCH framework stands for:

Table 2 IIMARCH Framework

I	Information
I	Intention
M	Method
A	Administration
R	Risk Assessment
C	Communication
H	Human Rights, Legal and Ethical

Through the use of this standardised methodology, Work Package 6 will ensure that each exercise has a clear and detailed plan, developed in conjunction with eNOTICE colleagues, which is: a) consistent with the human rights and legal requirements of the project, and; b) flexible and reactive enough to evolve (given the flexibility required in organising large scale emergency preparedness exercises, and particularly due to changes enforced by the COVID-19 pandemic) right up to the conduct and evaluation of the exercises.

As recommended by D8.1, *Legal and Ethical State-of-the-Art on CBRNe preparedness and response* (section 3 Ethical Framework), during the scenario development activities, the ethical standards that underline the response phase of CBRN emergencies have been considered; we list here in brief the standards that are relevant (EUR-OPA, Resolution 2011-1):

- Humanitarian assistance: all persons receive immediate assistance, including the benefit of basic health services. Humanitarian assistance is provided fairly, impartially and without discrimination, showing due regard for the vulnerability of victims and for individuals' and groups' specific needs.
- Information and communication during disasters: all persons, local and regional authorities and non-governmental organisations affected by disasters are informed of and are entitled to participate in making decisions in response to disasters. They receive, in their own language, easily understandable information about the nature and extent of the disaster, the emergency measures planned in response to it, the times and places at which food and drink will be distributed, the location of emergency medical facilities, temporary housing arrangements and the arrangements for and destination of any population movements that are planned.
- Compulsory evacuation of population: compulsory evacuation can only take place if a clear explanation has been given of the potential risks involved in the case of non-evacuation.
- Respect of dignity: the dignity of all persons who are victims is respected, particularly concerning his/her security, physical safety, access to food and clean water, hygiene, temporary housing, clothing and if necessary essential emergency medical and psychological care
- Respect of persons: personal rights are respected, particularly the right to one's own image and the right to privacy, so that the presence of the media does not result in abuses
- Emergency assistance for the most vulnerable persons: allowing for local circumstances and without prejudice to the priority assistance to be given to all who have a chance of survival, priority for humanitarian assistance, first aid and emergency evacuations go in priority to the most vulnerable people, such as pregnant women, children, people with disabilities, elderly people, the ill and the wounded.
- The importance of rescue workers: Irrespective of their nationality, theirs status or their function and regardless of the seriousness and nature of the disaster, both civilian and military rescue workers, including any private security forces, behave with dignity, keep their anxiety of fear under control, keep calm and ensure that they never infringe the fundamental rights of the people they are rescuing.

From the point of view of research ethics (see D8.3 *Materials and briefing for PROACTIVE exercises*), given the involvement of human volunteers in the exercise programme, it is critically important to ensure that there are robust legal and ethical provisions in place. The IIMARCH framework allows for the robust assessment of ethical and legal issues tailored to each exercise (see

D6.1) and through the work of Work Package 8. More detail is provided in D8.3 and D6.1, but in short, the following key points form the basis for the PROACTIVE ethical assessment:

- All participants in PROACTIVE field exercises will provide informed consent to participate in the exercises;
- All participants will be briefed in person prior to the start of a field exercise in order to allow for any questions and make sure all roles are clear;
- Welfare support will be provided (and clearly signposted) to all participants;
- All participants will have the right to withdraw themselves (and/or any personal data) from the field exercises at any time;
- An Ethics and Data Protection Supervisor (supported by the External Ethics Advisory Board) has been appointed to ensure that all exercises are conducted in an ethically compliant fashion.

Through tailoring these broad principles to each specific field exercise, PROACTIVE will be able to anticipate the vast majority of potential ethical and legal issues, and will be able to react in a rapid and decisive way to tackle any unforeseen issues that may arise during exercise planning/ implementation. In particular, despite setting out a detailed ethical recruitment strategy (D10.1: H - Requirement No 1] there is still a risk that fieldwork participants (in particular volunteers representing vulnerable groups) might be harmed or disrespected during the field exercises. Therefore, the PROACTIVE consortium will carry out an incidental finding analysis with the involvement of external ethics experts. The PEO will share a summary of the environment and safety conditions defined in WP6 for each exercise scenario together with the categories (citizens, end-users, etc.) and relevant characteristics (age, gender, vulnerability, etc.) corresponding to the final list recruited participants with the External Ethics Advisory Board (EEAB). Feedback will be requested regarding possible non-anticipated risks so that mitigation measures can be put in place before the field exercises. Outcomes of this collaborative assessment will be integrated into the exercise plans developed within WP6. As such, explicit provisions in the event of complications during the field exercises will be provided before the running of the field exercises.

Overall, the exercises are designed to:

- benchmark current practice against the best practice identified through the PROACTIVE project (e.g., relating to communication with the public, incident management, etc.);
- examine the utility of best practice tools (e.g., Apps for communication, clear pre-incident information);
- ensure the acceptability and accessibility of these tools based on the criteria established in D8.2 (e.g., disability inclusion, ease of use and perception of data security). D8.2 operationalised data protection and acceptability requirements into targeted recommendations for registered users, end-users and policymakers. Based on these outcomes, certain operational and societal variables will be assessed during the PROACTIVE validation process. Issues to be examined include how PROACTIVE technologies are used throughout emergency scenarios for transmitting information between the above stakeholders, how these actors perceive technological mediation and how personal data is protected within these interactions. Participant observation and questionnaires used with this purpose will follow responsible research criteria described in D8.3 and D6.1. In this way, the project will ensure ethics and privacy by design. Results and

guidelines derived from these analyses will be reflected in D8.4, “Ethical and societal assessment of PROACTIVE outputs”.

The exercises will build iteratively upon one another, first testing an Initial Operational Response scenario, before testing a Specialist Operational Response scenario, before drawing these together for a combined scenario to be deployed during the large-scale final exercise. Furthermore, each exercise will be thoroughly evaluated, with learning from previous exercises informing the revision of PROACTIVE materials designed to meet the recommendations and fill the gaps identified through Work Packages 1-5 (e.g., the provision of pre-incident information, the use of an app/ platform, methods of effective communication with the public, etc.). The specific plans, scenarios, and resulting evaluation of each exercise (with recommendations and iterations to be incorporated into the subsequent exercise) will be outlined in the relevant deliverable for each exercise.

The scenario development process (detailed in the next section) aimed to reveal a list of key components for testing and examination across the PROACTIVE exercises, with an emphasis on optimising communication given the central role of communication in compliance detailed through Work Package 1. As noted in the Introduction, the scenarios themselves need to be co-developed for each exercise with eNOTICE and the local response partners, we therefore focus on developing a high-level template scenario that can be used to inform discussions around the scenario for each exercise.

3. SCENARIO DEVELOPMENT

The iterative process of scenario development whilst actively engaging Practitioners, EU LEA’s and Policy Makers was carried out using a revised four-step process, based on the three step process initially proposed.

3.1. SCENARIO DEVELOPMENT PROCESS - STEP 1

Alongside the human factors analysis conducted as part of Work Package 1 (and reported in Deliverables 1.1 and 1.2) a stakeholder engagement activity, parallel to the PROACTIVE project, was held with experts from emergency service, health and Government organisations. This engagement activity took the form of a one-day workshop, hosted and facilitated by Frank Long (PhD student at Imperial College London), and involved expert input from PROACTIVE Work Package 1. During the workshop, experts were asked to discuss factors that might result in the best and worst case CBRNe scenarios in terms of the impact of the scenario on public behaviour. This focus was commensurate with the PROACTIVE project aim to “enhance societal CBRN preparedness by increasing Practitioner effectiveness in managing large, diverse groups of people in a CBRN environment”. Several variables which could affect scenario severity were identified. These included: weather, dependants, and type of communication provided by responders (a full list of identified variables are presented in Table 3).

Table 3 Variables which can impact scenario severity

Variable	Best Case (Easiest to manage)	Worst Case (Hardest to manage)
Location	No consensus	Enclosed and Unfamiliar
Communication by Responders	Good clear and consistent with casualties understanding the situation	No communication
Public Awareness and Knowledge	Good awareness and Knowledge	No Awareness or Knowledge
Agent Effect	No-Effect or Severe Reaction	Mild Painful Reaction
Weather Conditions	Dry and warm	Wet and cold
Dependents	None involved	Dependants involved
Confidence in Responders	High	Low
Crowd and other casualties actions	Remain	Leave

These variables were then used to develop five detailed scenarios. Experts from the stakeholder workshop ordered these scenarios in terms of their severity (i.e. 'best case' to 'worst case'; these can be found in Appendix A). It is important to note that these factors and the associated scenarios were constructed by subject matter experts (including PROACTIVE expert input) during the independent workshop. They are not, therefore, a full and systematic account of all possible factors. There are, therefore, other factors that were not identified and are not included in these scenarios that are being considered in the context of the PROACTIVE project (for example, multi-site or consecutive incidents). These are briefly discussed in the PROACTIVE/eNOTICE Scenario Development subsection.

3.2. SCENARIO DEVELOPMENT PROCESS - STEP 2

The scenarios developed as part of the workshop were shared with the PSAB using an online survey to ensure active engagement of Practitioners, EU LEAs and Policy Makers in making the scenarios fit for purpose within the PROACTIVE project.

As part of the survey, PSAB members were asked to rank how difficult the scenario would be to manage using a five-point Likert scale (with 1 representing not at all difficult, and 5 representing very difficult), and to provide general thoughts about the scenario. Fifteen PSAB members participated in the online survey and provided feedback on the scenarios. Outcomes from this review process were discussed as part of the Pre-Exercise Workshop held on the 19th of March 2020.

The quantitative survey feedback provided by PSAB members is displayed in Table 4. The scenario that was ranked hardest to manage was consistent across the engagement activity carried out in step 1 (with members from emergency services, health and Government organisations) and the PSAB (mean score of 4, $n = 13$).

Table 4 Feedback from the PSAB regarding how difficult each scenario would be to manage from a responder perspective

	Workshop	PSAB		
Scenario	**Ranking**	**Ranking**	**Average**	**Responses (n)**
It's midday on a sunny July day and you are doing your weekly food shop with your family at your local supermarket. Suddenly, there is a loud crash and you find yourself covered in a white powder. As you look around you, your eyes start to sting slightly, and you realise there is a strange taste and smell in the air. A few people around you seem to have started rubbing their eyes. But like you, they are standing still. No one is leaving the area. As you stand there, the emergency responders begin to arrive. They are quick to come and talk to you and the other people, giving you clear instructions and explanations.	Best Case (1)	2	2.94	16
It's 9am on a warm and dry summer day and you are wandering alone down your local high street where you often shop. You're in no hurry and are enjoying the walk. Suddenly	Not Worst (2)	1	2.38	13

there is a loud explosion. Looking down you see you are covered in a white powder. The air around you smells and tastes slightly strange. But you feel fine. The people around you all seem to have stopped and are staying put as the emergency services start to arrive. As you stand there with the others who have been part of the incident, the emergency responders begin to communicate with you all, explaining what is going on and helping you understand what will happen next as they help you.				
It's early afternoon on a cloudy autumn day and you and a friend are walking into the centre of a town you've been to a few times before. As you're walking, there's a loud bang/explosion and you look down to find yourself covered in a White powder. Very quickly your eyes start to sting slightly. Looking around you can see other people who are covered in the powder and are starting to rub their eyes. Some of them are hurrying away from the area, but others are staying put as the emergency responders arrive. Their communication doesn't really seem to be either good or bad.	Middle (3)	4	3.79	14
It's lunch time on a frosty day and you're on your way to meet a family member for lunch. As you're walking through a shopping centre that you've never been to before, an explosion occurs. Looking down you find you are covered in white powder. You don't feel any different. The powder doesn't seem to be affecting you or the people around you. But many of them have started to leave the shopping centre. As you're watching, you see that the emergency services have started to arrive. They don't seem to be saying much to you or the others about what is going on. The emergency responders aren't helping you understand what's happening or what will happen next.	Very Bad (4)	3	3.71	14

On a cold, wet late evening, you are travelling alone to meet a family member. As you are stood alone on the platform of an unfamiliar train station, an explosion occurs. Ears ringing, you look down to find that you are covered in a white powder. Quickly you feel your eyes start to sting and a cough rises in your throat. It is feeling increasingly difficult to breathe and your skin has started to burn. All around you, you can see your fellow passengers starting to suffer. Some have collapsed. Many have started to head for the exits to leave the station. As you're standing there, the first emergency responders begin to arrive. They're not really saying anything. They're just moving people around.	Worst Case (5)	5	4	13

PSAB members were asked to provide their general thoughts about each scenario, and the qualitative feedback provided covered a wide range of topics. Key themes arising from PSAB feedback (and illustrative examples) regarding each scenario can be found in Table 5.

