Multi-Criteria Evaluation of Mobile Triage Decision Systems

Pari DELIR HAGHIGHIa,1, Frada BURSTEINa, Leonid CHURILOVb and Akrati AMIT PATELA

aMonash University, Australia
bThe University of Melbourne, Australia

Abstract. Mobile Clinical Decision Support Systems (CDSS) aim to assist medical practitioners with decision making and provide an easy access to information on the move using mobile devices such as a mobile phone. Despite many benefits of mobile CDSS, these systems need to undergo a comprehensive evaluation to determine the correctness of the actions that decision maker take and validate the system’s knowledge and recommendation. In this paper, we discuss the impact of mobile CDSS on decision making process and outcomes, and describe the multi-criteria evaluation of a mobile field triage decision support system in terms of effectiveness, efficiency, satisfaction and use. In this research, data was collected from a group of qualified paramedics using both quantitative and qualitative methods. The results and findings shed new light on the role of mobile decision support systems in improving healthcare delivery and provided better understanding of the barriers to the adoption of these systems and the impact of different stakeholders’ views in evaluating the success of the DSS.

Keywords. DSS evaluation, mobile DSS, clinical DSS, medical triage

Introduction

Clinical decision support systems (CDSS) are information systems designed to assist health practitioners with clinical decision making [1]. They provide “knowledge and person-specific information, intelligently filtered and presented at appropriate times, to enhance health and health care” [2:1]. They support clinical diagnosis and treatment plan processes and promote use of best practices and condition-specific guidelines [3]. While CDSS are beneficial in improving health practitioners’ performance and patient outcomes, there is a limited access to such systems at the point of care [4]. Mobile CDSS aim to address this challenge. They provide portability, timely decision making and an easy access to the information at the right time and in the right place using handheld devices such as a PDA or mobile phone [5]. Mobile CDSS can be used for patient tracking, electronic prescribing, clinical and drug references, patient-specific care reminders, and system-generated guidance for clinicians [4, 5]. An example of system-generated guidance systems is mobile triage DSS that aims to assist with triage decision making during emergencies and mass casualty situations [6, 7].

1 Corresponding Author: Pari Delir Haghghi, Faculty of Information Technology, Monash University, Caulfield East, 3145, Victoria, Australia, E-mail: Pari.DelirHaghighi@monash.edu.
Like any other healthcare innovation, mobile CDSS need to undergo a comprehensive evaluation to validate the system's knowledge and determine the correctness of the system recommendations [1]. CDSS can directly affect the patient outcome and it is highly important to conduct careful validation of these tools before their adoption in clinical practice [2]. The evaluation of decision support systems traditionally considers the organizational costs, quality of information and the impact on users’ decisions and actions [8, 9]. Yet, in healthcare, measuring intangible costs and the health impacts can become a challenging task. One of the common approaches for validating CDSS is evaluating accuracy of recommendations [7, 10]. While accuracy is critical to the success and acceptance of CDSS, the evaluation of mobile CDSS should measure other aspects that could be relevant to the decision task and important to the success of CDSS.

Previous research has shown that often more than one measure has to be taken into consideration when evaluating DSS. Moreover, the view of the system can differ depending on the role of the evaluator. Maynard et al. [11] propose a flexible and rich multi-faceted evaluation approach for decision support systems. The framework considers four main domains including effectiveness, efficiency, satisfaction and use, and each domain supports a number of criteria. The authors suggest the views of the following DSS constituencies need to be considered: the developer, the user, the decision-maker, the chauffeured decision-maker, the decision-consumer and the management, and the evaluation criteria should be tailored accordingly.

In this work, we adopt the evaluation approach introduced by Maynard et al. [11] to validate a mobile CDSS for field triage named iTriage [12]. The iTriage system aims to assist healthcare practitioners with determining the priority of patients based on standard SIEVE guidelines at mass gathering events [12]. In the earlier study [12], the iTriage system was evaluated by the developer constituency in terms of only accuracy. In this work, we extend this study and present our validation of iTriage by conducting a more comprehensive multi-criteria evaluation that involved highly trained paramedics with the constituency roles of the DSS user and decision-maker. According to the constituencies involved, we apply a set of suitable evaluation criteria from the four domains of effectiveness, efficiency, satisfaction and use [11].

