International Journal of Disaster Risk Reduction 50 (2020) 101695

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

Decision support for medical disasters: Evaluation of the IMPRESS system in the live Palermo demo

Fabio Cibella [a,*], Simona Panunzi [b], Valerio Cusimano [b], Andrea De Gaetano [b]

[a] National Research Council of Italy, Institute for Biomedical Research and Innovation, Palermo, Italy
[b] National Research Council of Italy, Institute for System Analysis and Computer Science, Biomathematics Laboratory, Rome, Italy

ARTICLE INFO

Keywords:
Validation of Decision Support System
Live exercise
Crisis management

ABSTRACT

Background: In medical disasters, coordination, information flows, and decision making are crucial for response and management. Different factors contribute to thwart the response efforts. Some are due to the coordination of the many agencies active in disaster response. Support tools for gathering and analysing data may support task assignment, resource allocation, and acquisition as well as training at different decision levels (in the field and in command-rooms). Validation of Decision Support Systems (DSS) in simulated contexts, simulating real situations, becomes mandatory. In the framework of testing and validation of the IMPRESS project (and of its INCIMOB and INCIMAG tools), one scenario was planned in Palermo, a city of 700,000 inhabitants in the Mediterranean Area of Southern Italy, simulating the sudden liberation of high concentrations of toxic compounds from a fire in Palermo harbor. Emergency Agencies, a real and a simulated Hospital and operators in the field used the system during the response phase. A group of 20 external Observers participated for evaluation purposes. During a joint debriefing session, ad-hoc questionnaires were administered. IMPRESS was useful in improving the execution of important functions during the DEMO; Users agreed about the advantages of the use of IMPRESS tools for conducting crisis activities. INCIMOB use resulted more problematic from an operational point of view. Shortcomings were detected and criticisms were raised due mainly to the lack of training and direct voice communication. Evaluation of DSS in Emergency medicine can benefit from live exercises to highlight weaknesses in both the response system and decision support.

1. Background

Coordination, information flow, integration, and timely decision making are crucial for an efficient response to and management of medical disasters [1]. A number of factors contribute to thwart the response efforts, especially in connection with the necessary coordination of the many agencies having an active role in disaster response. For early preparedness it is necessary that the rescue personnel know their roles and act according to standard operating procedures beginning with the arrival of the very first victim [8]. Mass casualties create a tremendous amount of stress for any medical system, thus, standard operating procedures and a Decision Support System (DSS) constitute a fundamental guidance for involved personnel [22]. In the case of health crises and critical health events, Kulling et al. proposed to realize guidelines aimed at promoting standardized methodologies for facilitating analysis comparison of findings in order to improve preparedness and response planning [19]. Support tools for gathering, processing, and analyzing data can be a useful instrument for task assignment, resource allocation and acquisition, as well as for training purposes, at different decision levels (in the field and in the command and control rooms). The validation of DSS in simulated contexts, as similar as possible to likely real situations, becomes a mandatory task.

To this end, the application of Use Case Scenarios can be an optimal solution to showcase the ability of the DSS to take into account organizational procedures, roles and functions within the Emergency Management System, as well as those design features responding to the requirements and preferences expressed by stakeholders and experts with a view to address existing shortcomings of current health emergency response [20].

A conceptual framework to measure the performance of a response system during emergency exercises has been described by Savoia [5]. As it can be easily seen from the following description, it is strongly tailored to the DSS validation framework described by Lamy [9], even though it seems to better reflect real life events, such as those which are to be tested in case of emergency response. In this context the use of

* Corresponding author.
E-mail address: fabio.cibella@cnr.it (F. Cibella).

https://doi.org/10.1016/j.ijdrr.2020.101695
Received 9 January 2020; Received in revised form 20 May 2020; Accepted 25 May 2020
Available online 26 June 2020

List of acronyms

IMPRESS	IMproving Preparedness and Response of HEalth Services in major criseS, EU FP7-SEC Project
CRI	Italian Red Cross Provincial Command
DHC	Data Harmonization Component
DSS	Decision Support System
DVI	unit for Disaster Victims Identification of State Police
INCIMAG	Incident Management, a tool and relevant environment for managing emergency incidents (desktop solution)
INCIMOB	The mobile extension of the INCIMAG system for on-field operations, patient tracking, receiving notifications, etc.
KPIs	Key Performance Indicators
LOGEVO	Model based component for the prediction of the provision of health care resources
PATEVO	(PATient EVOlution) a model based platforms allowing the forecast of the evolution of the vital physiological functions of the victims of the crisis, determining the time-curve of each Physiological State Variable, based on a modelling analysis through which both the effect of injuries and the effect of administered treatments are described
PEIMAF	Emergency hospital plan for major crisis
PMA	Advanced Medical Post
SORIS	Regional Department of civil protection
SORLOC	(SOuRce LOCalization) a tool for estimating time, location and spatial extent of a biological release, based on data from the first few cases presenting at hospitals or other health-care facilities, as results of a disease outbreak
SUES 118	Emergency Health Service
VVF	Provincial Fire Department
WARSYS	the IMPRESS data base, which provides the interfaces to import data from medical and logistics repositories (such as hospital information systems); further it can be used to store and view the extracted data and information generated inside the IMPRESS system

benchmarking is mandatory. Benchmarking of a DSS is usually referred to the study of the performance of the system, using as Key Performance Indicators (KPIs), for example, the throughput of information during an established period of time, or metrics quantifying how the system reacts to a considerable increment in the demands placed on it [21].

Evaluation of DSSs is difficult because of their dynamic nature [4]. Kim [3], in a fundamental historical article, described four methods for evaluating a DSS: the cost benefit analysis, the value analysis, the multi-attribute method and the Analytical Hierarchy Process Method. A combination of these methods has often been applied. All these methods are still employed at a conceptual level: most modern techniques for evaluation rely however on the definition of benchmarks for DSSs through a complete computing environment where a population of users with different functions executes tasks against a database utilizing the DSS to be validated. The benchmarks are focused on the main activities performed with the help of the DSS (questions/answers, recordings, checking status of conditions, monitoring levels of resources, etc.), according to predefined KPIs.

It must be remembered, on the other hand, that it is exceedingly uncommon for jurisdictions and agencies to use common evaluation tools or metrics. Evaluation of public health system exercises is frequently done *ad-hoc*, and the data gathered is often narrative in form, rather than quantitative and standardized. To date, in the US public health system there are no recognized benchmarks of performance of agencies for emergency situations, hence it is very difficult to define the benchmarks for the DSSs to be employed during these events [6].

It is clear that benchmarking of decision support systems for disaster situations is an open problem, that no general consensus exists, that in the existing literature KPIs vary greatly in the number and types of elements to be considered, and that there exists a lack of consistency among the KPIs which are considered in different reports. Basically in the health management system response to disasters, two domains have been identified for the evaluation of performance, one relative to the Emergency Medical Systems in the field and one relative to the Hospital Response, this last being usually referred to as Hospital Surge ([10,11]).

Despite some questions on using KPIs as measures of system performance in health care settings have been recently raised [18], balanced sets of KPIs are considered useful for administrators and decision makers to establish comprehensive performance measurements [21].

Ingrassia [20] has published a Disaster Simulation Suite (iNovaria, Novara, Italy), which is a computerized system to evaluate training in case of simulation exercises of mass casualty situations. The system utilizes as benchmarks the key times in the Pre-Hospital phase (time to first triage, time to first move, time to treatment, length of stay in the scene) and in the Hospital phase (time to ED triage, time to first medical assessment, emergency department length of stay). These KPIs are derived from those proposed by Green [7] for evaluating disaster drills in developing countries.

Real-time scenarios (table-top exercises) seem to be the best choice for evaluating real time actions, since the time spent when performing any task can be gauged appropriately in the field. In such an exercise, the input data and the outcome are predictable and reflect the actual situation; in this way the outcome of the scenario can be appropriately evaluated at the same time that the exercise affords an opportunity to extend the participants' knowledge and hone their skills. In fact, in scenario-based simulations the input can be programmed and the output measured: the simulation coordinator can adjust the intensity, severity and content of the exercise to adapt it to the experience of the participants; the simulation is an effective opportunity to combine teaching, exercise and training for the different operational phases (command, control, communication); the consequences of each decision made can be identified and discussed. Starting from the assumption that real word and its problems and systems are dynamic, Paul and Hlupic argue that therefore models cannot be validated against the real-world system they represent, as the real-world system itself is not static [14]. While such a viewpoint is certainly debatable, it does emphasize the difficulty of validating any simulation model and the need to situate the system to be validated in a context that is as near as possible to the real situation, testing it in a framework where casuality, human error, and unexpected developments condition the course of the events. Only a live exercise, involving for example several agencies with people and aerial, maritime, and terrestrial resources, as it would happen in a real emergency situation, could provide an effective test of a system set up for tackling crisis situation. Conversely, table top exercises - certainly less troublesome as concerns the evaluation process, and which are useful for the reasons reported above - cannot reproduce the effects of all the variables potentially interfering with the operational procedures and result therefore insufficient to demonstrate the real usability, usefulness and efficacy of a DSS in the real world. However, the evaluation of DSSs in live exercise presents serious issues and requires the use of a valid approach. This is due to the impossibility to perform a rigorous statistical comparison of quantitative indicators collected during parallel, randomized, prospective, blind testing procedures, also comparing results with and without the use of a DSS. Moreover, a lack in literature of rigorous methods of evaluation makes the task even more difficult.