Table 5 Themes arising from PSAB qualitative feedback on initial proposed scenarios

Scenario	Theme	Example Quotes
Best Case	Unrealistic	*"The scenario seems very simple, as I do not believe that after such an incident most of the people would be too stunned to act. Probably most of them will start running in panic."* *"That would be easy to manage but not realistic (people would panic and leave the area before our arriving), depend on the number of people"* *"Not sure that the people will stay still or will not move for a long time except if the situation hinders any move which may rather create some panic."* *"Normally Firefighters/Police would arrive within 10 minutes. Is it possible that in 10 minutes people are still standing inside the shop and not trying to go home or just run in the street?"*

Not Worst	Communication	*"Communicate as soon as possible with correct information and orders will be the most difficult for first responders but for me the best way to keep the people calm and to manage the situation as easy as possible. It will also help to evacuate the people orderly and to make it more easy during the decontamination procedure."* *"Effective comms* [sp] *- although depending on numbers there may be people that can't hear or understand directions from emergency services"*
	Compliance	*"People would be collaborative"* *"All affected get information and treatment."*
Middle	Challenges managing public movement	*"Open spaces are difficult to manage the people."* *"maybe the people running away would add some challenge (for the first responder?)"* *"Outside in fresh air;2. Difficulties in managing contamination spread;3. Comms* [sp] *appear to be having some effect on casualties remaining in the local area;4. Difficulties in creating a cordoned off area to keep those affected in one place;5. Warm day beneficial for wet decon* [sp]*;"*
Very Bad	Realistic	*"That would be more realistic but the weather could be a real obstacle for the decontamination (but that's why we need some training, for that kind of challenge) and the people who has leaving would be another issue (but not for first responders?)"* *"This scenario approach more the reality. There will always be people/victims who will have left the place."*
	Panic	*"People's fear and uncertainty may cause panic."* *"After the explosion people will be in panic and will run around to evacuate."*
	Public leaving the scene	*"Without interaction seeking form the emergency services the possible CBRNe recipients may spread around the city and won't find necessary help on time."* *"the scenario rises* [sp] *the question to me of security perimeter and buffer zones to put in place so that the people leaving the shopping*

		center [sp] *don't spread to much in town and to get a diagnosis to identify this white powder and to be decontaminated and/or treated."*
Worst Case	Realistic	*"Much more the reality, especially when there are more people involved. It will be very difficult to know what happened."* *"It's realistic, First responders should secure triage and evacuate people to a CP."*
	Difficult to manage	*"That would be not easy to manage because of the weather (decontamination) and the severity of the situation but that's more realistic than the scenario before"* *"Incident exacerbated by panicking members of public;2. Crush injuries could potentially be worse than the white powder;3. P1, P2 and P3 casualties requiring treatment;4. Clinical wet decon* [sp] *as well as Mass Decon* [sp] *required;5. Inability to provide effective communication. 6. Difficulties in managing spread of contamination;7. Impact on transport network and resulting effect on the local area;8. Impact on local A&E's;9. Cold - impacts on undertaking effective wet decon;10. Dark/evening - assumption that this is after hours when businesses are closed preventing use of local facilities/amenities (water/blankets/shelter etc)."*
	Communication	*"Those affected are not communicated to, and not even contained."*
		"Information to be given to the people involved in this event seem to be necessary for them to understand if the situation is under control or not and if the instructions (Moving people around) have a rationale (for evacuation, for a medical care, etc..)"

Given the relevance of the railway scenario to the Rieti exercise and the importance of planning for realistic worst-case scenarios, the 'worst case' scenario was deemed the most appropriate to be carried forward for further development. This reasonable worse-case scenario provides an excellent opportunity for learning and development within a realistic context. PSAB members feedback on the 'worst case' scenario was centred around being realistic (e.g. "Much more the reality, especially when there are more people involved. It will be very difficult to know what happened."), being difficult to manage from a responder point of view (e.g. "That would be not easy to manage because of the weather (decontamination) and the severity of the situation but that's more realistic than the scenario before") and the importance of communication (e.g. "Information to be given to the people involved in this event seem to be necessary for them to understand if the situation is under control or not and if the instructions (Moving people around) have a rationale (for evacuation, for a medical care, etc..)").

As communication was a key theme that arose from the feedback provided, the scenario was further developed to provide different communication strategies that could be deployed and examined within an exercise context. Specifically, two additional paragraphs were developed to add on to the

reasonable worst-case scenario (displayed in Table 2): a 'poor communication strategy' (displayed in Figure 1) and an 'optimised communication strategy' (displayed in Figure 2). These revised scenarios were presented to the PSAB during the Pre-Exercise Workshop.

3.3. SCENARIO DEVELOPMENT PROCESS - STEP 3

As part of step three, the revised scenarios were workshopped with the PSAB as part of a PROACTIVE Pre-Exercise Workshop. The Pre-Exercise Workshop took place on the 19th of March and PSAB representatives from 18 organisations attended (additional details can be found in Table 6). The session focussing on scenario development was led by PHE and is detailed below. Additional topics covered at the Pre-Exercise Workshop are documented in D2.2.

Table 6 Pre-Exercise Workshop PSAB Attendees

Organisation	Type	Country
VIA Rail Canada	Rail Expert	Canada
National Public Health Centre	CBRN Specialist	Lithuania
National Fire Chiefs Council (NFCC) National Resilience - CBRN Capability	LEA	UK
Hellenic Ministry of National Defence	LEA	Greece
INERIS (National institute for Industrial Environment and Risks)	Rail Experts	France
Polish State Railways (PKP SA)	Rail Experts	Poland
Polskie Koleje PaÅ stwowe S.A.	Rail Experts	Poland
National Resilience	CBRN Specialist	UK
CNVVF - Italian Firefighters Corp	LEA	Italy
I4-Flame OU (LLC)	LEA	Estonia
Fire Department of Dortmund	Exercise Leader	Germany
ENEA	Project manager	Italy
Universita Cattolica del Sacro Cuore	Medical Responders	Italy
TH Köln	CBRN Specialist	Germany
Spanish National Police	LEA	Spain
University (IHU-DIPAE)	Scientific/Technical Officer - MAG of PROACTIVE project	Greece
SAFE	CBRN Specialist	Italy
Europol	LEA	Netherlands

The session opened with a description of how the baseline scenarios (which can be found in Appendix A and detailed in section 2.2) were created and a description of the PSAB feedback task (i.e. both rating on a five-point Likert scale, and by providing general written feedback on each scenario, see section 2.2). This session was to facilitate PSAB members to provide expert review of the revised scenarios with a focus on: a) whether the scenarios are representative of a realistic event, and; b) whether the scenarios would be beneficial to aid learning.

Within the session, the PSAB members were provided with the two revised 'worst case' scenarios, one with an optimised communication strategy and one with a poor communication strategy (which are shown in Figures 1 and 2). The PSAB were asked to provide general thoughts on the scenario and then also specifically their opinions on how realistic and feasible the scenarios are.

Poor Communication Strategy
Emergency responders tell you, and others nearby, to follow them to a location outside the train station. They then ask you to remove the top layer of your clothes, down to your underwear. They give you some blue roll (paper towel) and tell you to use it to wipe the liquid off your skin. While you are wiping the liquid off your skin, emergency responders set up a large shower system, using hoses from their fire engines. Emergency responders tell you and the others affected to form a queue and wait to enter the large shower system. You are told to remove your underwear, and then enter the shower one by one.

Figure 1 Poor communication strategy adaptation added to the worst-case scenario in the Scenario Development session

Optimised Communication Strategy
***You explain to members of the public that they have potentially come into contact with a harmful chemical**, and to **prevent further contamination** they should follow you to a location outside of the train station. Once there you ask them to remove the top layer of their clothing. **You explain that by removing the top layer of their clothing, they will be removing 80-90% of contaminant which they may have come into contact with and removing their clothing will therefore help to protect them**. You provide members of the public with blue roll (paper towel) and tell them to use it to wipe the liquid from their skin. **You explain that using the blue roll to wipe to remove the liquid from their skin will remove any contaminant which may remain on their skin, and will therefore help to reduce their risk from the contaminant**. You set up the shower system and ask members of the public to form a queue and to wait until called to enter the showers. **You explain that going through the shower will remove any contaminant which may remain on their skin**. You tell members of the public to remove their underwear before entering the shower one by one, **as this will help to ensure that all remaining contaminant is removed from their skin during the shower**.*

Figure 2 Optimised communication strategy adaptation added to the worst-case scenario in the Scenario Development session. Bold text highlights additional information that has been added to optimise communication.

There was agreement that the scenario was feasible. The discussion around feasibility raised the following points:

"Both scenarios are feasible of course the optimised-detailed will be more precise on what will be the outcome."

"The scenario is feasible. We must consider the use of the agent to be more realistic, otherwise we will have only casualties."

"Previous exercises with 30 plus casualties has raised challenges in effectively getting blue roll [paper towel] *to casualties."*

"A powder explosion is difficult to spread, generally it takes a fire to spread powders using air."

Several contextual changes were raised by the PSAB. This included: preference for a chemical incident, with more detail about how the contaminant is delivered:

"We can use Chemical Incident Scenario maybe?"

"Worst case – explosion and chemical."

requests for additional information:

"There is no information about all the affected people can be reached and put together..."

"Can you speak to how the white powder is delivered: Reference "Explosion"?"

an additional challenge:

"Attach a crime or criminal activity. Responders would be difficult though, e.g. forensics, police."

and, one PSAB member presented the idea of taking a holistic approach to scenario development:

"Could it be possible to approach it holistically, meaning to understand what will be the prioritisation from both sides? Could that be helpful for all of us to identify as more as possible gaps?"

PSAB members also showed a preference for ensuring the volunteers used within the exercise should be varied:

"Who are we asking in terms of volunteers? … All ages need to be used, dogs etc"

"Adequate to the variety of audience (teenagers, senior citizens), technical resources (possibility or capability to access or use) and to the level of use of them (TICs)"

"Other option: involving the people that would be called to the scene, e.g. religious leaders, different numbers of casualties"

In relation to variable levels of communication in the scenario, the PSAB raised the following points:

"The challenge is not the "usual" communication but the communication with special groups of victims (e.g. disabled etc.)."