1. Triage Decision Support System

Triage is a process of sorting patients based on their need for immediate medical treatment according to certain guidelines [13]. It enables prioritizing patients on the basis of the severity of their condition to receive medical facilities in a timely manner [14]. The word “triage” originates from a French word “trier” which means to sort, pick out, classify, or choose [15]. There are two main types of triage: i) hospital triage (or the triage sort) which is generally conducted in at a casualty clearing station or hospitals, and ii) field triage which is performed out the hospital, e.g. in environments like mass gathering events where a large number of people gather, and involves a rapid assessment [16]. There are commonly four labels (R, Y, G, and B) and categories of triage [15]. These include: i) the red label (R) indicates the immediate priority casualties (Priority 1), ii) the yellow/orange color (Y) which is used for those who require urgent treatment (Priority 2), iii) the green color (G) representing delayed priority causalities where patients have minor injuries (Priority 3), and iv) the last category presented by black/white (B) is used for dead causalities (No Priority).
Studies show that triage nurses often take decisions based on past experiences or encounters of certain symptoms, and the use of a standardized guideline in triage CDSS can improve accuracy of triage decisions [16]. Mobile field triage DSS can speed up the patient categorization and enable easy access to collected data on the move [8, 12]. This paper focuses on the evaluation of iTriage [12].

1.1. An Overview of the iTriage DSS

The iTriage system is a mobile field triage decision support system developed for Android-based smart phones. The system enables efficiently sharing and visualizing triage data at mass gatherings and assisting administrators in determining the exact state of the situation [12]. Figure 1 shows a simple user interface of iTriage where the triage decision is presented to the user (e.g. Priority 3) based on the patient information.

![Figure 1. Screenshots of main screens of the iTriage system.](image)

2. Evaluation of Clinical DSS

Computerized and mobile clinical DSS assist medical practitioners with decision making, and provide patient-specific recommendations in a timely manner [7]. Despite many benefits of CDSS, these systems need to be evaluated thoroughly to assess their correctness and their impact on healthcare delivery and patient outcomes [7, 8]. Prugnet et al. [5] conducted a systematic review to assess the impact of mobile handheld systems on hospital physicians’ work practices and patient care. The authors assessed the impact of mobile devices based the user (who), the activity (what), the location and time of activity (where and when), and how the activity was conducted.

There is limited existing research that concentrates on evaluating mobile triage CDSS. Michalowski et al. [7] conducted an evaluation of a Mobile Emergency Triage (MET) for patients with abdominal pain through a clinical trial in a hospital. The extended version of this system named MET-AP supports triage for acute pediatric abdominal pain [10] which was evaluated based on accuracy of triage decisions. San Pedro et al. [6] evaluated a PDA-based triage prototype to validate the underlying heuristic model in terms of its capability to provide effective recommendations. They also focused on studying any issues that a clinician might experience during the trial.
The triage prototype was further evaluated in another study using a lab experiment that involved twenty-nine nursing students [8]. The participants first performed the triage using the mobile triage system and then attempted the paper-based triage. The evaluation aimed to discover the differences between two groups. While, the findings showed that the PDA group overall performed better than the paper-based group in most test case scenarios, there was no significant difference found across the groups in terms of accuracy and consistency. The iTriage prototype developed by Jayaraman et al. [12] was evaluated by the system developer based on accuracy. The recommended decisions were compared to the actual triage categories of test cases.

Table 1 presents a summary of current triage CDSS studies. These works are analyzed according to the four domains proposed by Maynard et al. [11]. These domains include *efficiency, effectiveness, satisfaction, use*. The comparison table also considers the DSS constituency roles (i.e. *evaluators*) in the analysis.

The definitions of the four domains [11] are as follows: i) *Effectiveness* is defined as an extent to which a particular system is capable of producing the expected and accurate result and measured in terms of 29 criteria such as accuracy, ii) *Efficiency* focuses on the decision support systems’ level of performance and includes 12 criteria such as user productivity, iii) *Satisfaction* assesses the adequacy of the DSS from the evaluators’ perspectives and considers 29 criteria including user confidence, and iv) *Use* directly concentrates on how the DSS is used to achieve its intended goals.