Bearing all these aspects, for the testing and validation of the IMPRESS system two different test cases were foreseen. The first one was a table-top simulation based on an earthquake scenario at E79 motorway near Greek-Bulgarian border. The second test case, which chronologically was held before, the one we are concerned with in the present work, was a live exercise, planned in Palermo, a city of 700,000 inhabitants located in the Mediterranean Area of Southern Italy. The simulation involved the sudden liberation of high concentrations of toxic compounds from a fire developing on-board of a large ship moored in the Palermo harbor. Several Emergency Agencies, about 20, one real and one simulated Hospital, and hundreds of operators in the field used the system during the response phase.

In such a complex framework it was mandatory to develop an ad-hoc methodology for the IMPRESS evaluation: it involved the use of a set of KPIs; the set up of an Evaluation Committee, composed by a group of high-level external expert evaluators involved in crisis management and belonging to different European Agencies; the involvement of a group of 20 Observers from different agencies who participated in the demo for evaluation purposes; the involvement of personnel from the different agencies who used the system during the demo; organization of a debriefing session where all Observers and system users participated in and filled ad-hoc questionnaires.

The aim of the present work is twofold: to describe the entire framework in which the IMPRESS system was evaluated, that is the live exercise itself, along with the organizational plan put in place (information campaign, meeting …) for the Demo preparation (Pre-event phase); to show the approach set up for the evaluation of IMPRESS in the live setting, outlining all the steps performed and the questionnaires prepared for the event.

The organizational procedure and the evaluation framework could represent a guide for the organization of such a live region-wide exercise and a possible feasible way for the evaluation of DSS systems in real scenarios, going further the classical table-top exercises.

2. Methods

2.1. The IMPRESS solution: logical description

During an incident management stakeholders and in particular emergency health service providers have to deal with two basic challenges: the disproportion between the needs and the available human/material resources in the response capacity and the inherent time constraints of an emergency. These critical factors play a crucial role in the decision-making process during a crisis event, which affects all levels of command & control (strategic, operational, tactical).

The European IMPRESS project was aimed at designing a platform capable of supporting decision-making process during a health emergency event.

The basic components and modules embedded into the IMPRESS solutions are reported below:

- WARSYS, the IMPRESS data base, which provides the interfaces to import data from medical and logistics repositories (such as hospital information systems); further it can be used to store and view the extracted data and information generated inside the IMPRESS system.
- Reference Semantic Model, which defines an ontology related to the health emergency management domain.
- Data Harmonization Component (DHC), which achieves the harmonization and the homogenization of the data of different structure and nature, provided by the WARSYS component and coming from a variety of external sources.
- LOGEVO, a models based component for the prediction of the provision of health care resources.
- SORLOC (SOuRce LOCalization), a tool for estimating time, location and spatial extent of a biological release, based on data from the first few cases presenting at hospitals or other health-care facilities, as results of a disease outbreak.
- PATEVO (PATient EVOlution), a model based platform allowing the forecast of the evolution of the vital physiological functions of the victims (individuals involved in the mass casualties incident) of the crisis, determining the time-curve of each Physiological State Variable, on the basis of a modelling analysis through which both the effect of injuries and the effect of administered treatments are described. The model predicts the evolution of ten Physiological State Variables describing some human vital functions in agreement with the ABCDE Primary Survey and Resuscitation [2]. The evolution of the physiological variables is determined by the effect of “Assets” (on the scene, ambulance, Emergency room, etc …), which deliver treatments (oxygen, cortisone, surgery, etc …) in order to improve individual’s conditions and to restore the patients’ physiology towards pre-crisis levels.
- INCIMAG (Incident Management), a tools and environment to manage emergency incidents (it is a desktop solution).
- INCIMOB, the mobile extension of the INCIMAG system for on-field operations, patient tracking, receiving notifications, etc.
- Training Component for online training purposes of potential users of the IMPRESS system.

For a more detailed description refer to Ref. [16,17].

At the bottom level of the architecture is the Data Storage, which consists of WARSYS and the DHC. While WARSYS collects the data (via interfaces to external systems utilizing standard data formats such as HL7-RIM or EDXL-HAVE), the DHC harmonizes it and other types of data originating from the Web (utilizing the Semantic Reference Model). On the top level is the Incident Management. It consists of the INCIMAG user interface (desktop application) and the INCIMOB (mobile application). Several incident management instances can be connected to each other to achieve a distributed crisis management. External medical devices (like pulsimeter, sphygmomanometer, or glucose meter) deliver their data to INCIMOB, which in turn passes this information to INCIMAG and consequently to WARSYS. Responders can access the decision support systems (LOGEVO, PATEVO and SORLOC) through INCIMAG and INCIMOB to be facilitated and able to make more informed decisions according to the recommendations of the DSS engines. A schematic representation of the IMPRESS architecture is presented in Fig. 1.

2.2. Overview of the exercise

The first validation of the IMPRESS system was performed by means of a real test scenario. The test case took place in Palermo, a city of 700,000 inhabitants, located in the south Mediterranean area and simulated the sudden liberation of high concentrations of toxic compounds from a tank fire developing on-board of a ship moored in front the Palermo harbor, close to the Palermo promenade. The severity of the fire was high since it was impossible to extinguish it by means of on board equipment. The ship had mainly a cargo of plastic materials that, during the combustion, released different toxic substances. The presence of a wind from North East (a very common wind in Mediterranean sea) carried the toxic cloud in the direction of the "Kalsa" District, a densely populated area of Palermo. Moreover, victims were on board the ship with burns, wounds and crushed limbs, while others fell into the sea while trying to abandon the ship. In the urban coastal area impacted by the fire scenario, several institutional buildings are located: a school (The Palermo Nautical Institute), an office of Harbor Authority, the Central Regional Administrative Court, the Tax Office, the administrative offices of the University of Palermo, and a church. All these institutions were involved in scenario development through meetings in which aspects relevant to the scenario were presented and discussed in the weeks preceding the demo. The areas immediately outside that involved by the contamination were set as suitable for triage facilities and for the landing of helicopters for the evacuation of most severe victims. The nearest hospital was 2 km away, accessible directly through the seafront boulevard. Another important element of the Scenario was