"Communication to the civil society must be always "multichannel" to reach people from traditional means of communication to social networks"

"COVID-19 has demonstrated that the civil society is very dependent on social media and looking for short guidelines from telecom operator. In COVID-19, it took time to set up this channel of communication. Perhaps, the States and South Korea were the quickest."

"We stick mostly to traditional critical communications means. In this trial, perhaps, victims can receive instructions directly into their mobiles too. Assuming that they have exited the area of contamination and waiting for assistant. A kind of emergency robot on telecom operator."

"In a scenario like this, in respect of communication issue, the victims initiate this channel of communication, I suppose to communicate this reality to the emergency responder and after that, depending on the scenario (city, village, rural area) can intervene the witnesses though the social network, local journalists... In order to value the impact over the local society."

"I believe it doesn't have to be so detailed. It can describe that the Blue lights will give guidance to the infected- casualties on what and how to do. the answer of the population depends on good communication and education on CBRN events, done in peacetime."

The idea of adapting and amending the scenario to show development across the three scheduled live exercises was apparent:

"One without the tools and one with the tools would be good."

"Maybe one suggestion: In a first phase, there shouldn't be a communication plan or instruction for the practitioners … in a second round with recommendations from PROACTIVE."

One PSAB member believed the current scenario was common and other similar exercises have already taken place.

"Somehow, I have the feeling is that this scenario is very common. Similar exercises have been conducted earlier. I am trying to Figure out what new am I going to learn afterwards?"

However, other PSAB members proposed the idea of injects to be added to the scenario, which would provide variability and would allow for evaluation to take place to ascertain the effect of introducing additional variables as well as manipulation effects.

> *"This could be a good starting point to determine a CBRNe info. This is the previous info that can be completed with additional injections in other to determine the capabilities to implement."*

> *"The use of additional injects are necessary for the execution of the exercise so everybody will be aware of what to do and how the others will react. "*

> *"In Italy in table top exercises we often use communication injects like fake news about an incident, to check how the communication function of the authority can react and correct the information."*

3.4. SCENARIO DEVELOPMENT PROCESS - STEP 4

Following the iterative process of practitioner stakeholder feedback, PHE further revised the prototype scenario to: a) standardise the information presented across both the initial scenario paragraph (presented in Table 2) and subsequent communication information (presented in Figures 1 and 2), and; b) incorporate comments made by the PSAB workshop attendees. Specifically: the scenario was standardised as a liquid chemical incident throughout; further information concerning the delivery of the contaminant was provided, and; the scenario was specifically contextualised as a terrorist attack.

Furthermore, the scenario was also redrafted to represent the responder's perspective, creating two exercise scenarios: 1) responder-focused scenario; 2) public-focused scenario. The responder-focused scenario will be used as the main exercise scenario and will guide initial exercise play, while the public-focused scenario will be used to inform the development of a pre-exercise information briefing for members of the public who participate in the exercise. The public-facing scenario will be finalised following discussion with representatives of the PROACTIVE Civil Society Stakeholder Advisory Board (CSAB) during the rescheduled September 2020 workshop.

The final prototype scenarios are presented in Figure 3.

3.5. SCENARIO DEVELOPMENT PROCESS - KEY OUTCOMES AND RECOMMENDATIONS

These scenarios are the result of several iterations of stakeholder-driven feedback and close consideration of the needs of both PROACTIVE and eNOTICE. As a function of this process, the above scenarios represent a realistic assessment of a reasonable worst-case scenario, tailored to fit the needs of the PROACTIVE project while still closely adhering to the requirements of the eNOTICE exercise in Rieti.

As scenario development for the three exercises is a fully collaborative endeavour between PROACTIVE, eNOTICE and exercise organisers, and given delays imposed by the COVID-19 pandemic (resulting in the rescheduling of the Rieti exercise), it has not been possible to fully execute

the scenario development plan as detailed in Task 6.1. Instead, PROACTIVE WP6 have developed an initial prototype scenario that can be used to inform ongoing discussions between PROACTIVE and eNOTICE, both in relation to the rescheduled Rieti exercise and future exercises. This scenario can be adjusted, amended, extended, and tailored to fit a range of exercises and so represents a strong base from which to revise and develop the scenarios across the duration of the PROACTIVE project.

In addition to the developed scenarios, throughout the iterative process of scenario development with practitioners, PSAB members, EU LEAs, and Policy Makers, detailed in the previous sections, several key components of critical importance for the PROACTIVE project emerged. When considered alongside the final list of guidelines and recommendations for mitigation and management of CBRNe terrorism (initially drafted within D1.3 and revised for D2.2), it is possible to provide a series of conclusions and recommendations for key components that should be included in future scenario development discussions with eNOTICE to ensure that: a) the requirements of PROACTIVE are met; b) the exercises represent a reasonable worst-case scenario, thus demonstrating maximum value for participating practitioners, and; c) opportunities to maximise learning from the exercises are taken.

These key insights, conclusions and recommendations are structured in the following sections according to the step of the process in which they were identified.

3.5.1. Step 1

The stakeholder engagement activity (a one-day workshop held with experts from emergency service, health and Government organisations) conducted parallel to (and involving representatives from) PROACTIVE revealed a series of variables that all have the ability to influence how difficult a scenario is perceived to be (see Table 1). These include:

- location;
- communication provided by responders;
- public awareness and knowledge;
- agent affect;
- weather conditions;
- dependants;
- confidence in responders;
- actions of the crowd and casualties.

It is therefore possible to vary these different elements of any scenario to influence the perceived difficulty. Indeed, scenarios that were developed following this parallel activity used these elements to provide a series of variably difficult scenarios from ‘best case’ to ‘worst case’. This activity was validated by the PSAB feedback in which the scenario designed to represent the ‘worst case’ was also deemed the most difficult to manage. Future scenario development in collaboration with

eNOTICE can therefore use these elements to tailor the scenario to be more or less difficult as required by the demands of both projects and the exercise organisers.

3.5.2. Step 2

Qualitative feedback provided by the PSAB on these initial scenarios emphasised the lack of responder communication as key to the perceived difficulty of the 'worst case' scenario. Given the emphasis of PROACTIVE on considering CBRNe preparedness and response, focusing on the intersection between responders and the vulnerable civil society, responder communication must be a critical consideration across all three exercises. To this end, a prototype draft of both a poor and optimised communication strategy was prepared. These can be incorporated into the exercises as required (either at the outset, or as an inject). Furthermore, several guidelines and recommendations for mitigation and management of CBRNe terrorism developed within the project (refined within D2.2) make reference to features of responder communication. Specifically:

- Communication should aim to reduce anxiety, by providing information to enhance self-efficacy.
- Communication should: 1) inform the public about loved ones' whereabouts in relation to family, friends and pets; 2) provide information about active police and security efforts to apprehend terrorists; 3) provide information on the importance of complying with instruction (including health specific information to address public health concerns; 4) and be delivered by a credible spokesperson (e.g. local resources, hazard groups and health departments).
- Information provided by authorities should be pre-planned, where applicable, to ensure prioritisation and consistency, provide uniformity and advocate cohesion.
- Information should be available in writing (i.e. print form), where possible, using non-complex language.
- Information should be provided in multiple languages, pictographic form, and sign language.
- It is necessary to establish whose duty it is to inform the public of CBRNe events, and who should be responsible in communicating during incident information.
- Official communication should be honest, empathic, assertive and reliable.

These recommendations should be considered while developing any future iterations of scenario drafts or communications injects.

3.5.3. Step 3

Expert feedback provided during the Pre-Exercise workshop held with Practitioners, EU LEAs and Policy Makers highlighted several additional considerations for scenario development.

Firstly, more information should be provided about the affected individuals (i.e. those contaminated or injured) and the type of incident (i.e. whether the incident is likely to be criminal or accidental, and the potential cause of contamination/ injury, etc.). Expert feedback also advocated the presentation

of information to responders via an App during the exercise (app development and the development of pre-incident information forms an integral part of the PROACTIVE project).

Secondly, expert feedback emphasised the importance of ensuring learning from the exercises, through the manipulation of specific variables (such as those detailed in the preceding section) to vary the difficulty of the scenario and the examination of the efficacy of interventions (based on the guidelines and recommendations developed in D1.3 and D2.2) for improving exercise outcomes. This could involve, for instance, running the first exercise without intervention and then comparing the outcomes to the second exercise in which an intervention is implemented. However, there are difficulties in comparing outcomes across exercises (especially where they are in different settings and with likely different scenarios as in the current case). It would be preferable to test the effectiveness of an intervention *within* an exercise (e.g., with half of responders and/or casualties receiving a specific intervention and half not). Given the emphasis of PROACTIVE on effective responder – casualty communication, and the development of pre-incident information materials within WP5, an intervention based on information provision to casualties is recommended and can be developed as exercise planning progresses.

The final proposed scenario for both members of the public and responders (with potential communication variation) is shown in Figure 3. It should be noted that the titles of 'poor' and 'optimised' communication have been included in this document in order to provide clarity for the reader; when the live exercises take place, neither the responders nor the public will be made aware of whether they are delivering or receiving the 'poor' strategy or the 'optimised' strategy.

The final scenario (presented in Figure 3) has the potential to be amended, adjusted and tailored for optimal learning by manipulating the following aspects:

- Number of people directly affected (e.g. collapsed)
- Number of other people leaving (to be managed by responders)
- Percentage of vulnerable crowd members

Responder Scenario	
It is a cold, wet late evening, and you are called to a train station following reports of a malicious incident. Available information suggests that an individual has run through a crowd of people waiting on the station platform, and sprayed members of the public with a liquid. As you arrive at the station you see that there are around 30 people who are covered in the liquid and are coughing and struggling to breathe; some have collapsed. You notice that many other people are starting to head for the exits to leave the station.	
Responder Communication Strategy	
Poor	**Optimised**

You tell members of the public to follow you to a location outside of the train station. Once there, you tell members of the public to remove their top layer of clothing, down to their underwear. You provide them with blue roll (paper towel) and tell them to use it to wipe any liquid from their skin. You set up a decontamination shower system and tell members of the public to form a queue and to wait until called to enter the showers. You tell the public to remove their underwear before entering the shower one by one.	You explain to members of the public that they have potentially come into contact with a chemical, and to prevent further contamination they should follow you to a location outside of the train station. Once there you ask them to remove the top layer of their clothing. You explain that by removing the top layer of their clothing, they will be removing 80-90% of contaminant which they may have come into contact with and removing their clothing will therefore help to protect them. You provide members of the public with blue roll (paper towel) and tell them to use it to wipe the liquid from their skin. You explain that using the blue roll to wipe to remove the liquid from their skin will remove any contaminant which may remain on their skin, and will therefore help to reduce their risk from the contaminant. You set up the shower system and ask members of the public to form a queue and to wait until called to enter the showers. You explain that going through the shower will remove any contaminant which may remain on their skin. You tell members of the public to remove their underwear before entering the shower one by one, as this will help to ensure that all remaining contaminant is removed from their skin during the shower.
Public Scenario	
On a cold, wet late evening, you are travelling alone to meet a family member. As you are standing on the platform of an unfamiliar train station, someone wearing dark clothing and a backpack runs through the crowd spraying people with liquid. You look down and realise that your clothing is wet. Your eyes sting and you start to cough. You look around and see that other members of the crowd are also starting to suffer.	
Public Communication Conditions	
Poor	**Optimised**
Emergency responders tell you, and others nearby, to follow them to a location outside the train station. They then ask you to remove the top layer of your clothes, down to your underwear. They give you some blue roll (paper towel) and tell you to use it to wipe the liquid off your skin. While you are wiping the liquid off	Emergency responders explain that you have potentially come into contact with a chemical, and to prevent further contamination, they ask you to follow them to a location outside the train station. They then ask you to remove the top layer of your clothes, down to your underwear. They explain that removing the top layer of your

your skin, emergency responders set up a large shower system, using hoses from their fire engines. Emergency responders tell you and the others affected to form a queue and wait to enter the large shower system. You are told to remove your underwear, and then enter the shower one by one.	clothing will remove 80-90% of any contaminant which you may have come into contact with, and will therefore help to protect you. They give you some blue roll (paper towel), and ask you to use it to wipe the liquid off your skin. They explain that using the blue roll to wipe yourself down will remove any contaminant which may remain on your skin, and will therefore help to reduce any risks from the contaminant. While you are wiping the liquid off your skin, emergency responders set up a large shower system, using hoses from their fire engines. Emergency responders explain that going through the shower will remove any contaminant which may remain on your skin. Emergency responders ask you and the others affected to form a queue and wait to enter the large shower system. You are asked to remove your underwear, as this will help to ensure that all remaining contaminant is removed from your skin during the shower. You are asked to enter the shower one by one.