The results in Table 1 show that most of the existing works perform evaluation of CDSS using a limited number of criteria. They mainly evaluate the mobile CDSS’ impact on healthcare delivery and patient outcomes in terms of accuracy without considering other aspects. Further, most works on evaluation of triage DSS involve domain experts such as physicians and nurses who are highly trained and qualified. There are also mobile CDSS like iTriage that aim to assist minimally-trained users with decision making in the resource-limited environments where there is a scarcity of qualified healthcare providers [4]. It is important that the evaluation of such systems considers different DSS constituencies and reflect their perspectives. The evaluation criteria also need to be selected according to the role of the evaluators. In the study by Jayaraman et al. [12], iTriage was evaluated by the developer based on accuracy. This evaluation does not fully assess the mobile system and its results are limited to only one view and measure. Conducting evaluations of mobile CDSS that involve different constituencies is surely an advantage because it enables to compare the results obtained from the different evaluations and avoid any unnecessary biases viewpoints [11].

<table>
<thead>
<tr>
<th>Studies</th>
<th>Effectiveness</th>
<th>Efficiency</th>
<th>Satisfaction</th>
<th>Use</th>
<th>Evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michalowski et al., 2005</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Physicians and residents</td>
</tr>
<tr>
<td>Farion et al., 2008 [10]</td>
<td>Yes (accuracy)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Physicians and residents</td>
</tr>
<tr>
<td>San Pedro et al., 2005 [6]</td>
<td>Yes</td>
<td>No</td>
<td>Yes (confidence)</td>
<td>Yes</td>
<td>One clinician</td>
</tr>
<tr>
<td>Padmanabhan et al., 2006</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Nursing students</td>
</tr>
<tr>
<td>Jayaraman et al. 2013 [12]</td>
<td>Yes (accuracy)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>A non-expert</td>
</tr>
</tbody>
</table>
3. Experimental Evaluation Design and Method

The experimental evaluation consisted of two stages. In the first stage of the trial, the participants were asked to enter their triage decisions for each test case on the paper. In the second stage, they used the iTriage system to enter the test cases data, observe the mobile system’s decision, and provide their own decisions if disagree with the device.

The multi-faceted evaluation approach [11] was adapted as a framework for this evaluation. It includes four domains where each domain provides a list of criteria. Figure 2 shows the criteria that we considered appropriate for the constituencies involved and relevant to the decision task and evaluation environment.

We selected accuracy for the effectiveness domain. Accuracy here refers to measuring the correctness of the triage decisions that the system and paramedics provided. Under the efficiency domain, we identify the user productivity and time taken for task accomplishment as the suitable criteria. User productivity assesses how the decision support system can improve the efficiency. The second criterion aims to assess if the mobile system allows users to perform triage within a shorter time frame compared to the paper-based triage. For the satisfaction domain, we consider four criteria. The user confidence criterion is used to measure the confidence and trust level of the evaluators and study how strongly the constituencies believe that the system performs correctly. Information quality focuses on the evaluators’ perspectives on the quality of information that the system provided. Interface applicability concentrates on the system interface and if it matches the triage guidelines. Information relevance assesses whether the information and output are related to triage decision making situations. Finally, ease of use is selected for the use domain and measures how easy the participants find the system to use.

3.1. Test Cases

Fifty test cases were selected randomly using a Kutools toolbox\(^2\) in Excel out of the test bank that consisted of 102 anonymous test cases provided by Ambulance Victoria. There was not any chance for a learning bias because of how data was randomly

allocated to each participant. Each test case contained the real patient data which comprised of the patient age, gender, patient condition, respiratory rate, pulse rate, symptoms and injuries. The cases had the same level of difficulty. These test cases were the same cases used in the study by Jayaraman et al. [12]. The authors identified five test cases that were misclassified by the iTriage system. In our experiment, we used those cases and deliberately included three randomly selected misclassified test cases in each experiment in order to investigate whether the participants would correct the misclassified cases by overwriting the answers suggested by the system.

3.2. Subjects

Participants included 11 qualified paramedics from Ambulance Victoria. Out of the eleven participants, six were females and five were males. All except one of the participants were frequent users of smart phones. Seven participants were under the age group of 18-30, two were under 31-40 and two were under 41-50. Five participants had more than five years of experience whereas the rest had a lesser experience. All the participants took part in both paper-based and mobile-based triage. A unique ID was assigned to each participant and test case.

3.3. Data Collection

Data was collected using three different forms: 1) response sheets, 2) questionnaire, and 3) open-ended questions.

The response sheets were used during both the paper-based and mobile-based triage experiments to allow participants to enter their triage decisions for each test case. Each response sheet contained 50 randomly-selected test cases. The response sheets were designed in a table format but differently for each experiment. In the paper-based triage, the participants were asked to enter their triage decisions for each test case along with a confidence factor value out of 100%. In the mobile-based triage, the participants first entered the triage decision provided by the mobile system. Then if they disagreed with the system decision, they provided their own decision with a confidence factor.