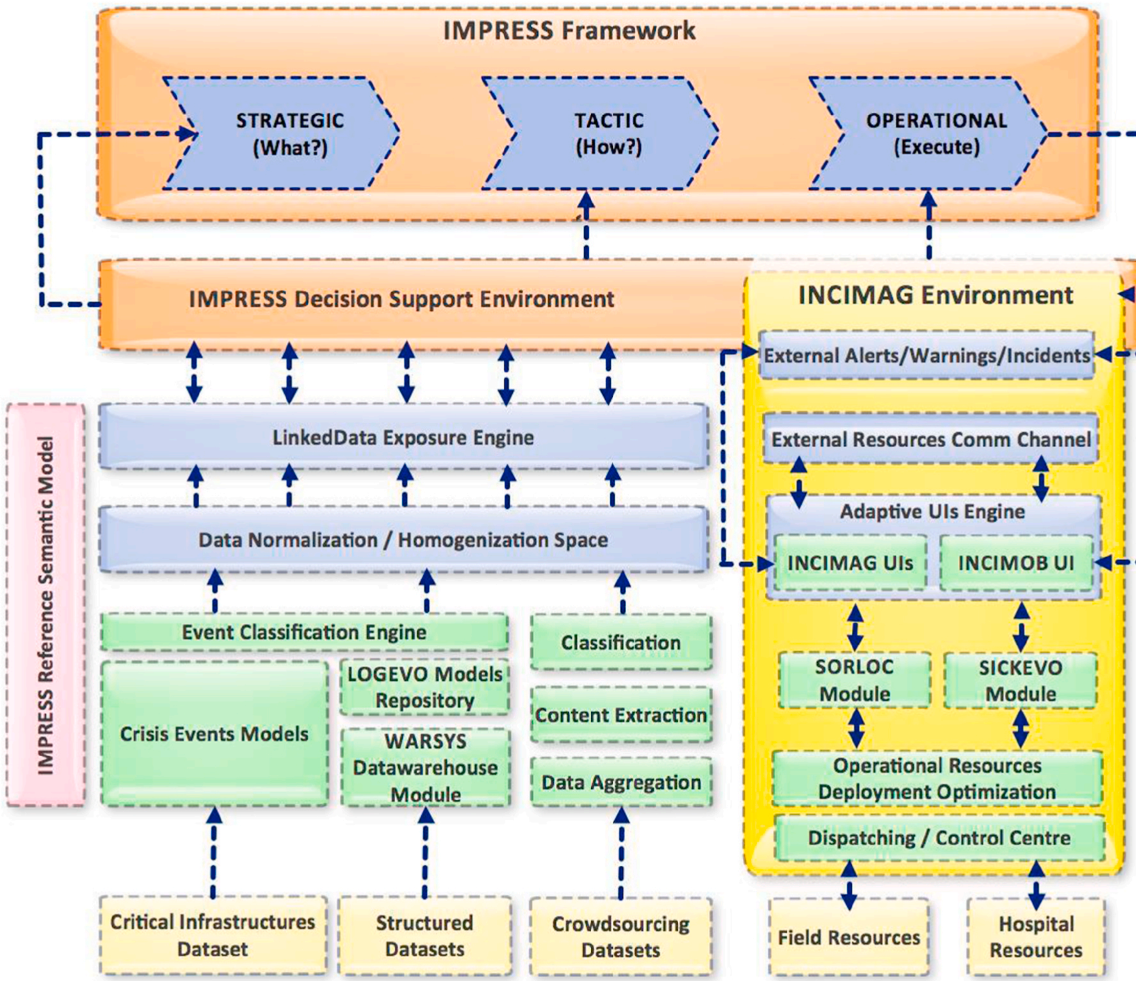

Fig. 1. Schematic representation of the IMPRESS architecture. INCIMAG: Incident Management, a tools and relevant environment for managing emergency incidents (desktop solution). INCIMOB: the mobile extension of the INCIMAG system for on-field operations, patient tracking, receiving notifications, etc. SORLOC (SOuRce LOCalization): a tool for estimating time, location and spatial extent of a biological release, based on data from the first few cases presenting at hospitals or other health-care facilities, as results of a disease outbreak. SICKEVO (SICKness EVOlution) and PATEVO (PATient EVOlution): model based platforms allowing the forecast of the evolution of the vital physiological functions of the victims of the crisis, determining the time-curve of each Physiological State Variable, based on a modelling analysis through which both the effect of injuries and the effect of administered treatments are described. LOGEVO: a model based component for the prediction of the provision of health care resources. WARSYS: the IMPRESS data base, which provides the interfaces to import data from medical and logistics repositories (such as hospital information systems); further it can be used to store and view the extracted data and information generated inside the IMPRESS system.

the traffic conditions. The crisis happened in a geographic area characterized by narrow roads and heavy traffic, mainly during working hours, while the seafront boulevard presents a heavily congested traffic during working hours. A forecast of the diffusion of the toxic cloud was carried out by means of ALOHA (Areal Locations of Hazardous Atmospheres, version 5.4.5, United States Environmental Protection Agency), widely used to plan for and respond to chemical emergencies. Meteorological data such as wind direction, wind speed, wind height above ground, cloud cover, average air temperature, humidity, etc. were also used for a better prediction of the diffusion of the toxic cloud. The Palermo test Scenario allowed both the demonstration and validation of the main features and advantages of the use of the IMPRESS-solution and the identification of limitations and shortcomings as well as of possible area of improvements.

2.3. Pre-event phase

2.3.1. Event organization phase

The planning activities for the IMPRESS Palermo Demo started in 2015. It still represents the largest regional exercise of Civil Protection ever realized in Sicily, and one of the largest in Italy, for the number of involved agencies and for deployed human and material resources. The demo took place on June 7th, 2016 in the morning hours (from 8:00 a.m. to 2:00 p.m.). A planning document was firstly produced indicating the involvement of each Agency. Different meetings were held to define the demo storyboard and the exercise design, to quantify and coordinate the available (human and material) resources from each Agency. The material resources deployed on the field were three Advanced Medical Posts (PMA), two Nuclear, Biological, Chemical and Radiological NBCR units, one Accommodation Area, DVI unit (Disaster Victims Identification of State Police), ten ambulances, one PEIMAF (emergency hospital plan for major crisis), one CRI (Italian Red Cross Provincial Command) medical team, one VVF (Provincial Fire Department) BLSD team on board, gates of the security cordon around the involved area (controlled by State and Municipal Police), twenty Info points. A total of about 500 people from different agencies were involved. The agencies equipped with the INCIMAG tool were:

- Coast Guard: as the first agency involved in this emergency that commands and manages the rescue at sea. The Coast Guard is the agency that first collected the information about the crisis (type of emergency, severity, involved victims).

- Prefecture of Palermo, the Government agency which activates the Crisis Unit, representing the core of the emergency coordination, where officers from all the other agencies are convened.
- Regional Department of civil protection (SORIS) which after receiving the alert activates the internal procedures to support the emergency.
- Emergency Health Service (SUES 118) Control Room: the agency that manages the entire health emergency and coordinates all the available health resources (hospitals availability, ambulances, emergency departments). SUES 118 has a physician on the field communicating in real time with the control room and taking all the decision according the priorities. It is the Agency that avails of the Decision Support System of IMPRESS for a better management of the casualties and for an optimization of the resources.
- Buccheri La Ferla hospital: the real hospital involved in the demo, who activated its internal PEIMAF.

The agencies equipped with the INCIMOB devices were:

- Fire Brigade: this agency includes some BLSD operators that, during an emergency, may provide a first support. During the demo a team of two BLSD operators, sailing on a rescue vessel, used one INCIMOB for the rescue of the first victims.
- CRI: one INCIMOB was handled by CRI medical team on board of a Coast Guard patrol boat to rescue victims from the sea.
- Emergency Health Service (SUES 118): three INCIMOBs for the physicians of emergency, one for each of the three PMAs.
- SUES 118, CRI and Palermo Provincial Health Agency (ASP): one INCIMOB in each of the ten ambulances.
- Buccheri La Ferla hospital: one INCIMOB at the emergency department and one at the PEIMAF.

An incident profile (individual cards) was created for each one of 20 mock victims in the ship and in the sea, along with further 20 incident profiles relevant to mock victims among population due to the inhalation of toxic smoke or accidents caused by panic (for a total of 40 incident profiles). Profiles were defined on the basis of the library of the ten Physiological State Variables, whose evolution is predicted by the PATEVO DSS component. For each victim profile a maximum survival time was determined in the absence of needed intervention/therapy: each mock victim was provided with a waterproof sheet indicating the requested interventions and the possible death occurrence if an intervention was not provided in time. Individual cards were used by CRI operators, to put makeup on the actors (moulage) on the day of the Demo, according to each victims profile. Incidents profiles were designed to ensure a complete spectrum of severity for each mock victim, to account for changes in patient characteristics as a result of different combinations of interventions [12]: they included i) victims found dead; ii) subjects that would have died despite prompt medical intervention due to the severity of the lesions; iii) subjects who would have died if adequate life support had not been ensured within an adequate time-span; iv) subjects who would have arrived alive at the Emergency Department of the hospital; v) and finally victims not needing any treatment.

2.3.2. *Validation framework and benchmarking*

The evaluation framework was aimed at the following objectives: i) minimizing the gaps between the user requirements and the design and implementation of the IMPRESS solution; ii) demonstrating the IMPRESS solution to possible final users during a real exercise. The evaluation methodology followed an iterative process, where improvements and modifications at each step led to the next version of the product, converging towards a configuration fulfilling most of the needs of the users and stakeholders (Fig. 2).

The first evaluation of the IMPRESS subcomponents (i.e., communication platform, data warehousing, DSS subcomponents) allowed to demonstrate the IMPRESS interoperability and functionalities to the end-users in their real environment, providing end-users a chance to assess the added value of IMPRESS. The test and evaluation process taking place during the Palermo Demo addressed the following two different aspects:

- Finding inconsistencies and unintended bugs (related to technical aspects);
- Getting feedback about the usefulness of the platform (related to the interaction user-platform).

While the first aspect was important for bug reporting, the second

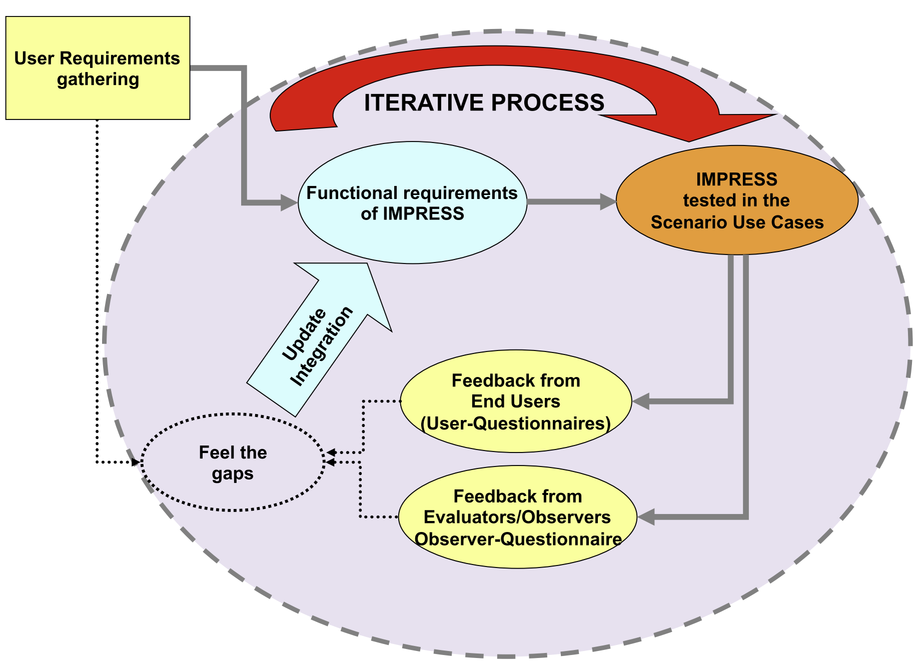

Fig. 2. The IMPRESS Evaluation - Improvement process.

one had relevance for producing improvements in IMPRESS usability and usefulness.