Figure 3 Final responder scenario containing a chemical incident on a railway platform

3.6. SCENARIO DEVELOPMENT - SUMMARY

PROACTIVE WP6 have developed an initial prototype scenario that can be used to inform ongoing discussions between PROACTIVE and eNOTICE both in relation to the rescheduled Rieti exercise and future exercises. This scenario can be adjusted, amended, extended, and tailored to fit a range of exercises and so represents a strong base from which to revise and develop the scenarios across the duration of the PROACTIVE project. Future work will involve close liaison with eNOTICE and exercise organisers as the projects progress in order to help develop these bespoke scenarios and tailored injects. The public-facing scenario (see Figure 3) will be finalised following discussion with representatives of the PROACTIVE Civil Society Stakeholder Advisory Board (CSAB) during the rescheduled September 2020 workshop. This workshop will also facilitate discussion of the public-focused communication strategies (presented in Figure 3) to ensure all the communication needs of relevant vulnerable groups are represented in the strategies.

4. PROACTIVE/ ENOTICE EXERCISE SCHEDULE AND SCENARIOS

As outlined in D6.1, an outline format for the three PROACTIVE exercises was agreed with eNOTICE. The first two exercises are intended to be modular and focus on: 1) a Specialist Operational Response scenario; 2) an Initial Operational Response scenario, which is then brought together into a final large-scale exercise. In this way, the exercises will each examine different types of response and so may be more or less relevant for addressing the gaps and incorporating the recommendations identified through PROACTIVE WP1 – WP3.

Prior to the COVID-19 pandemic, the Rieti exercise was the first scheduled exercise and so was at the most advanced stage of development. This was planned as a single-day event with a 2-hour duration for the field exercise. A maximum capacity of 35 members of the civil society was specified, and the focus was to be on a railway scenario linked to a terrorist attack. Specifically, the planned scenario involves a simulated dispersal of a chemical agent through an air-cooling system with some pyrotechnics and smoke added for realism. The exercise was not planned to involve an ongoing terrorist threat, and discussions were underway concerning the potential for developing a further scenario involving the discovery of a terrorist device within a mock underground station at the exercise site. This exercise was designed to focus on Initial Operational Response. Participants were to be identified from the local community (and particularly the schools) in order to ensure a good cross-section of the civil society (including representatives from members of vulnerable groups). In order to ensure ethical conduct, volunteers will need to be fully briefed on the plans for the exercise, however this does not mean that there will be no opportunity for a surprise effect with participants (within reason). Finally, this exercise was also planned to represent the initial testing of pre-incident public information materials developed as part of T5.1.

The template scenario developed through the aforementioned process is critical for ensuring that the needs of the PROACTIVE project are well represented through collaborative exercise planning discussions between the PROACTIVE team, eNOTICE, and local exercise organisers. The close similarity between elements of the template scenario developed and presented within this document and the details of the Rieti exercise plan demonstrates the clear synergy between the organisational teams. Further additional information (e.g., communication information, disambiguation, whether the response is early/ late, any potential additional incidents/ scenarios, etc) will be developed as the exercise planning continues, with several elements being included within the exercise as injects to supplement the base scenario (delivered to the exercise players as the incident unfolds).

Unfortunately, the COVID-19 pandemic has resulted in substantial delays to the conduct of the three exercises detailed in the PROACTIVE Description of Action. A revised preliminary timetable for the exercises is presented below in Table 7.

Table 7 Updated Exercise Schedule

Exercise Location	Scheduled Date
Dortmund, Germany	April 2022
Rieti, Italy	October 2022
Ranst, Belgium	May 2023

This rescheduling of exercises has meant that the Dortmund exercise will now be the first exercise, with Rieti following second. Initial Consultation and engagement has taken place regarding the planning of the Dortmund exercise and it will focus on a Specialist Operational Response scenario.

5. SPECIFICATION OF THE EVALUATION METHODOLOGY

A rapid evidence review was conducted to scope out the literature on evaluations of field exercises. The outputs of this review are combined with relevant project outputs and subject matter expertise in Section 3.3.

5.1. RAPID EVIDENCE REVIEW - METHOD

5.1.1. Selection Criteria

To be included within the review, articles must have met at least one of the first four criteria listed in Table 8: they must contain evaluation data relating to an exercise; the exercise must be a field exercise; the exercise must include some aspect of emergency preparedness; and the exercise must involve members of the public, or others with no particular expertise in emergency management (e.g. medical students). Articles also had to be available in English. Additionally, animal data were excluded due to the PROACTIVE project's focus on human factors. Finally, articles published prior to 2001 were excluded from this review to limit articles to those that reflect the shift in emergency response planning following the 9/11 USA terrorist attacks.

5.1.2. Information Sources

Searches were conducted on EMBASE and Medline on the 2nd June 2020.

5.1.3. Search

The search terms included terms relating to types of scenarios (e.g. Scenario), types of emergency situations (e.g. Preparedness), and descriptive words for evaluation (e.g. Evaluation). The search was conducted using Boolean logic: OR was used for each search term with these three searches, which were subsequently combined using the AND operator (see Appendix B for full search conducted).

Table 8 Final inclusion criteria used for full text screening

Inclusion criteria	Notes
Exercise involves the public	This could include medical students but not staff members. Additionally, public must be involved in a non-medical capacity. Excluded articles which pertain to animal rather than human data
Exercise is a field exercise	Table top or virtual simulation exercises were excluded.

Article contains data on the evaluation of an exercise	Articles that did not give details on the method of evaluation were excluded
Exercise related to emergency preparedness	This excludes exercises to practice clinical drills
Available in English	
Published since 2001	

5.1.4. Study Selection

The search identified 372 articles (246 from EMBASE). Around a third of the papers (n = 129) were duplicates and so discarded. The titles of the remaining 243 articles were then assessed on whether they reported evaluation data from a field exercise involving members of the public. At this stage, only papers that were clearly outside the scope of this review (e.g. reported results of a biological experiment) or were not published articles (e.g. conference abstracts) were removed. The abstracts of the remaining 83 articles were then reviewed, with 36 articles being removed. Articles were excluded for 1) reporting surveys or literature searches on emergency preparedness, 2) reporting data from exercises that were not conducted in the field (e.g. table top exercises), 3) not involving members of the public, or 4) reporting on the development of exercises rather than their evaluation. The full texts of the remaining 47 articles were then reviewed to ensure they met the inclusion criteria. Of the 47 articles in the full-text review stage, only nine met the inclusion criteria. A PRISMA diagram can be found in Figure 4 detailing the full screening process. Additionally, the authors of this review were aware of 22 papers from other sources that met the inclusion criteria, and these were included for data synthesis.

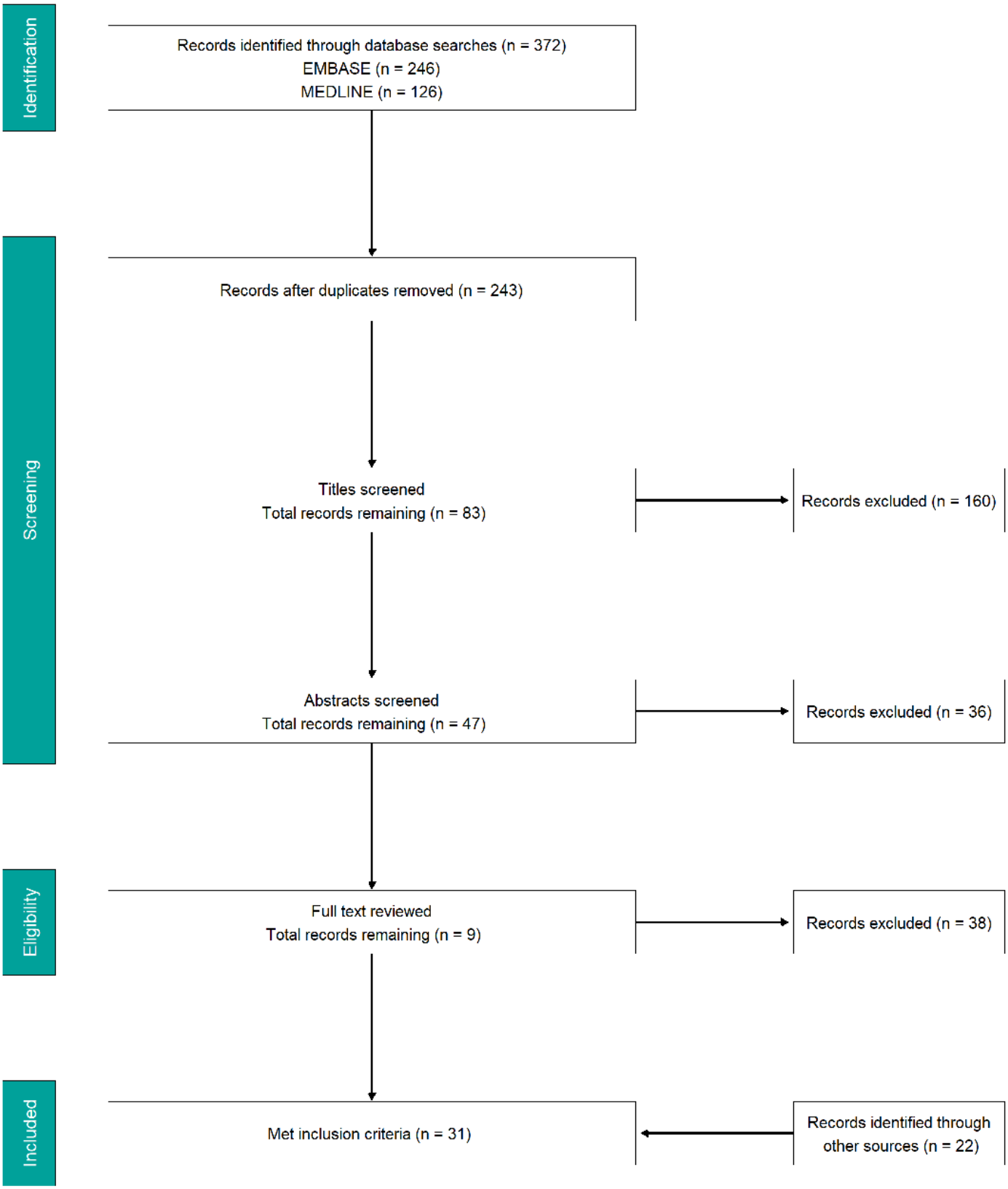

Figure 4 PRISMA diagram detailing the stages of the review process

5.1.5. Data synthesis

A standardised data extraction process was carried out by WN on all 31 papers to extract the following information: authors; date of publication; institution location of the first author (to ascertain study origin); location of study; incident described (if applicable); evaluation method; sample information (consisting of N, male to female ratio and age data) and specific sample characteristics (see Appendix C for full Table). As studies used within this review are highly varied and non-comparable, meta-analysis was not attempted. Instead, a narrative synthesis of the included papers was conducted. Thematic analysis was undertaken to extract the types of evaluation methods used in each article.