In order to evaluate a DSS in decision making, we should be able to compare consequences of decisions taken without the aid of the DSS with those that would have been taken with the aid of the DSS. In complex situations that are unique, irreproducible or that occur rarely, as with emergency situations, validation of a DSS output becomes extremely difficult. Sometimes it is possible to make use of historical data, but very often data are inexistent or incomplete. Another method could be that of comparing the performance of the proposed DSS to existing decision support tools, but they should be addressed to the same objectives and should be evaluated in terms of the same outputs ([15, 23]). According to Zeigler however, the only way these tools can be validated is by placing them in the context of its intended use. Following Lamy [9] the evaluation of a DSS should happen in two phases: testing the DSS in controlled conditions; evaluating the DSS in real use, during a randomized trial.

Validation should ideally be done through statistical comparison of KPIs collected during parallel, randomized, prospective, blind testing procedures, one including IMPRESS, the other excluding it. This was however impossible to do, also because of the impossibility of repeating the same live region-wide exercise with and without the use of the IMPRESS platform. For this reason an ad-hoc validation assessment was conducted based on the following elements:

1. A group of high-level external expert evaluators was identified (the Evaluation Committee). They were experts in crisis management from different European Agencies (EKAB, Athens prehospital EMS, Greece; GAGS, the HZM-incident officers, the Netherlands; CAA “Area Chirurgica”, DEAS, IRCCS San Raffaele, Milan, Italy; SIMNOVA - “Centro di Simulazione in Medicina e Professioni Sanitarie”, University of Piemonte Orientale, Novara, Italy) and helped with the contents and methodology of the validation process. In particular, the evaluators contributed to the definition of target criteria and to the development of questionnaires to be administered during the post-exercise phase;
2. A number of senior professionals, representing the several specialties involved in the management of a crisis (for example experts from the field of Emergency response at the Ministry of the Interior and at the Center for reaction to crises), were invited to participate in the demo as “Observers”, placed at different strategic points (on the field and at decision stations), whose task was that of annotating the main functionality-related issues. The “Observers” were therefore persons working at the tactical/operational level as well as at a strategic level, belonging to those organizations who are involved in the response phase should similar crises happen. An “Obsever Oriented Questionnaire” was prepared and administered to the Observers;
3. A “User Oriented questionnaire”, one for INCIMAG and one for INCIMOB users, was administered to the users of the IMPRESS System immediately after the end of the demo;
4. A set of quantitative, automatically recorded benchmarks were defined, the KPIs, Table 1. They should have been all collected during the demo but at the time the Palermo Demo occurred, not all the indicators were already been included into the IMPRESS platform to be automatically recorded. Quantification of some of them are reported in Table 9, and refers to the number of patients tracked/triaged with the IMPRESS system. Since their measurements during the active IMPRESS demo could not be compared with corresponding measurements in a control demo, no statistically valid conclusions could have been drawn with respect to the efficacy of IMPRESS in modifying them. However, the analysis of the resulting scores gave indirect, supporting evidence and/or suggested specific weaknesses of the system;
5. Each active operator with decision roles and many active operators with supporting roles were matched with a Ghost, an operator from the same kind of service or agency, who was trained in the use of the IMPRESS platform. The reason for performing this first demo without each active operator having direct access to the IMPRESS hardware was that operators would have been too busy with the emergency operations to be able to afford the time needed to handle another information device. Supporting each operator with a human Ghost, able to suggest to the active colleague what modifications to the normal routine may be indicated by IMPRESS, would have provided proof-of-concept that extra information was indeed useful. After the demo, therefore, each Operator-Ghost couple was asked to fill in the corresponding User Oriented Questionnaire.

The questionnaires represented the main instrument to gather feedback from the users and evaluators and gathered information about the following aspects: usefulness of IMPRESS in the execution of important functions during the crisis, compatibility with procedures, usability, shortcomings, dysfunctions, the perceived efficacy of the system, limitations occurring in the use and in the functionalities, suggestions for possible improvements. Four versions of the evaluation questionnaire of the Palermo Demo were prepared: one was addressed to the INCIMOB users, one to the INCIMAG users, one to the ICIMOB observers and one to the INCIMAG observers. The User INCIMAG (see Appendix A) and ICIMOB (see Appendix B) questionnaires differed only for a set of questions concerning the utility of having employed the system in the demo, addressing specific questions about the usefulness of IMPRESS focusing either on the activities proper to a Command/Control Station or on the activities related to the field response (first set of questions). The User and Observer versions differed instead only for a series of questions addressing the usability of the system and that were included only in the User questionnaire. A Statistical Report containing descriptive summarizing statistics of the questionnaire scores and KPIs was then produced. The Evaluation Committee reviewed the Statistical Report and formulated its conclusions on the performance of the IMPRESS platform in a Final Report.

2.3.3. *Information campaign*

During the planning of an exercise implicating the population over an urban area it is very important to provide correct information to prepare the people about the event, in order to avoid real accidents due to panic: it does happen in fact that, not notwithstanding the repeated announcements, some inhabitants mistake the exercise for a real emergency and behave accordingly. The Kalsa district involved in the Palermo Demo is located in the historical center of the city. It covers an area of 0.8 km^2 with about 128,000 inhabitants. The buildings are ancient and the streets are narrow and tortuous. The population is mixed: Sicilian people and non-EU citizens are present. Furthermore, many tourists are present due to the many attractions of the district (museum, old churches, gardens, and so on) and to its proximity to the harbor. For the above-mentioned reasons, a diffuse information campaign was activated some days before the demo. The information campaign was developed as follows:

1. Joint press conference of the Prefect and the Mayor of Palermo on June 3rd in the Prefecture. Reporters of national and regional newspapers and television attended the press conference and the resulting information was reported by the media over the following days. Representatives of each agency participated in the press conference, during which the organizers described the IMPRESS project and the platform to be tested during the demo.
2. Posters and brochures were distributed during the week before the demo in most public places (bars, restaurants, shops, markets, hotels, churches, etc.) located in the district and in its proximity, in all the schools and offices within the district and in the agencies involved in the demo. The information provided by the posters and brochures concerned the date of the exercise, the simulated emergency, the area involved by the demo, and indications about areas and streets off limits for vehicle traffic.

3. Two training meetings were planned to inform the general population and the safety personnel in the public buildings located in the red area, providing in particular basic information about the procedures to be activated in case of toxic cloud, panic management during an emergency, basic procedures for first aid.

2.3.4. First training phase

On May 31st the first online training session took place in Palermo. Throughout the session, the trainees had the opportunity to use the training platforms, examine the core functionality, learn how to browse, download and view the available teaching material, as well as how to utilize the collaborative features of the online environment. For the meeting 20 computers with internet access were available to navigate the online training platform: 40 trainees participated in the online training session. All instructions given in English by the IMPRESS technical trainers were simultaneously translated into Italian language for the Italian native speaker trainees. Each trainee accessed the training platform using their unique username and password that were prepared beforehand by the Training Suites System Administrator. The structure of the online course was explained and the capabilities of the platform listed. The last session of the training day was dedicated to questions about the use of the online platform. During the session the INCIMOB application was showed to the responders (field operators).

2.3.5. Physical training phase

On June 6th, 2016, a physical training was organized. The audience was composed of end-users involved in the day after exercise: medical and paramedical personnel as well as responders from Coast Guard, firebrigades, Civil Protection. For the training session, 20 smartphones and tablets, equipped with INCIMOB application and mobile data connectivity, were handed out to the responder teams in order to allow them to familiarize themselves with the tools. A session was dedicated to the technical concepts and the functions of the INCIMAG command & control platform.

2.3.6. Ethical aspects and data protection

All the ethical issues that arose or that might have arisen in the context of the Palermo Demo were analyzed in depth, including those relevant to participant data protection. No volunteers belonged to a vulnerable group and all participants were able to give their signed informed consent. All volunteers were informed about the project, its nature and purpose and received documentation in a language they could understand, with particular attention to their right to be able to withdraw at any time and to the insurance coverage. No participant was pressured into participating. Sensitive information was not stored, unless data relevant to the mock victims used for the demo purpose. Finally, with the help of the Ethical Review Committee (ERC) of the project, a proper disclosure to the Authority for the Protection of Personal Data was submitted.

2.4. Event phase

The exercise started at 9:37 a.m. on June 7th with the MAYDAY alert communicated from the ship Vincenzo Florio to the Coast Guard (CG) control room. The CG took the command and coordination of the rescue operation at sea, ordered the Search & Rescue patrol boat to approach the vessel on fire and contacted the Fire Brigade (VVF) in the harbor to order their action. The CG called the Prefecture to give the alert and the Prefecture activated the Crisis Unit in accordance with the usual procedure. After 15–20 min, the Crisis Unit was operational at the Prefecture. At 9:47 a.m., the captain of the vessel communicated the presence of a seriously injured seaman, who was rescued by the CG helicopter. At 9:55 a.m., the captain communicated the presence of several seriously injuried seamen. At 10:00 a.m. the master communicated that the crew was not able to extinguish the fire and that he had to initiated the abandonment of the ship. At 10:10 a.m. in the Puntone dock in the harbor, the PMA (PMA A) was ready to receive the victims from the ship and the sea. At 10:15 a.m. for the presence of a black cloud towards the urban area, a fast air analysis was required from the VVF, who in few minutes communicated the presence of a toxic cloud, also providing information relevant to the involved area (red area). The Prefect ordered to the Police forces the isolation of the red area. All the buildings involved in the red area were warned about the risk of the toxic cloud from the Municipal civil protection, which asked them to activate the procedure against that type of risk. The first victims from the sea arrived at Puntone dock at 10:30 a.m. and then passed through the decontamination unit, received the first triage at PMA A. The first victim was transported to the Emergency department of Buccheri La Ferla Hospital. At 10:30 a.m. PMA B (located in Piazza Marina) and PMA C (located in Foro Umberto I) were activated to receive the victims from the population. At 10:30 a.m., the Buccheri La Ferla hospital activated its PEIMAF. The outputs recorded in the IMPRESS Platform during the demo showed the activities of INCIMOB operators in the field and the messages sent to INCIMAG (Tables 4 and 5). At 10.45 a.m. the first victim among the population on the field was taken to PMA B. At 11.30 a.m. the last victims from the ship arrived at Puntone dock. At 12.15 p.m. the VVF communicated that the work of extinguishing the fire on board allowed the ship to be towed to a safe area. At 12.25 p.m. the last victims were triaged at the Emergency department of Buccheri La Ferla Hospital. At 01.00 p.m. the air quality was declared within limits. At 01.05 p.m. the Prefect declared the “END EXERCISE”.