5.2. RAPID EVIDENCE REVIEW - RESULTS

5.2.1. Date of Publication

Papers included in this review were first published in 2003, with over half (16) of the papers being published since 2010 indicating a growing interest in the inclusion of members of the public in emergency preparedness exercises but also in the evaluation of conducted exercises.

5.2.2. Country of Origin

To display the geographical spread of the papers used within this review, data were extracted relating to the location of the first authors institute at the time of publication. Seventeen of the papers originated from the USA. Nine papers were published by UK based authors. Two papers originated in Australasia (see Appendix C for full details). Country of origin was not restricted when conducting the search, and so the countries of origin presented in this section represent all eligible articles identified from the literature search process.

5.2.3. Incident Described

Twenty papers described a CBRN exercise, predominantly of a chemical nature. Four papers described a disaster drill (e.g. house fire), three papers described exercises simulating natural events (e.g. earthquakes), while four papers described a drill within a school setting.

5.2.4. Evaluation method

In terms of the methods used to evaluate the exercise, nine of the papers used multiple methods. Questionnaires were widely used to gather feedback with seventeen papers using this method. Ten papers used observations to assess exercises and eight papers used interviews/focus groups. Finally, seven papers used performance measures when evaluating a drill.

5.2.5. Casualty sample

The number of casualties ranged from n = 2 to n = 18,211. Samples often included those drawn using a convenience sampling method (e.g. a nursing college using undergraduate nurses, Primeau and Benton, 2019) but others included members of the public recruited via advertisement. A table displaying extracted data in relation to sample size and characteristics, including unique characteristics and gender distribution can be found in Appendix C.

5.3. EXERCISE EVALUATION ANALYSIS – RAPID EVIDENCE REVIEW AND SUBJECT MATTER EXPERTISE

This section draws together the articles from the rapid evidence review with relevant project outputs and subject matter expertise held within Public Health England to present various evaluation tools suitable for the evaluation of a field exercise. Throughout, noting the purpose of PROACTIVE, the analysis will focus on assaying evaluation methods in terms of their suitability for vulnerable members of civil society.

5.3.1. Observations

Published literature

One method of evaluating exercises is through observations. This method can be open or structured. In the latter a predetermined criterion is used to assess the completion of expected outcomes/ actions. In the former an observer records all decisions and activities undertaken in the exercise to build up a timeline of events. In terms of an evaluation method, four factors are likely to impact the quality of evaluation: the number of observers, interrater reliability, the behaviours one wants to observe, and the type of observation used.

In terms of the number of observers, it is important that any evaluation has sufficient observers to provide coverage to key areas within an exercise. For example, while Allen, Lorek, and Mensia-Joseph (2008) trained 17 observers to assess a school-based mock drill, other studies only had two to four observers per site (see Ramirez, Kubicek, Peek-Asa, and Wong, 2009). Clearly, the quantity of data gained from 17 compared to four observers is likely to be higher and will ensure that a wider range of events is captured. However, an increased number of observers will also necessitate higher costs, both in time to train and to pay for observers' time.

Linked to the number of observers is interrater reliability. This is a measure of agreement between observers. High interrater reliability shows that observers are in agreement with how to grade an event, with low interrater reliability demonstrating that there is a discrepancy between observers (Lange, 2011). Only one study in this analysis reported their interrater reliability score (Kaji & Lewis, 2008), which was relatively low. Other studies, such as Allen et al. (2008) commented on the importance of interrater reliability but did not assess it. This means it is not possible to comment on whether observers were using checklists provided in the same fashion, and so undermines the overall results. As Kaji and Lewis (2008) note, longer training can help improve interrater reliability – though this would also lead to higher costs.

The type of behaviour under observation also affects the suitability of this method. In many studies the observers are assessing outcomes. For example, Klima et al. (2012) used observers to assess if best practice was being followed, e.g. the correct handling of contaminated patients. Such behaviours are well suited to evaluation through observation as they are in effect binary events, they either did, or did not happen. In other cases, however, it may be questioned whether an observation method is the most sensible approach to evaluate an exercise. For example, Fertel, Kohlhoff, Roblin, and Arquilla (2009) used four trained observers to assess the mood of children participating in a decontamination drill at a New York hospital. It was found that younger children were more likely to display fearful mood compared to older children. Yet it is difficult to assess the mood of an individual

and perhaps other methods (such as a mood questionnaire given to the child or caregiver) would have been a more appropriate evaluation method.

The last aspect to consider when using observations to evaluate an exercise is how one records observations. Some evaluations (such as Cicero et al., 2017 and Kaji and Lewis, 2008) used structured checklists. As Kaji and Lewis (2008) note, such checklists enable comparisons between sites (e.g. different hospitals in the same drill), and indeed, could allow for performance of a site to be tracked over time (e.g. performance of a hospital in yearly drills). Cicero et al. (2017) ran a series of emergency exercises with emergency medical services over a six-month period. The exercises staged a multiple casualty house fire, a school shooting and a school bus collision. Each event had four actors with differing injuries, including an uninjured child with special health care needs. Observers used a standardised checklist and video recordings to assess personal performance after each of the three scenarios. Findings showed that paediatric disaster triage improved with time. The use of the standardised checklist assessment tool in this exercise enabled performance to be consistently measured over time thereby facilitating a reliable evaluation of personal performance – demonstrating the gains from participating in the exercise. However, a structured checklist only records events that were pre-determined to be of importance for the evaluation. The alternative is to use an unstructured checklist where all events can be recorded (or more likely, a mixed checklist).

Exercise reports and subject matter expertise

The reports produced by PHE have made use of both structured and unstructured observational methods. For example, as with Klima et al. (2012), in their evaluation of a CBRN exercise the Health Protection Agency (2009) assigned observers to key areas of the drill. Evaluators received a briefing prior to the event and were able to evaluate the exercise against set best practise. While giving observers training before the drill is likely to have helped with the interrater reliability this was not reported, nor was the number of observers assigned to each area (though a comparable exercise ran by the Health Protection Agency in 2008 used a similar number of evaluators as Allen et al., 2008). Additionally, the Health Protection Agency (2009) report used structured observations. Dacey et al. (2011) evaluated a mass decontamination exercise and highlighted the benefits of an unstructured checklist. Using this method, the observers were able to raise issues around the management of disabled casualties. The use of an open-ended checklist allowed for these issues to be captured in more detail then perhaps might have been recorded in a structured checklist.

Observation is therefore a useful method for evaluating a field exercise. However, for this to be effective, there must be a sufficient number of trained observers, and interrater reliability must be measured in order to establish an agreement between observers on key outcomes. Careful consideration should also be given to whether the outcome measured is easily observable (e.g. practical challenges) as opposed to being something more subjective or less readily observable (e.g. participant mood). For the exercises carried out within the PROACTIVE project, unstructured observation may be particularly useful for capturing any practical challenges associated with the management of members of vulnerable groups during the exercise. It may be that some form of structure can be provided to the observation method, though this will depend on the scenario used within each exercise. Depending on the number of observers available to observe the exercise on the day, consideration will also be given as to whether each exercise can be filmed in order to enable observation to be carried out after the exercise.

5.3.2. Questionnaires

Published literature

Another method to evaluate exercises is through questionnaires. Unlike observations which solely record events, questionnaires can be used to understand the experience of the exercise from the perspective of those involved (i.e., responders or volunteers).

One key aspect of questionnaire design is the type of scale used. Many studies reported in this analysis use a Likert scale, where respondents choose one of the given options. For example, Unver et al. (2018) asked those delivering medical care (all senior nursing students) to provide feedback on a simulated earthquake drill using a 3-point scale. A Likert scale questionnaire was also used in evaluations by Stergachis et al. (2007), and Charney, Lehman-Huskamp, Armbrecht, and Flood (2011). However, they used a differing number of responses. For example, Unver et al. (2018) used a 3-point scale (partly agree, agree, fully agree), whereas Charney et al. (2011) used a 5-point scale (1 = not at all important, 5 = extremely important). Research has shown that scales with fewer than four responses can suffer from reliability issues, with between four and seven responses being optimum (Lozano, García-Cueto, & Muñiz, 2008). Another paper, Beaton et al. (2003), used a questionnaire with a visual analogue scale (VAS) to evaluate members of the publics' experience of an anthrax exposure drill. A VAS normally uses a scale of 0-100, with respondents choosing to place a mark between these points that represent their opinion. Such scales have been found to generate more reliable responses (Voutilainen, Pitkäaho, Kvist, & Vehviläinen-Julkunen, 2016), but some evidence suggests that Likert scales are more appropriate for children (van Laerhoven, van der Zaag-Loonen, & Derkx, 2004).

The audience of a questionnaire is also an important factor to consider. Johnson, Johnston, Ronan, and Peace (2014) designed a questionnaire to assess children's knowledge of safety behaviours during an earthquake and tsunamis. Importantly, the questionnaire was designed to be child friendly. This was achieved by using short and simple questions and ensuring they were matched to the appropriate reading level. However, despite the authors giving consideration to ensure that the questionnaire was age appropriate and piloting the questionnaire before deployment, it seems children might have misunderstood the meaning of some questions. This highlights the difficulty of designing a truly age-appropriate questionnaire.

Another issue to consider was raised by Currie and Heslop (2018). The authors gave a 26-point questionnaire to simulated patients (mostly medical students) immediately after an exercise. Overall, 46 participants completed the survey with respondents highlighting the lack of communication they received as casualties as a key issue. The authors raise the problem that as data was collected through participation in an exercise it could be the case that participants were able to mentally prepare for the event (as opposed to a real incident for which there may be little or no preparation time). This could lead to bias in response in the survey, with responses not accurately reflecting participants 'true' feelings. This issue could potentially be lessened by interviews or focus groups that would enable a closer examination of participants feelings surrounding participation in the exercise. A similar limitation was highlighted by Carter, Drury, Amlôt, Rubin, and Williams (2013). The authors used questionnaires to assess participants (who were members of the public) experiences of partaking in a decontamination exercise, with results indicating that poor communication and concerns about privacy can reduce public compliance with a decontamination process. However, it may be the case that participants were showing demand characteristics to

answer survey questions in a certain manner. In this case, as the authors note, using direct observation to assess for behavioural variables could be used to substantiate the findings from the questionnaires.

A clear benefit of using questionnaires to evaluate exercises is that they can be quick (helping to minimise disruption to exercise organisers and participants) and facilitate cost-effective large-scale data collection. However, it is still important that sufficient time is set aside for the completion of questionnaires. For example, Digregorio et al. (2019) asked medical participants to complete the Interprofessional Collaborator Assessment Rubric (ICAR) before and after taking part in their drill. While completion rates of the pre-drill questionnaire were high (87.74%), only 54.72% of exercise participants completed the post-event questionnaire. The authors hypothesised that this could have been due to the limited, one week, time frame for the response that fell within an examination week. Such low completion rates could have been mitigated by giving participants time to complete follow up evaluations.