3. Results

3.1. Post-event phase: questionnaire analysis

Overall 60 questionnaires were collected: 21 were INCIMAG-OBSERVER Evaluation Questionnaires; 10 were INCIMAG-USER Evaluation Questionnaires; 12 were INCIMOB-OBSERVER Evaluation Questionnaires; 17 were INCIMOB-USER Evaluation Questionnaires.

The 21 INCIMAG observers were from four different agencies: 52.4% (11/21) was from Public Institutions, the 23.8% (5/21) was from Medical (Hospital) Agencies, the 19% (5/21) from Medical Emergency Services and the 4.8 (only one case) was from the Rescue Agency. The percentage of males was higher (71.4%, 15/21) than the percentage of females (28.6%, 6/21). Most of Observers (42.9%, 9/21) were in the higher age class (> 50 yrs), while the remaining were equally (28.6%, 6/21) distributed in the first two classes (31–40 and 41–50).

The 10 INCIMAG users were distributed in the four Agencies according to the following: 30% (3/10) was from Public Institutions, 30% was from Medical Emergency Service, 20% (2/10) from the Rescue Agency and the 10% (1/10) was from both the Medical (Hospital) Agency and Police. Males accounted for 90% (9/10). Also in this group most of the users (4 people) were in the higher age class (> 50), 3 were in the class 31–40, 2 in the class 41–50 and only one person was aged less than 30.

The 12 INCIMOB Observers (75% males, 9/12) were from the following agencies: 42% (5/12) from Public Institutions, 25% (3/12) from Medical (Hospital) Agency, 17% (2/12) from Medical Emergency Service and the 8.3% (one person) both from the Rescue and Military Agencies. They were mainly aged between 31 and 40 (42%, 5/12). 33% (4/12) were in the class > 50 years, 17% (2/12) were aged between 41 and 50 and 1 person (8.3%) was younger than 30 years.

The 17 INCIMOB users were distributed among the four Agencies according to the following: 53% (9/17) were from Rescue Agency, 17.6% (3/17) were from Public Institutions, 11.8% (2/17) both from Medical Emergency Service and Fire Agency, 6% (one person) from the Medical (Hospital) Agency. 88% (15/17) were female. Most of the users (47%, 8/17) were aged 31–40 years, 30% (5/17) were aged greater than 50 years and the remaining 23% (4/17) were equally distributed in the two classes 41–50 and < 30 years.

3.1.1. Improvement of function excution questions

The first set of questions, which differed between the INCIMAG and INCIMOB questionnaires, were aimed at evaluating to which extent IMPRESS improved the execution of some important functions to be executed during the demo, in comparison with similar exercises performed without the use of IMPRESS. For each identified function, observers and users were asked to give a score from Strongly Disagree to Strongly Agree to express their opinion. The qualitative answers were recoded as numerical scores (from 5 = Strongly Agree to 1 = Strongly Disagree, leaving out the "Not Applicable" score) and differences between Observers and Users were evaluated by means of a Mann-Withney *U* test. Tables 2–5 report the questionnaires' items with the relative percentages of answers in the five scores and the corresponding median score.

For the INCIMAG questionnaire no significant difference emerged between observers and users on any of the administered questions related to the improvement of some important functionalities with the use of IMPRESS system. On the other hand only for one INCIMOB item the difference in responses between Observers and Users resulted to be significant: *Impress improves on scene victim prioritization* (P = 0.0336), while for the item *Impress improves on scene coordination & control* a borderline significance was obtained (P = 0.0599) with the Observers median being "Neutral" and the User median being "Agree" for both items.

3.1.2. System usability questions

Another part of the IMPRESS questionnaires concerned the usability of the system. The usability questions were addressed only to the INCIMOB and INCIMAG Users. A Mann-Withney *U* test was used to test if the usability of the two tools was perceived to be different. Tables 6 and 7 report the results related to the two administered questionnaires.

Only for the item *The system worked as I expected* a borderline significant difference in the responses was recorded (P = 0.0591) between INCIMAG and INCIMOB users (the median values were "Agree" and "Disagree" respectively).

3.1.3. Suggested functionalities

Some questions were addressed to understand which functionalities should be added to the IMPRESS Solution. Each answer was categorized according to its content and was assigned to one of the following classes: "Alerting", "Functionalities", "Communication", "Data entry", "Training", "Searching capabilities", "Visualization/Information display and sharing". "Alerting" referred mainly to the introduction of ring tones to attract attention on new events occurring during the crises. The "Functionalities" category included suggestions about system simplification and consolidation of already existing functions, such as optimization of human and material resource allocation, visualization on maps of resources and information streams improvement of victim related information transmission, victim location and delivered treatments. "Communication" suggestions on one hand referred to the possibility to share information among all the agencies involved in the crisis, on the other hand, were focused on the need of having a system based on a pyramidal information sharing. "Data entry" suggestions were aimed to have a speedier data entry system, including pre-compiled text and speech recognition. As for "Training", we received requests for a thorough training of the system users. "Searching capability" functions were required for location visualization on the map, for information of recorded victims in the system and for a help when users are in need with the system utilization. "Visualization/Information display and sharing" was the category where most of the suggestions were collected. Suggestions referred either to a summary of the crisis status (number of victims with their triage code, tracking status of the victims, allocated/free resources, hospital capacities) or to validation data and decision, providing users with alert or suggestions on patient management, and with a forecast of time of arrival to the hospital based for example on traffic information. The need for a multilanguage translation and for a facilitate multidisciplinary information exchange also emerged. The number of answers obtained in each category by Users and Observers were summarized into Table 8.

Table 1
KPIs (Key Performance Indicators).

KPIs	Definition
t0	The time the scene was declared safe by the fire brigade - Time of initial notification call
T1stTg	**Time to first triage**: Time until the first ill/injured victim has been triaged in the field
T1stMv	**Time to first move (CA/AMP)***: Time until each victim was transported in CA or AMP
T1stTx	**Time to first treatment**: Time until first treatment was performed
LosSc	**Scene length of stay (LOS)**: Time until victim is evacuated from scene
TtED	**Time to ED arrival**: Time until victim arrives to emergency department ED
TtMedMgr	**Time to notification** of the first appropriate staff person who assumes medical management coordination role
T1stAmb	**Time to arrival** of the first EMS ambulance on scene
TLstEvac	**Time to transportation/evacuation** of the last ill/injured survivor from the scene
T1stEDtr	**Time to first ED triage**: Time until first triage assessment in emergency department
TLstEDtrAss	**Time to last ED triage**: Time until last triage assessment in emergency department: Time until initial medical ass
TInTrans	**Average Time in transit**: average time spent by victims on ambulances and helicopters
NoVictEvac	**Number of victims evacuated from scene**
NoVictTr	**Number of victims that receive first triage**
NoVictED	**Number of victims transported to emergency department**

3.1.4. Questionnaire analysis - the IMPRESS effectiveness

Responders were asked if the problems presented in the exercise were partially or totally solved by the IMPRESS capability. For both INCIMAG and INCIMOB (see Fig. 3) the percentages of positive responses were higher than those of negative responses. The Chi-Squared test was used to assess a possible association between Observers and Users in perceiving the usefulness of the IMPRESS Solution in solving the encountered problems. The test resulted to be not significant for both comparisons. Users and Observers were also asked to identify motivations according to which IMPRESS proved not useful in solving the problems presented during the exercise. The most indicative answers can be collected into four macro-areas:

- Communication: lack of communication by radio through the IMPRESS system; text messaging was too slow and then not efficient. Some users complained the lack of sharing information among the different INCIMAG stations.
- Feasibility: a direct victim assessment sometimes resulted to be faster than querying the IMPRESS system for an automatic triage of the patient.
- Functionality: the system sometimes crashed and hardware problems prevented the use of the devices. Use of INCIMOB sometimes presented difficulties due to the operations to be performed on the field.
- Training: lack of an adequate training on the use of the IMPRESS system.

3.1.5. Open questions section

Questionnaires reported also open questions (see the Appendices). Open questions resulted to be redundant with respect to the information already collected, but helped to better understand shortcomings and areas of improvement of the IMPRESS system. Encountered problems were due above all to the limited familiarity with the system, which caused slowdowns in system use or requests for functionalities already installed but that the users were unable to utilize. From the open questions emerged again what already reported in two above paragraphs.

Table 2
INCIMAG observers.

INCIMAG	Observers						
IMPRESS improves	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
alert notification	19	42.9	19	4.8	9.5	4.8	Agree
data collection	28.6	47,6	14,3	0	4,8	14,3	Agree
disaster management plan activation	4.8	42.9	33.3	4.8	9.5	4.8	Neutral
incident area assessment	23.8	33.3	19	4.8	4.8	14.3	Agree
needs assessment	19	38.1	33.3	0	4.8	4.8	Agree
resources mobilization	28.6	38.1	19	9.5	0	4.8	Agree
resources allocation	19	38.1	22.8	14.3	0	4.8	Agree
evacuation priorities	4.8	42.9	19	14.3	4.8	14.3	Agree
victim distribution	23.8	38.1	19	9.5	4.8	4.8	Agree
decision making	9.5	23.8	47.6	9.5	4.8	4.8	Agree
coordination & control	9.5	61.9	14.3	4.8	4.8	4.8	Agree
communication with other organizations during response	23.8	38.1	28.6	4.8	4.8	0	**Agree**
collaboration with other organizations during response	19	23.8	47.6	4.8	4.8	0	Neutral
continuity of operations	19	28.6	33.3	9.5	4.8	4.8	Neutral
situation awareness	28.6	23.8	23.8	14.3	9.5	0	Agree
IMPRESS is useful as a training tool within your own service	23.8	19	33.3	14.3	9.5	0	Neutral
as a training tool with other involved organizations (interoperability)	14.3	33.3	33.3	9.5	9.5	0	Neutral
emergency public information	14.3	19	42.9	4.8	4.8	14.3	Neutral
time management	19	28.6	42.9	9.5	0	0	Neutral
time saving	23.8	33.3	28.6	4.8	9.5	0	Agree

Table 3
INCIMAG users.

INCIMAG	Users						
IMPRESS improves	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
alert notification	10	50	10	0	20	10	Agree
data collection	30	20	20	10	10	20	Agree
disaster management plan activation	10	60	10	0	10	10	Agree
incident area assessment	30	50	0	0	10	10	Agree
needs assessment	20	30	30	0	10	10	Agree
resources mobilization	20	10	50	0	10	10	Neutral
resources allocation	20	20	40	10	0	10	Neutral
evacuation priorities	10	40	10	20	10	10	Agree
victim distribution	30	30	20	10	0	10	Agree
decision making	10	30	20	10	20	10	Neutral
coordination & control	20	40	0	30	0	10	Agree
communication with other organizations during response	30	30	10	10	10	10	Agree
collaboration with other organizations during response	10	50	10	20	0	10	Agree
continuity of operations	20	30	30	10	0	10	Agree
situation awareness	10	50	10	10	10	10	Agree
IMPRESS is useful as a training tool within your own service	10	50	20	10	0	10	Agree
as a training tool with other involved organizations (interoperability)	10	60	10	0	10	10	Agree
emergency public information	10	40	20	10	0	20	Agree
time management	10	40	20	20	0	10	Agree
time saving	10	40	20	20	0	10	Agree

Table 4
INCIMOB observers.

INCIMOB	Observers						
IMPRESS improves	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
on scene decision making	8.3	33.3	50	0	8.3	0	Neutral
on scene data collection	8.3	66.7	8.3	0	0	16.7	Agree
on scene victim assessment	8.3	33.3	25	25	0	8.3	Neutral
on scene victim prioritization	0	33.3	33.3	25	8.3	0	Neutral
on scene victim treatment	25	25	41.7	0	8.3	0	Neutral
on scene victim tracking	8.3	25	58.3	8.3	0	0	Neutral
evacuation priorities	16.7	25	33.3	16.7	0	8.3	Neutral
victim distribution	0	25	58.3	16.7	0	0	Neutral
on scene coordination & control	0	41.7	33.3	8.3	8.3	8.3	Neutral
on scene communications	8.3	25	50	0	16.7	0	Neutral
on scene collaboration	8.3	25	50	8.3	8.3	0	Neutral
IMPRESS is useful as a decision support tool for on scene management	8.3	33.3	33.3	16.7	8.3	0	Neutral
as a training tool within your own service	8.3	25	33.3	8.3	16.7	8.3	Neutral
as a training tool with other involved organizations (interoperability)	16.7	25	25	16.7	16.7	0	Neutral
time management	25	8.3	33.3	16.7	8.3	8.3	Neutral
time saving	16.7	16.7	25	25	8.3	8.3	Neutral

Table 5
INCIMOB users.

INCIMOB	Users						
IMPRESS improves	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
on scene decision making	0	47.1	29.4	11.8	0	11.8	Agree
on scene data collection	5.9	47.1	29.4	11.8	0	5.9	Agree
on scene victim assessment	0	41.2	23.5	11.8	5.9	17.6	Neutral
on scene victim prioritization	0	52.9	23.5	0	0	23.5	Agree
on scene victim treatment	5.9	52.9	5.9	11.8	5.9	17.6	Agree
on scene victim tracking	5.9	58.5	11.8	0	5.9	17.6	Agree
evacuation proprities	17.6	47.1	5.9	5.9	5.9	17.6	Agree
victim distribution	5.9	25.3	5.9	17.6	11.8	23.5	Agree
on scene coordination & control	23.5	41.2	17.6	5.9	0	11.8	Agree
on scene communications	0	41.2	11.8	17.6	11.8	17.6	Neutral
on scene collaboration	0	47.1	29.4	11.8	0	11.8	Agree
IMPRESS is useful as a decision support tool for on scene management	0	29.4	41.2	5.9	5.9	17.6	Neutral
as a training tool within your own service	0	29.4	29.4	11.8	11.8	17.6	Neutral
as a training tool with other involved organizations (interoperability)	5,9	47.1	11.8	11.8	0	23.5	Agree
time management	0	41.2	23.5	17.6	0	17.6	Neutral
time saving	0	23.5	11.8	35.3	11.8	17.6	Disagree

Table 6
System usability INCIMAG.

System Usability	INCIMAG						
ONLY System User	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
the system was easy to use	30	30	10	10	20	0	Agree
the font size was appropriate	10	40	0	30	10	10	Agree
the font style was appropriate	10	30	20	20	10	10	Neutral
the labels that described the functions made sense to me	0	60	10	20	0	10	Agree
I was able to find functionality where I expected it	20	50	0	20	0	10	Agree
the amount of information was appropriate (i.e., not overwhelming or too sparse)	30	20	10	30	0	10	Agree
I could become productive quickly using the main menu in the system	20	30	20	10	0	20	Agree
the system was easy to learn to use	10	20	30	10	10	20	Neutral
the functions of the system met my needs	20	30	20	20	0	10	Agree
the available functionality of the system was complete	10	20	20	30	10	10	Neutral
the system was enjoyable to use	10	20	30	0	10	30	Neutral
I was able to navigate easily while using the system	0	30	30	10	10	20	Neutral
the organization of functions made sense to me	10	30	40	0	10	10	Neutral
I was able to develop a care plan using the system	0	30	40	0	10	20	Neutral
interventions presented in the system were relevant	0	40	30	10	10	10	Neutral
I have enough education to use the system	10	30	40	0	10	10	Neutral
overall, I am satisfied with the system's capabilities	0	70	10	10	0	10	Agree
the system worked as I expected	0	50	30	10	0	10	Agree
the system was flexible	0	20	50	10	0	20	Neutral
balance between training time and the benefits of using the system	10	10	40	20	0	20	Neutral

3.1.6. Palermo DEMO outputs

The IMPRESS system automatically recorded data related to the operations conducted with the use of IMPRESS, as for example the number of triaged victims and other key indicators useful to evaluate the DEMO results. The most relevant results showed that only 32 patients out of 40 mock victims (20 at ship or sea and 20 among the population) were triaged and then tracked with the use of IMPRESS, 19 were victims from the ship and the sea and 13 victims from the Kalsa district; 21 patients were transported to the hospital, 15 were yellow codes and 6 were red codes; the patients tracked from ambulance INCIMOBs were 8; five ambulances did not use INCIMOB intensively. Table 9 reports the main outputs, in terms of collected KPIs, from the DEMO, while Table 10 reports the distribution of the triage codes at the beginning of the DEMO.

4. Discussion

Decision Support Systems (DSSs) for crisis management are becoming useful instruments for activity coordination, resource allocation and efficient information sharing in crisis situations, but their validation must be conducted and an analysis of their effective utility and efficiency must be performed. The management of mass casualties and the application of effective and accurate triage are multi-disciplinary activities based on team coordination [13]. Table-top exercises are generally implemented to this end, but evaluation of Decision Support System in Emergency medicine can greatly benefit from large live exercises, to highlight both critical issues in the response system and weaknesses in the decision support tools to be evaluated.

Even if well planned, a live exercise conducted on full-scale and involving therefore different agencies and disciplines must contemplate the occurrence of unexpected events, just like in a real situation. This character of randomness makes the evaluation of the DSS more real and effective and, if on one hand more shortcomings of the tool may be evidenced during the execution of the demo, on the other hand a larger basis for further improvements is obtained. The very problem however resides in the impossibility to perform a formally correct system validation: this should ideally be conducted through statistical comparison of Key Performance Indicators collected during parallel, randomized, prospective, blind testing procedures, one including the DSS, the other excluding it. This is however impossible to do, also because of the impossibility of repeating the same live region-wide exercise with and without the use of the tested platform (it would be impossible to reproduce the demo under identical conditions, same operators, same

Table 7
System usability INCIMOB.