Lastly, it is important to ensure any questionnaire measures the behavioural aspect of interest. Carter, Drury, Rubin, Williams, and Amlôt (2012; 2013) assessed members of the publics experience of participating in five live drills. The questions used by the authors were designed to assess the exercises, rather than to examine the psychological constructs outlining witnessed behaviour. However, in Carter, Drury, Rubin, et al. (2013) the data was used in a hypothesis driven analysis. Though there was strong prima facie reason to assume that questions did link to a psychological construct, some of the constructs of interest, such as anxiety, were only measured once which may reduce the reliability of measures. On a related matter, Charney et al. (2011) used a survey to evaluate the experience of 22 caregivers of children during a disaster drill. Overall, caregivers were happy with the experience of the drill. However, the survey was not piloted and so there may have been issues in the survey (e.g. interpretations of the questions) that affected the results.

Exercise reports and subject matter expertise

Many of the issues raised in the published literature are mirrored in exercise reports. For example, Krieger et al. (2014) used questionnaires to assess public perceptions of emergency response pre- and post- a CBRN exercise. However, the questionnaires were not piloted, and participants noted that there were some questions they found difficult to understand. Such difficulties are likely to impact upon the quality of data collected. Krieger et al. (2014) also report concerns about questionnaires not asking the *right* questions. In the study, while the exercise was CBRN based, questions asked about emergencies in general. Therefore, it might be the case that the questionnaire lacked sensitivity to address the area of interest for the researchers. This paper highlights the need for questionnaires to undergo piloting and for researchers to ensure that questionnaires are fit for purpose, in that they provide the data to address the questions of interest to the researcher/ exercise organiser.

For their exercise report, Turner, Amlôt, and Simpson (2007) sought feedback from the 65 children acting as patients (aged from 6 – 14 years old) to review a mass decontamination drill. Unlike Johnston et al. (2011), Turner et al. used pre-existing questions which have been validated extensively as being appropriate for children. For one of the questionnaires, the Positive and Negative Affect Scale for Children, older children reported lower positive scores compared to young children. This finding could represent true differences in experience of the exercise or it could reflect that older children are more able to distinguish between responses on a 5-point scale and so are

less likely to endorse higher responses. This finding demonstrates the difficulty in using questionnaires with children and suggests it is important to see groups as containing sub-groups that each require a different approach, for example, adjusting questionnaires for different age ranges.

Questionnaires can provide valuable insight into participants' experiences during an exercise. If used pre- and post-exercise, they can be a valuable tool for understanding how participants' perceptions change over the course of the exercise, in a way that would be difficult to achieve using a different method of evaluation. However, for questionnaires to be effective, they must be validated to ensure they can be easily understood and are addressing the areas of interest, and they must be appropriate for the audience with whom they are being used. Given the advantages associated with effective use of pre- and post-exercise questionnaires, these will be used within all the PROACTIVE exercises to capture information about participants' experiences and perceptions. Within the PROACTIVE exercises, there will be a percentage of volunteers from vulnerable groups. It is possible that some members of vulnerable groups may face specific challenges in completing questionnaires. For example, as noted above, it is particularly difficult to design a questionnaire that is appropriate for children. Additionally, participants may face physical or communication-related challenges in filling out questionnaires. When designing questionnaires for use within the PROACTIVE exercises, consideration will be given to ensuring that these questionnaires are validated, fit for purpose, and able to be completed by all exercise participants; extra support will be given to those who require assistance with completing the questionnaires.

5.3.3. Debrief

Published literature

Another evaluation method is a debrief conducted at the end of an exercise. Hot debriefs can be conducted immediately after the exercise to collect instant thoughts or cold debriefs can be conducted sometime following the exercise to consolidate reflections and lessons learnt. Both types of debriefs were used in the papers highlighted by the rapid evidence review.

Hot debriefs can be used to highlight immediate issues. For example, Rehmani (2005) evaluated a small aircraft crash exercise in Pakistan involving 30 patients drawn from a hospital volunteer group. A debrief was conducted immediately following the exercise with the patients. This highlighted the need for communication to be given in multiple languages (rather than just English). However, Digregorio, Graber, Saylor, and Ness (2019), in their evaluation of a chemical spill exercise, highlight the importance of the debrief being led by someone who is sufficiently experienced. The authors comment that the individual leading the debrief following this exercise was inexperienced and focused more on areas of improvement (such as patient management) rather than self-reflection.

Cold debriefs, such as Vinson (2007) and FitzGerald, Sztajnkrycer, and Crocco (2003), can be used to highlight wider issues. For example, Sweeney, Jasper, and Gates (2004) reported the lessons learnt from a large-scale disaster drill, with key areas of improvement identified as: managing victims' expectations and ensuring victims receive enough training to feel comfortable with their role. Johnston et al. (2011) used five observers to evaluate a New Zealand school earthquake drill, which involved 200 children aged from 5 – 13 years. Following the exercise, the observation team and school staff carried out a cold debrief. While cold debriefs can help to develop long term goals they can often come at the expense of losing the voice of participants. For example, in Johnston et al. (2011) the debrief did not include any discussions with the children involved.

Exercise reports and subject matter expertise

A debrief is often carried out as a matter of course following an exercise and is the most basic level of evaluation that should be expected. As noted above, a debrief can be hot or cold, and each method has its advantages. Ideally, everyone involved in the exercise should have the opportunity to attend a debrief, to give them an opportunity to reflect on their experiences, understand areas for improvement, and identify lessons learned. A hot debrief will be carried out for each of the PROACTIVE exercises. As a minimum, this debrief will include exercise organisers, emergency responders who take part in the exercise, and members of the PROACTIVE and eNOTICE project teams. Depending on the number of people involved, it may be appropriate to carry out separate hot debriefs with members of different groups. Whilst it can be useful for different groups to reflect together on their experiences of an exercise, in order to understand others' perspectives, if the group becomes too large people may not get an opportunity to contribute or may not feel comfortable sharing their views. A debrief can create the potential for conflict, particularly when discussions focus on areas that have not gone well, or aspects that could be improved. To ensure that any debrief carried out is as beneficial as possible, and to reduce the risk of conflict or confrontation, each debrief will be facilitated by an experienced individual who is able to direct discussions effectively.

5.3.4. Interviews and focus groups

Published literature

Interviews and focus groups allow for more in-depth data to be collected than either debriefs or questionnaires and facilitate a better understanding of participants' experiences and perceptions during the exercise. However, they are also more resource intensive.

The resources required to carry out interviews and focus groups mean it is often not possible to interview everyone who takes part in an exercise. For example, in their evaluation of an earthquake preparedness drill Alim, Kawabata, and Nakazawa (2015) conducted short (10min) one-to-one structured interviews with 40 (15%) of participants (student nurses). While feedback was positive the interviews were only conducted on a proportionally small sample of participants and so it is possible differing views were not captured.

Another issue is how to choose participants. To evaluate the effectiveness of emergency drills at schools Ramirez et al. (2009) conducted group interviews with students in schools across Los Angeles County in the USA. The authors used a purposive sample, with teachers selecting students based on the child's leadership and communication skills. The selection of children by teachers could have created a bias, with teachers selecting children who would be more likely to represent the school in a positive light.

Interviews with individuals with additional needs can also raise further complications. Taylor, Balfanz-Vertiz, Humrickhouse, and Jurik (2008) used interviews and focus groups to evaluate a decontamination drill in Chicago that involved people with physical disabilities, people with hearing difficulties and foreign language speakers. A benefit of using interviews to evaluate this exercise is that participants were able to provide in-depth information about their experiences and raise issues that the research term had not considered, thus helping improve decontamination for all. However, this process was lengthy, with focus groups lasting approximately an hour. Additionally, the need to conduct the session in the language appropriate for the individuals, including Spanish, American

Sign Language, and English for people with disabilities is likely to impose high financial cost when compared to other methods.

Exercise reports and subject matter expertise

The available exercise reports suggest various ways to overcome the issues raised through the published literature. Krieger et al. (2014) argue that it is possible to mitigate the issue of non-representative sampling by using data from interviews and focus groups to enrich knowledge gained from other methods, rather than by using them as a method to gain data on how whole populations might behave. Furthermore, Dacey et al. (2011) included all members of the public (who acted as casualties in the exercise) in their focus groups thus reducing any bias in the sample. With respect to evaluating experiences of children, while Ramirez et al. (2009) indicated issues around purposive sampling, Turner et al. (2007) also used a focus group as part of their evaluation and reported that children gave honest accounts of their experience. Importantly, the children highlighted negative aspects of the exercise indicating that social desirability was not necessarily biasing responses.

Focus groups and interviews are arguably the best way to capture data about participants' experiences during an exercise. In order for focus groups and interviews to be most effective several things must be in place. First, focus groups and interviews must be carried out by an individual who is experienced in this method of evaluation. This will help to ensure that discussions are appropriately focused, whilst participants are free to express their views in an open and non-judgemental environment. Second, focus groups and interviews must be tailored to the needs of participants. For example, interpreters should be available when running focus groups or interviews with those who do not speak English. Third, focus groups and interviews should be carried out alongside other methods of evaluation, to enrich, rather than replace, other forms of data. The exercises carried out as part of the PROACTIVE project are deliberately designed to capture (amongst other things) the experiences and perceptions of members of vulnerable groups. Focus groups and interviews can be a key way to gather in-depth data on participant experiences and will be an essential part of the way in which PROACTIVE exercises are evaluated. Given the highly resource-intensive nature of interviews, it is anticipated that focus groups will be used to capture data relating to participant experiences during the PROACTIVE exercises. However, there is potential for interviews to be conducted with individuals of interest to clarify key points. For the first exercise, it is anticipated that, given the relatively small number of participants, all participants will have the opportunity to take part in a focus group to discuss their experiences. Careful consideration will be given to the different needs of members of vulnerable groups taking part in the exercises, and additional support will be provided to ensure that all participants who want to take part in a focus group are able to do so. For later exercises in which there will be a larger number of participants, it may be necessary to randomly select a sample of participants to take part in focus groups, rather than giving all participants the opportunity to do so.

5.3.5. Physical quantitative measures

Published literature

Some of the articles in the rapid evidence review used time as a measure to evaluate drills. For example, Cone et al. (2009) assessed medical staff response times in a mock air craft crash drill. In the exercise observers used digital stopwatches to time specific aspects of casualty care. Such an approach can be easy to use and can offer an objective measure of key outcomes from an exercise. Ramirez et al. (2009) collected data on the time it took schools to evacuate the students from

classrooms. Findings showed that Elementary schools were faster to evacuate than Middle and High schools. However, in many cases these times were collected via self-report from schools. Importantly schools were not provided with set instructions in order to ensure that schools calculated evacuation start and stop times from the same points. This study highlights the benefits of using quantitative measures, in that they can provide objective measures, but also emphasises the importance of ensuring that they are prepared appropriately. In this instance, having independent observers at each school might have led to more accurate data.

Exercise reports and subject matter expertise

In their review of a mass decontamination drill, which included 37 individuals with disabilities, Dacey et al. (2011, also see Dalby et al., 2012) used tracking chips worn by casualties to track their movements as they passed through various treatment stages in the exercise. These trackers allowed the authors to provide an objective measure of the progress of patients. Though not presented in the report it would be possible to assay treatment times between those with a disability compared to those with not. Such data would help to highlight any time differences, which could be further explored through questions in a focus group.

In their evaluation of a field exercise involving mass casualty decontamination, Dacey et al., (2011) also used a physical measure of the efficacy of the decontamination process. This was done by applying a harmless fluorescent chemical simulant to the skin of casualty volunteers and using a specialist UV facility to examine how much simulant remained on their skin following decontamination. This therefore facilitated an understanding of the effectiveness of decontamination, an aspect that would be very difficult to capture using more traditional methods of exercise evaluation.