System Usability	INCIMOB						
ONLY System User	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Not Applicable	Median
the system was easy to use	18	53	18	12	0	0	Agree
the font size was appropriate	17.6	58.8	5.9	17.6	0	0	Agree
the font style was appropriate	17.6	64.7	5.9	11.8	0	0	Agree
the labels that described the functions made sense to me	12	41	35	12	0	0	Agree
I was able to find functionality where I expected it	0	41.2	29.4	23.5	5.9	0	Neutral
the amount of information was appropriate (i.e., not overwhelming or too sparse)	5.9	23.5	29.4	23.5	17.6	0	Neutral
I could become productive quickly using the main menu in the system	0	35.3	41.2	5.9	11.8	5.9	Neutral
the system was easy to learn to use	11.8	58.8	17.6	5.9	0	5.9	Agree
the functions of the system met my needs	5.9	23.5	35.3	23.5	5.9	5.9	Neutral
the available functionality of the system was complete	11.8	35.3	23.5	17.6	5.9	5.9	Neutral
the system was enjoyable to use	17.6	41.2	23.5	11.8	0	5.9	Agree
I was able to navigate easily while using the system	11.8	52.9	17.6	5.9	11.8	0	Agree
the organization of functions made sense to me	18	35	35	12	0	0	Agree
I was able to develop a care plan using the system	0	23.5	35.3	23.5	5.9	11.8	Neutral
interventions presented in the system were relevant	11.8	23.5	41.2	11.8	5.9	5.9	Neutral
I have enough education to use the system	29.4	41.2	11.8	5.9	5.9	5.9	Agree
overall, I am satisfied with the system's capabilities	11.8	17.6	35.3	23.5	5.9	5.9	Neutral
the system worked as I expected	5.9	17.6	11.8	35.3	17.6	11.8	Disagree
the system was flexible	5.9	23.5	17.6	23.5	17.6	11.8	Neutral
balance between training time and the benefits of using the system	11.8	5.9	35.3	23.5	23.5	0	Neutral

Table 8
Suggested functionalities. MAG-O: INCIMAG Observers, MAG-U: INCIMAG Users, MOB-O: INCIMOB Observers, MOB-U: INCIMOB Users.

Suggested functionalities	MAG-O	MAG-U	MOB-O	MOB-U	Total
Alerting			2	1	3
Functionalities	5	2	3	10	20
Communication	6	1	1		7
Colouring	2	1			3
Data entry	1				1
Training	1			1	2
Searching capabilities	3		1	1	5
Visualization/ Information display and sharing	13	1	1		15

modus operandi, same traffic conditions …).

4.1. Methodologic implications

For these reasons we set up and followed a different approach. We are certain that a DSS validation must: i) be performed on the field; ii) fulfil the needs of users and stakeholders; iii) provide an iterative introduction of improvements and modifications in subsequent steps, just following the suggestions of the final users and stakeholders. Thus, we followed the approach of evaluating IMPRESS, and more in specific each its subcomponent (communication platform, data warehousing, DSS subcomponents) during both a real-life exercise and during a table top demo. This allowed us to demonstrate the IMPRESS interoperability and functionalities to the end-users in their real environment and to give end-users a chance to assess the added value, as well as shortcomings, of IMPRESS for further improvements. The test and evaluation process which took place during the Palermo Demo aimed at addressing the

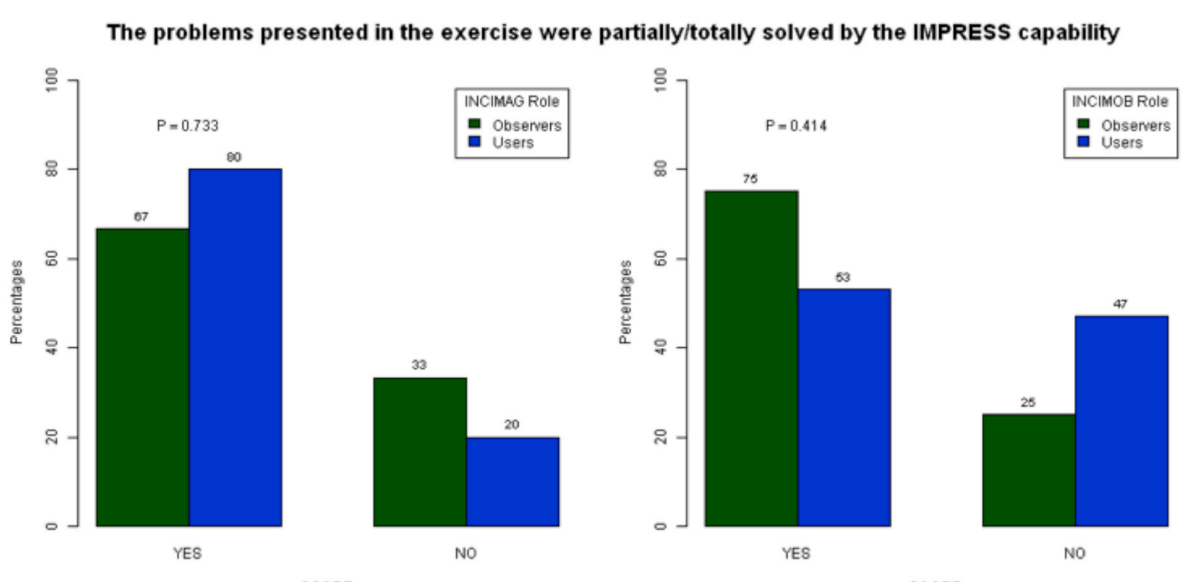

Fig. 3. The problems presented in the exercise were partially or totally solved by the impress capability.

Table 9
Palermo DEMO Outputs in terms of KPIs.

IMPRESS - Palermo Exercise Key Performance Indicators	Value
Number of patient tracking messages sent from INCIMOB to INCIMAG	105
Number of patients tracked by INCIMOB and sent to INCIMAG	32
Number of patients from the ship	12
Number of patients from sea	7
Number of patients from Kalsa district	13
INCIMOB First Triage/Track - First Encounter	Value
VVF on board	9
PMA A (Puntone dock)	7
CRI on board	3
PMA B (piazza Marina)	9
PMA C (Foro Umberto I)	0
Hospital	3
AMB 9	1
Triaged/Tracked in Hospital	Value
Triaged in hospital	19
Lost tag in hospital	2
Patients triaged/tracked by INCIMOB	Value
VVF	9
CRI	3
PMA A	9
PMA B	10
PMA C	0
AMB 1	1
AMB 3	1
AMB 8	2
AMB 9	3
AMB 10	1
Hospital Team A	11
Hospital Team B	9

Table 10
Patients color codes distribution.

INCIMOB	Initial Triage color			Total
	Green	Yellow	Red	
CRI on board	2	1	0	3
VVF on board	4	3	2	9
PMA A	4	2	1	7
PMA B	0	7	2	9
Amb 9	0	0	1	1
Hospital Team A	0	2	0	2
Hospital Team B	1	0	0	1
Total	11	15	6	32

following two different aspects: find inconsistencies and unintended bugs (more related to technical aspects); get feedback about the usefulness of the platform (more related to the interaction user-platform). While the first aspect is important for bug reporting, the second aspect has relevance for improvements in IMPRESS usability and usefulness. The objective of the live Palermo Demo was therefore the assessment of the relevance for the end users of the IMPRESS components and their features, of the system as a whole, and of the functionalities it aims to deliver. For all these reasons, we presented the ad-hoc validation assessment as reported in the manuscript.

Thus an ad-hoc validation assessment was planned. This work aims at describing the methodology, as it was applied in the Palermo DEMO but extendable to similar live exercises, as well as the obtained results. The validation process was addressed to:

- assess the relevance i) of the IMPRESS components and their features for the end users; ii) of the system as a whole; and iii) of the functionalities it aims to deliver;
- find inconsistencies and unintended bugs;
- get feedback about the usefulness of the platform.

The IMPRESS validation during the Palermo Demo was therefore based on the following elements:

1. A group of high-level external expert evaluators were identified for the setup of the methodology of the validation process: target criteria were decided and questionnaires to be administered post-exercise were drawn up.
2. A number of professionals, representing the several specialties involved in the management of the crisis and in the preparation of the software tools for it, were called to participate in the demo as observers, dislocated at different points (on the field and at decisional stations) with the objective of pointing out the main functionality-related issues.
3. A User Oriented questionnaire was administered to the users of the System immediately after the end of the demo.
4. A set of quantitative, automatically recorded benchmarks were defined and collected during the demo.