Physical quantitative evaluation methods can be a useful and objective way to measure the outcomes from an exercise. As noted, such methods can include timings for how long different activities take (measured by physically collecting timings, or using technological solutions), or collecting physical measures to understand the efficacy of processes which would be difficult to evaluate using traditional exercise evaluation methods. The exercises carried out as part of the PROACTIVE project are likely to include physical interventions (e.g. dry decontamination, wet decontamination etc) which may benefit from evaluation using physical quantitative measures. It is anticipated that this will include timing data, for example, timing how long it takes for exercise participants to progress through the different stages of the exercise. Given that 15% of participants within each of the PROACTIVE exercises will be members of vulnerable groups, such timing data will enable an objective measure of any differences in length of time taken by members of vulnerable groups in comparison to other casualty volunteers. Other aspects that may be difficult to measure using more traditional exercise evaluation methods include the effectiveness of any decontamination procedures, and casualty volunteers' anxiety during the exercise. Both aspects lend themselves to evaluation using physical quantitative measures, and this is something that will be considered when further developing the bespoke evaluation strategies for each of the PROACTIVE exercises.

6. EVALUATION METHODOLOGY – SUMMARY AND LOGISTICS

It is increasingly recognised that effective evaluation is essential to identify lessons learned following an exercise and generate recommendations for improvement in future incidents. There are a number of evaluation methods that can be used to capture different aspects of the exercise, facilitating an understanding of casualties' and responders' perceptions during the exercise, measuring the physical efficacy of any technical procedures employed, and identifying any improvements in response as a result of exercise participation. To ensure that the exercises carried out within the PROACTIVE project are as useful as possible, we will take a mixed-method approach to evaluating the exercises and will ensure that appropriate measures are used to capture different aspects of exercise play. We will also ensure that each evaluation method is tailored to the needs of those participating in the exercise, with particular consideration to the needs of members of vulnerable groups. For each exercise, we will:

a) carry out a hot debrief to capture the views of emergency responders, exercise planners, and project partners;

b) use pre- and post-exercise casualty feedback questionnaires to identify any changes in casualty volunteers' perceptions as a result of taking part in the exercise;

c) observe exercise play to identify any challenges that arise (particularly in relation to the management of vulnerable groups);

d) and carry out focus groups with casualty volunteers (and potentially interviews with individuals of interest to clarify key points) to facilitate an in-depth understanding of their experiences during the exercise and to identify any areas for improvement.

Depending on the exercise scenario, consideration will also be given to using physical quantitative measures (e.g. timing data, measures of efficacy) to evaluate the exercise. A detailed evaluation plan will be developed for each of the PROACTIVE exercises, and the evaluation methods used will be tailored to each specific exercise scenario.

In terms of logistics for exercise evaluation, PHE will take the lead for standardising and co-ordinating this evaluation by:

- Developing standardised template questions/data capture forms for the questionnaires, focus groups, observational data collection, and hot debriefs. These will be provided to the local response organisations ahead of the exercises in order to ensure that they can be translated. Every effort will be made to deliver these in a way that is accessible to any members of the vulnerable civil society that participate in each session (e.g., verbal administration of questions to individuals who have visual impairment, etc) and are intended to take place both in the local language and in English where appropriate.

- Providing training to additional members of the evaluation team (recruited from the consortium and the local response organisations). This will involve light-touch training sessions (via video conferencing in the first instance and face-to-face in the immediate build

up for the exercise) and guidance materials in order to standardise the conduct of these evaluation tasks. These materials will be based on similar materials developed as part of PHE's COVID-19 response work and previous training courses for research involving human volunteers.

- Analysis of data for each exercise will be led by PHE and will involve participation from across the PHE consortium. Light-touch training and standardised guidance on data entry and analytical methods (specifically for the entry and analysis of qualitative data as this is the most resource intensive method of data collection) will be provided by PHE in order to ensure the timely and rigorous extraction and analysis of data. This data will subsequently be provided to all lead authors for the exercise-relevant deliverables D6.3-D6.6) with PHE working to supervise interpretation and integration into deliverables.

By using this combination of template materials and light-touch, standardised training, PROACTIVE will be able to ensure a clear and consistent theme of evaluation runs through all three field exercises. This will help the consortium to build upon lessons learned for each iterative exercise in order to deliver tools and recommendations that are fit for purpose and have been well evaluated at the end of the project.

7. CONCLUSION

This deliverable represents the output from two parallel activities conducted as part of PROACTIVE Task 6.2: scenario development and specification of the evaluation methodology.

First, through a process of iterative review and stakeholder engagement, we have developed an initial prototype scenario that can be used to inform ongoing discussions between PROACTIVE and eNOTICE both in relation to the rescheduled Rieti exercise and future exercises (which can be found in Figure 3). This scenario can be adjusted, amended, extended, and tailored to fit a range of exercises and so represents a strong base from which to revise and develop the scenarios across the duration of the PROACTIVE project. Future work will involve close liaison with eNOTICE and exercise organisers as the projects progress in order to help develop these bespoke scenarios and tailored injects. The public-facing scenario will be finalised following discussion with representatives of the PROACTIVE Civil Society Stakeholder Advisory Board (CSAB) during the rescheduled September 2020 workshop.

Second, following a process of rapid evidence review and synthesis with subject matter expertise, we have developed an initial plan for exercise evaluation. Specifically, to ensure that the exercises carried out within the PROACTIVE project are as useful as possible, we will take a mixed-method approach to evaluating the exercises and will ensure that appropriate measures are used to capture different aspects of exercise play. For each exercise, we will: carry out a

a) hot debrief to capture the views of emergency responders, exercise planners, and project partners;

b) use pre- and post-exercise casualty feedback questionnaires to identify any changes in casualty volunteers' perceptions as a result of taking part in the exercise;

c) observe exercise play to identify any challenges that arise (particularly in relation to the management of vulnerable groups);

d) and carry out focus groups with casualty volunteers (and potentially interviews with individuals of interest to clarify key points) to facilitate an in-depth understanding of their experiences during the exercise and to identify any areas for improvement.

Depending on the exercise scenario, consideration will also be given to using physical quantitative measures (e.g. timing data, measures of efficacy) to evaluate the exercise. In addition to this initial specification, a detailed evaluation plan will be developed for each of the PROACTIVE exercises, and the evaluation methods used will be tailored to each specific exercise scenario.

8. REFERENCES

Alim, S., Kawabata, M., & Nakazawa, M. (2015). Evaluation of disaster preparedness training and disaster drill for nursing students. *Nurse Education Today, 35*(1), 25-31.

Allen, K., Lorek, E., & Mensia-Joseph, N. (2008). Conducting a school-based mock drill: lessons learned from one community. *Biosecurity and bioterrorism: biodefense strategy, practice, and science, 6*(2), 191-201.

Beaton, R. D., Oberle, M. W., Wicklund, J., Stevermer, A., Boase, J., & Owens, D. (2003). Evaluation of the Washington State National Pharmaceutical Stockpile* Dispensing Exercise: Part I—Patient Volunteer Findings. *Journal of Public Health Management and Practice, 9*(5), 368-376.

Carter, H., Drury, J., Amlôt, R., Rubin, G. J., & Williams, R. (2013). Perceived responder legitimacy and group identification predict cooperation and compliance in a mass decontamination field exercise. *Basic and Applied Social Psychology, 35*(6), 575-585.

Carter, H., Drury, J., Rubin, G. J., Williams, R., & Amlôt, R. (2012). Public experiences of mass casualty decontamination. *Biosecurity and bioterrorism: biodefense strategy, practice, and science, 10*(3), 280-289.

Carter, H., Drury, J., Rubin, G. J., Williams, R., & Amlôt, R. (2013). The effect of communication during mass decontamination. *Disaster Prevention and Management: An International Journal*, *22*(2), 132-147.

Charney, R. L., Lehman-Huskamp, K. L., Armbrecht, E. S., & Flood, R. G. (2011). Impact of disaster drills on caregiver perception and satisfaction in the pediatric emergency department. *Pediatric Emergency Care, 27*(11), 1033-1037.

Cicero, M. X., Whitfill, T., Overly, F., Baird, J., Walsh, B., Yarzebski, J., . . . Auerbach, M. (2017). Pediatric Disaster Triage: Multiple Simulation Curriculum Improves Prehospital Care Providers' Assessment Skills. *Prehospital Emergency Care, 21*(2), 201-208.

Cone, D. C., Serra, J., Burns, K., MacMillan, D. S., Kurland, L., & Van Gelder, C. (2009). Pilot test of the SALT mass casualty triage system. *Prehospital Emergency Care, 13*(4), 536-540.

Currie, J., & Heslop, D. J. (2018). Operational systems evaluation of a large scale multi-agency decontamination exercise. *International Journal of Disaster Risk Reduction, 31*, 1054-1061.

Dacey, G., Amlôt, R., Carter, H., Rob, C., Joe, E., Rosemary, H., . . . Lorna, R. (2011). *Exercise MILO report*. Salisbury: Health Protection Agency.

Dalby, D., Amlôt, R., Beasley, L., Egan, J., Carter, H., Goddard, H., . . . Shuttleworth, C. (2012). *Exercise GREYSTOKE report*. Salisbury: Health Protection Agency.

Digregorio, H., Graber, J. S., Saylor, J., & Ness, M. (2019). Assessment of interprofessional collaboration before and after a simulated disaster drill experience. *Nurse Education Today, 79*, 194-197.

European Network Of CBRN TraIning CEnters. (2020). Retrieved July 21, 2020, from https://www.h2020-enotice.eu/

Fertel, B. S., Kohlhoff, S. A., Roblin, P. M., & Arquilla, B. (2009). Lessons from the “Clean Baby 2007” pediatric decontamination drill. *American Journal of Disaster Medicine, 4*(2), 77-85.

FitzGerald, D. J., Sztajnkrycer, M. D., & Crocco, T. J. (2003). Chemical weapon functional exercise--Cincinnati: observations and lessons learned from a" typical medium-sized" city's response to simulated terrorism utilizing Weapons of Mass Destruction. *Public Health Reports, 118*(3), 205-214.

Health Protection Agency. (2008). *Exercise ORPHEUS II*. Salisbury: Health Protection Agency.

Health Protection Agency. (2009). *Exercise Kent 2009*. Salisbury: Health Protection Agency.

Johnson, V. A., Johnston, D. M., Ronan, K. R., & Peace, R. (2014). Evaluating children’s learning of adaptive response capacities from ShakeOut, an earthquake and tsunami drill in two Washington State school districts. *Journal of Homeland Security and Emergency Management, 11*(3), 347-373.

Johnston, D., Tarrant, R., Tipler, K., Coomer, M., Pedersen, S., & Garside, R. (2011). Preparing schools for future earthquakes in New Zealand: Lessons from an evaluation of a Wellington school exercise. *The Australian Journal of Emergency Management, 26*(1), 24-30.

Kaji, A. H., & Lewis, R. J. (2008). Assessment of the Reliability of the Johns Hopkins/Agency for Healthcare Research and Quality Hospital Disaster Drill Evaluation Tool. *Annals of Emergency Medicine, 52*(3), 204-210.e8.

Klima, D. A., Seiler, S. H., Peterson, J. B., Christmas, A. B., Green, J. M., Fleming, G., . . . Sing, R. F. (2012). Full-scale regional exercises: closing the gaps in disaster preparedness. *Journal of Trauma and Acute Care Surgery, 73*(3), 592-598.