4.2. Lesson learnt from the Palermo Demo

The Users were persons working at the tactical/operational level as well as strategic level in crisis management and belonged to those organizations involved in the response to similar crises, should they happen. The Observers represented the many diverse professional expertise involved in the planning, realization, implementation and use of a crisis management support system. The Evaluators were senior people working in different organizations, experts in crisis management from different countries. By structuring the evaluation panels in this way, high-quality feedback from complementary perspectives were obtained. This procedure allowed the IMPRESS platform to be adapted and tailored, in the remainder of the project, according to the needs expressed and to the criticisms received. The Palermo Demo was indeed the first opportunity to test the IMPRESS solutions on the field. Problems were encountered in testing the IMPRESS prototype: some criticisms were raised, due mainly to the lack of training and to the inefficient communication facilities of IMPRESS, such as those related to the decision support functions. Some suggested improvements were subsequently introduced in the final version of IMPRESS and tested in a table top exercise. However, even in the face of a series of problems experienced during the DEMO, Users expressed high satisfaction with the use of the ICT systems and were very useful in providing indications for further improvement. Satisfaction was expressed with both INCIMAG and INCIMOB. The items of INCIMAG questionnaire were aimed at evaluating to which extent IMPRESS improved the execution of some important functions to be executed during the demo in comparison to similar exercises performed without the use of IMPRESS. Median values of the scores given by the Users and Observers were “Agree” for almost all the items. The score “Neutral” was the median for the remainder items. A comment recorded into one of the INCIMAG questionnaire read as it follows: “INCIMAG is an excellent tool which the user should be familiar with. Table-top exercises should be planned”. This comment highlights that such a tool could be very useful for people involved in the management of crisis situation and that, in spite of the encountered difficulties in the use of IMPRESS, due both to system shortcomings and insufficient training, Users and Observers appreciated the potential benefit and advantage from the use of the system. The use of INCIMOB resulted perhaps more problematic from an operational point of view, due to the fact that it must be used concomitantly with the normal activities to be carried out in the field, which can represent a burden for the normal operations. However, if Observers were essentially neutral, Users expressed a positive judgement with a median score of “Agree” for most of the items. Conversely, answer “Disagree” was the median score relevant to the question investigating if the use of INCIMOB improved time saving. Most of the comments, indeed, highlighted the need for voice communication, not only textual, for speeding up communications. Apart from communication problems and the need for more training sessions, suggestions were received on improvement of information visualization. In this regard it should be noted that some of the requests were made due to poor knowledge of the system functionalities.

Requests for summary of the crisis status (number of triaged patients, free and allocated resources, etc.) derived from the difficulty encountered in navigating and exploring INCIMAG, since the requested functionalities had already been implemented.

5. Conclusions

Results obtained from the Palermo Demo were very useful for the subsequent improvement of the IMPRESS system. Most of the shortcomings which emerged from the live exercise would be probably obscured in a table-top demo, where input data and outcome are in general predictable, the input is programmed and the emerged output can be real-time measured. In a table-top exercise, indeed, the effect of unpredictable variables cannot be considered and therefore a real evaluation of the usefulness, usability and efficacy of a DSS system, when used in the real world, cannot be derived. Therefore, while table-top exercises are easier to conduct, they are not for the individuation of problems and shortcomings which might arise only in the framework of a real operational setting. The organization of a live-demo, even if a cumbersome task, difficult to perform on a large-scale, becomes mandatory for effectively measure the performance and usefulness of a decision support tool. Evaluation of a DSS in decision making, should be done however in a way that the comparison between consequences of decisions taken without the aid of the DSS with those emerged with the aid of the DSS, is feasible. Rigorous system validation in a live exercise is however impossible to perform due to the non reproducibility of the same conditions. We suggest, therefore, to set-up an ad-hoc methodology, such as that implemented for the IMPRESS validation, able to represent a good compromise in situation where events are non-repeatable. The IMPRESS experience suggests that live exercises, coupled with an ad-hoc methodological approach, could be very useful for individuating directions to be addressed with the aim of delivering a product of substantial usefulness for the relevant stakeholders.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The IMPRESS Palermo Demo was part of the FP7 EU IMPRESS Project, FP7-SEC-2013-1, Project No. 608078 and was developed under the coordination of the Regional Department of Civil Protection of Sicily. We wish to thank Calogero Foti, Paolo Blandino, and Gianluigi Gioia for their effort in making the Demo possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2020.101695.

References

[1] A. Djalali, L. Ingrassia, F.D. Corte, M. Foletti, A.R. Gallardo, L. Ragazzoni, K. Kaptan, O. Lupescu, C. Arculeo, G. Von Arnim, T. Friedl, M. Ashkenazi, D. Heselmann, B. Hreckovski, A. Khorrram-Manesh, R. Komadina, K. Lechner, C. Patru, F.M. Burkle Jr., P. Fisher, Identifying deficiencies in national and foreign medical team responses through expert opinion surveys: implications for education and training, Prehospital Disaster Med. 29 (2014) 364–368.
[2] American College of Surgeons, Advanced Trauma Life Support Program for Doctors: ATLS, 6 edition, American College of Surgeons, Chicago, IL, 1997.
[3] C.S. Kim, T. Guimaraes, Selecting dss evaluation methods, J. Inf. Technol. Manag. 3 (1992) 29–37.
[4] F.N. de Silva, R.W. Eglese, M. Pidd, Evacuation planning and spatial decision making: designing effective spatial decision support systems through integration of technologies, in: Decision-Making Support Systems: Achievements and Challenges for the New Decade, vol. 358, IGI Global, 2003, p. 373.
[5] E. Savoia, F. Agboola, P.D. Biddinger, A conceptual framework to measure systemsâ™s performance, Int. J. Environ. Res. Publ. Health 11 (2014) 9712–9722.
[6] E.S. Lang, D.W. Spaite, Z.J. Oliver, et al., A national model for developing, implementing, and evaluating evidence-based guidelines for prehospital care, Acad. Emerg. Med. 19 (2012) 201–209.
[7] G.B. Green, S. Modi, K. Lunney, et al., Generic evaluation methods for disaster drills in developing countries, Ann. Emerg. Med. 41 (2003) 689–699.
[8] H. Admi, Y. Eilon, G. Hyams, L. Utitz, Management of mass casualty events: the israeli experience, J. Nurs. Scholarsh. 43 (2011) 211–219.
[9] J.B. Lamy, A. Ellini, J. Nobecourt, et al., Testing Methods for Decision Support Systems. Decision Support Systems. Chiang S.Jao, INTECH Croatia, 2010.
[10] J.D. Bayram, L.M. Sauer, C. Catlett, S. Levin, G. Cole, T.D. Kirsch, M. Toerper, G. Kelen, An expert consensus panel, PLoS Curr. 5 (2013).
[11] J.D. Bayram, S. Zuabi, I. Subbarao, Disaster metrics: quantitative benchmarking of hospital surge capacity in trauma-related multiple casualty events, Disaster Med. Public Health Prep. 5 (2011) 117–124.
[12] J.M. Franc, P.L. Ingrassia, M. Verde, F. Colombo Dand Della Corte, A simple graphical method for quantification of disaster management surge capacity using computer simulation and process-control tools, Prehospital Disaster Med. 19 (2014) 1–7.
[13] J.M. Ryan, D. Doll, in: G.C. Velmahos, E. Degiannis, D. Doll (Eds.), Penetrating Trauma, Springer, 2012, pp. 151–159.
[14] J.R. Paul, V. Hlupic, The casm environment revisited, in: 26th Conference on Winter Simulation, WSC '94, 1994, pp. 641–648.
[15] M. Mora, G. Forgionne, J.N.D. Gupta, Decision-Making Support Systems: Achievements and Challenges for the New Decade, IGI global Publisher of Timely Knowledge, 2002.
[16] N. Dobrinkova, A. Kostaridis, A. Olunczek, M. Heckel, D. Vergeti, S. Tsekeridou, G. Seynaeve, A. De Gaetano, T. Finnie, N. Efstathiou, C. Psaroudakis, Disaster reduction potential of impress platform tools 225–239 (2017) 11.
[17] N. Dobrinkova, T. Finnie, J. Thompson, I. Hall, C. Dimopoulos, G. Boustras, Y. Danidou, N. Efstathiou, C. Psaroudakis, N. Koutras, G. Eftichidis, I. Gkotsis, M. Heckel, A. Olunczek, R. Hedel, A. Kostaridis, M. Moutzouris, S. Panunzi, G. Seynaeve, D. Vergeti, Optimisation of Preparedness and Response of Health Services in Major Crises Using the IMPRESS Platform: Results of the Workshop on Computational Optimization WCO, 2017, pp. 15–34, 01 2019.
[18] N. Li, D.A. Stanford, A.B. Sharif, R.J. Caron, A. Pardhan, Optimising key performance indicator adherence with application to emergency department congestion, Eur. J. Oper. Res. 272 (2019) 313–323.
[19] P. Kulling, M. Birnbaum, V. Murray, G. Rockenschaub, Guidelines for reports on health crises and critical health events, Prehosp Disaster Med. 25 (2010), 377â–383.
[20] P.L. Ingrassia, D. Colombo, F.L. Barra, et al., Impact of training in medical disaster management: a pilot study using a new tool for live simulation, Emergencias 25 (2013) 459–466.
[21] R. Safdari, M. Ghazisaeedi, M. Mirzaee, J. Farzi, A. Goodini, Development of balanced key performance indicators for emergency departments strategic dashboards following analytic hierarchical process, Health Care Manag. 33 (2014) 328–334.
[22] Y. Bar-El, S. Tzafir, I. Tzipori, L. Utitz, M. Halbertal, R. Beyar, S. Reisner, Decision-support information system to manage mass casualty incidents at a level 1 trauma center, Disaster Med. Public Health Prep. 7 (2013) 549–554.
[23] B.P. Zeigler, Theory of Modelling and Simulation, John Wiley & Sons, New York, 1976.