Krieger, K., Jones, E., Rogers, B., Amlôt, R., Danielsson, E., & Johansson, C. (2014). *D8.15 Analytical report: Empirical evidence and analysis of the human behaviour data collection activities* [PRACTICE]; European Community’s Seventh Framework Programme.

Lange, R. T. (2011). Inter-rater Reliability. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 1348-1348). New York, NY: Springer New York.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology, 4*(2), 73-79.

Primeau, M. S., & Benton, A. M. (2019). Multilevel Disaster Simulation in Nursing: Lessons Learned in Undergraduate and Nurse Practitioner Student Collaboration. *Nursing Education Perspectives*. Advance online publication. doi: 10.1097/01.NEP.0000000000000602

Ramirez, M., Kubicek, K., Peek-Asa, C., & Wong, M. (2009). Accountability and assessment of emergency drill performance at schools. *Family & Community Health, 32*(2), 105-114.

Rehmani, R. (2005). Disaster drill at a university hospital. *Journal of Pakistan Medical Association, 55*(1), 28-32.

Stergachis, A., Wetmore, C. M., Pennylegion, M., Beaton, R. D., Karras, B. T., Webb, D., . . . Loehr, M. (2007). Evaluation of a mass dispensing exercise in a Cities Readiness Initiative setting. *American Journal of Health-System Pharmacy, 64*(3), 285-293.

Sweeney, B., Jasper, E., & Gates, E. (2004). Large-scale urban disaster drill involving an explosion: lessons learned by an academic medical center. *Disaster Management & Response, 2*(3), 87-90.

Taylor, K., Balfanz-Vertiz, K., Humrickhouse, R., & Jurik, C. (2008). Decontamination with At-Risk Populations: Lessons. *Internet Journal of Rescue and Disaster Medicine, 9*(1).

Turner, L., Amlôt, R., & Simpson, J. (2007). *Exercise YOUNG NEPTUNE: Mass Decontamination of Children Field Exercise*. Salisbury: Health Protection Agency.

Unver, V., Basak, T., Tastan, S., Kok, G., Guvenc, G., Demirtas, A., . . . Tosune, N. (2018). Analysis of the effects of high-fidelity simulation on nursing students' perceptions of their preparedness for disasters. *International Emergency Nursing, 38*, 3-9.

van Laerhoven, H., van der Zaag-Loonen, H., & Derkx, B. (2004). A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatrica, 93*(6), 830-835.

Vinson, E. (2007). Managing bioterrorism mass casualties in an emergency department: lessons learned from a rural community hospital disaster drill. *Disaster Management & Response, 5*(1), 18-21.

Voutilainen, A., Pitkäaho, T., Kvist, T., & Vehviläinen-Julkunen, K. (2016). How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing, 72*(4), 946-957.

9. APPENDICES

9.1. APPENDIX A: ORIGINAL SCENARIOS

Scenario 1 – Worst Case

- On a cold, wet late evening, you are travelling alone to meet a family member. As you are stood alone on the platform of an unfamiliar train station, an explosion occurs.
- Ears ringing, you look down to find that you are covered in a white powder. Quickly you feel your eyes start to sting and a cough rises in your throat. It is feeling increasingly difficult to breathe and your skin has started to burn.
- All around you, you can see your fellow passengers starting to suffer. Some have collapsed. Many have started to head for the exits to leave the station.
- As you're standing there, the first emergency responders begin to arrive. They're not really saying anything. They're just moving people around.

Scenario 2 – Very Bad

- It's lunch time on a frosty day and you're on your way to meet a family member for lunch.
- As you're walking through a shopping centre that you've never been to before, an explosion occurs.
- Looking down you find you are covered in white powder. You don't feel any different. The powder doesn't seem to be affecting you or the people around you. But many of them have started to leave the shopping centre.
- As you're watching, you see that the emergency services have started to arrive. They don't seem to be saying much to you or the others about what is going on. The emergency responders aren't helping you understand what's happening or what will happen next.

Scenario 3 – Middle

- It's early afternoon on a cloudy autumn day and you and a friend are walking into the centre of a town you've been to a few times before.
- As you're walking, there's a loud bang/explosion and you look down to find yourself covered in a White powder. Very quickly your eyes start to sting slightly.
- Looking around you can see other people who are covered in the powder and are starting to rub their eyes. Some of them are hurrying away from the area, but others are staying put as the emergency responders arrive.
- Their communication doesn't really seem to be either good or bad.

Scenario 4 – Not Worst

- It's 9am on a warm and dry summer day and you are wandering alone down your local high street where you often shop. You're in no hurry and are enjoying the walk.
- Suddenly there is a loud explosion. Looking down you see you are covered in a white powder. The air around you smells and tastes slightly strange. But you feel fine.
- The people around you all seem to have stopped and are staying put as the emergency services start to arrive.

- As you stand there with the others who have been part of the incident, the emergency responders begin to communicate with you all, explaining what is going on and helping you understand what will happen next as they help you.

Scenario 5 – Best Case

- It's midday on a sunny July day and you are doing your weekly food shop with your family at your local supermarket.
- Suddenly, there is a loud crash and you find yourself covered in a white powder. As you look around you, your eyes start to sting slightly, and you realise there is a strange taste and smell in the air. A few people around you seem to have started rubbing their eyes. But like you, they are standing still. No one is leaving the area.
- As you stand there, the emergency responders begin to arrive. They are quick to come and talk to you and the other people, giving you clear instructions and explanations.

9.2. APPENDIX B: SEARCH STRATEGY

1.	(Scenario or Simulation or Drill or Field exercise or Live exercise).ti.
2.	(Eval* or Analysis or Review or Feedback or Assessment or Observation or Debrief or Lesson*).ti.
3	(disaster* or weapons of mass destruction or Bioterrorism or CBRN or Chemical or Biological or Radiological or Nuclear or Terroris* or Emergency or Pandemic or Preparedness).ti.
4	1 and 2 and 3
5	limit 4 to (abstracts and english language and yr="2001 -Current")

9.3. APPENDIX C: SAMPLE CHARACTERISTICS

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
* Alim et al., 2015	Japan	Indonesia	Disaster drill (earthquake)	Post-drill interviews and observational assessment of medical staff	2	Unknown	Unknown	None (Medical students)
Allen et al., 2008	USA	USA	Emergency drill at schools	Observation of drills; survey of staff	Unknown	Unknown	Unknown	School children
Beaton et al., 2003	USA	USA	Medical dispensing drill from anthrax spray	Post exercise questionnaires for patients	214	65% were female	Mean age 41.5 years	Members of the public
Carter, Drury,	UK	UK	Large scale mass decontamination	Post exercise questionnaires for patients	402	Unknown	18 to - 85 years	Sample included members of the public

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
Rubin et al., 2012								
Carter, Drury, Rubin et al., 2013	UK	UK	Large scale mass decontamination	Post exercise questionnaires for patients	402	Unknown	18 to - 85 years	Sample included members of the public
Carter, Drury et al., 2013	UK	UK	Large scale mass decontamination	Pre- and post-drill questionnaires given to patients	115	62% were female	6 – 69 years, mean of 31 years	Sample included members of the public
Charney et al., 2011	USA	USA	Disaster drill at a hospital	Post-exercise survey of caregivers	22	Unknown	Unknown	Caregivers accompanying children in drill
* Cicero et al., 2017	USA	USA	Pediatric Disaster Triage	Observational assessment of medical staff	4	Unknown	Unknown	One child with special needs

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
Cone et al., 2009	USA	USA	Aircraft crash at airport	Performance assessment of medical staff	52	Unknown	Unknown	Drama club students
Currie et al., 2018	Australia	Australia	CBRN drill at a hospital	Survey of all participants; one week follow up phone call	46	36 were female	Mean age 34 years	None (Medical students)
Dacey et al., 2011	UK	UK	Large scale mass decontamination	Observations, debriefs and questionnaires for staff; focus groups and trackers for casualties	96	52% were male	18 – 60 years of age	37 casualties had a disability

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
Dalby et al., 2012	UK	UK	Large scale mass decontamination	Observations, debriefs and questionnaires for staff; focus groups and trackers for casualties	107	53% were male	18 – 70 years of age	8 casualties had a disability
* Digregorio et al., 2019	USA	USA	Train derailment and chemical spill	Debrief; assessment via performance measure for medical staff	120	Unknown	Unknown	Unknown
Fertel et al., 2009	USA	USA	Chemical spill	Observation of drills and mood of children	20 children (plus 5 caregivers)	Unknown	0.25 – 15 years	Young children
Fitzgerald et al., 2003	USA	USA	Large scale mass decontamination	Lessons learnt exercise	85	Unknown	Unknown	Members of the public, including children

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
Health Protection Agency, 2008	UK	UK	CBRN drill	Structured observations; debriefs; post event questionnaires for staff and casualties	60	58% were female	17 – 85 years of age	None (professional actors, medical students)
Health Protection Agency, 2009	UK	UK	CBRN drill	Structured observations; debriefs; post event questionnaires for staff and casualties	150	Unknown	Unknown	Unknown
Johnson et al., 2014	USA	USA	School earthquake and Tsunami drill	Pre- and post-drill questionnaires given to children	428	Unknown	10 years or older	Children
Johnston et al., 2011	New Zealand	New Zealand	School earthquake drill	Lessons learnt exercise	200 children	Unknown	5 – 13 years of age	Children

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
* Kaji et al., 2008	USA	USA	State wide CBRN drill	Observation of drills	Unknown	Unknown	Unknown	Unknown
Krieger et al., 2014	UK	UK	Large scale mass decontamination	Pre- and post-questionnaire and focus groups with casualties	40	50:50 ratio	19 – 69 years, mean of 37.8 years	Members of the public
Klima et al., 2012	USA	USA	Train derailment and chemical spill	Observation of drills	281	Unknown	Unknown	Members of the public
* Primeau et al., 2019	USA	USA	Disaster drill	Post drill survey given to medical staff	Unknown	Unknown	Unknown	Undergraduate nursing students
Rehmani, 2005	Pakistan	Pakistan	Aircraft crash drill	Debrief post session	30	Unknown	Unknown	Members of the public

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
* Ramirez et al., 2009*	USA	USA	Emergency drill at schools	Questionnaires/focus group; Observations; time measurements	18,211; including one adult school	Unknown	Unknown	School children
Stergachis et al., 2007	USA	USA	Medical dispensing drill from anthrax spray	Post exercise questionnaires for staff and patients	110	Unknown	One month – 92 years, mean age 31.3 years	Some children and elderly volunteers
* Sweeney et al., 2004	USA	USA	CBRN drill	Lessons learnt exercise	240	Unknown	Unknown	None (medical students)
Taylor et al., 2008	USA	USA	CBRN drill	Interviews and focus group with casualties	45	53% were male	Unknown	37 casualties had a disability

Short study citation	Institution location of the first author	Location of exercise	Incident	Evaluation method	Number of casualties in sample	Gender	Age	Relevant characteristics
Turner et al., 2007	UK	UK	Large scale mass decontamination	Reports from observers, medical staff; questionnaire given to medical staff and casualties	65 children	Unknown	6 – 14 years	Children
* Unver et al., 2018	Turkey	Turkey	Multiple disaster drills	Questionnaire given to medical staff; assessment via performance measure	20	All female	Unknown	None (medical students)
* Vinson, 2007	USA	USA	CBRN drill at a hospital	Lessons learnt exercise	Unknown	Unknown	Unknown	Student volunteers from a local college

*** = rapid evidence review paper**