The questionnaire contained twenty two questions. The questions focused on the individual participant’s demographic data. The last section consisted of questions related to the use and interface of the system. This section used a 5-point Likert scale ranging from ‘strongly agree’ to ‘strongly disagree’.

The data collection also included eight open-ended questions to collect qualitative information about the participants’ views on the mobile triage DSS. The questionnaire and the open ended questions were designed according to our evaluation framework.

3.4. Statistical analysis

The inter-raters agreement between the individual participant’s responses and correct triage responses were estimated using weighted Cohen’s Kappa with quadratic weights [17] separately for both paper-based and mobile-based methods. Weighted Cohen’s kappa is selected because it is widely used in the literature to measure the agreement between two or more raters or devices in the situations for ordinal outcomes where they can agree or disagree by chance [18]. To compare the agreement with correct triage responses between two methods, individual differences in respective weighted Kappa’s were calculated for individual participants. For descriptive purposes, to summarize
these differences at the group level, median difference with corresponding Interquartile Ranges (IQRs – the ranges that include the middle 50% of the scores) were presented. To statistically investigate whether there is a difference in agreement with correct triage responses between two methods, 95% confidence intervals for the median difference in agreement were presented. Wilcoxon Sign ranked test [19] was used to assess whether the median difference in agreement was different from 0.

4. Results

This section discusses the findings for effectiveness, efficiency, satisfaction and use.

4.1. Effectiveness

Effectiveness was tested by applying the criteria of accuracy. The accuracy was measured considering the correctness of triage decisions given by the system and by the participants in the paper-based experiment and the mobile-based method, where they were allowed to overwrite the system decisions. Table 2 shows the accuracy results for paper-based and mobile-based triage answers. As the table shows the accuracy of the decisions recommended by the mobile system was the highest (94%).

<table>
<thead>
<tr>
<th>Mobile iTriage Decisions</th>
<th>Paper-based Decisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Decisions</td>
<td>465 (87%)</td>
</tr>
<tr>
<td>Incorrect Decisions</td>
<td>85 (13%)</td>
</tr>
<tr>
<td>Total</td>
<td>550</td>
</tr>
<tr>
<td>Correct Decisions</td>
<td>517 (94%)</td>
</tr>
<tr>
<td>Incorrect Decisions</td>
<td>33 (6%)</td>
</tr>
<tr>
<td>Total</td>
<td>550</td>
</tr>
</tbody>
</table>

To investigate overwritten decisions, a comparative analysis was conducted. Table 3 shows the accuracy of all the overwritten answers, and also analyses them further based on the participants’ level of experience. Results show that the participants with an experience of more than five years entered fewer incorrect answers (i.e. 29) when overwriting, compared to those with less experience (i.e. 42). Further, Table 2 and 3 show that accuracy of decisions dropped from 465 (87%) in the paper-based trial to 16 (18%) in the mobile-based experiment when the participants overwrote the decisions.

Moreover, this result is a good example where the role of the system evaluator can impact the evaluation. In the study by [12], the same 102 test cases resulted in 95% accuracy in the system evaluation by the developer who was a non-expert, and the same five test cases were misclassified by the system. In our research, the DSS users/decision-maker constituencies were allowed to overwrite the system decisions. Yet, they attempted to overwrite the correct decisions provided by the mobile system.

<table>
<thead>
<tr>
<th>Overwritten decisions by all the participants</th>
<th>Participants with less than 5 years of experience</th>
<th>Participants with more than 5 years of experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total overwriten answers</td>
<td>87</td>
<td>50</td>
</tr>
<tr>
<td>Correct Decisions</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Incorrect Decisions</td>
<td>71</td>
<td>42</td>
</tr>
</tbody>
</table>

As a result, the number of the correct answers in the mobile-based experiment was reduced compared to the paper-based method. This also highlights that mobile CDSS
cannot be fully trusted as an effective decision making tool but they can assist decision makers with tasks such as information management and easy access to guidelines.

4.2. Level of agreement

The levels of agreement between correct and incorrect answers/decisions for both experiments using the weighted Cohen’s kappa are presented in Table 4.

Table 4. Kappa for measuring the level of agreement between the participants’ and triage responses.

<table>
<thead>
<tr>
<th>Description</th>
<th>Median (min, IQR, max)</th>
<th>95%CI and Wilcoxon signed-rank p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper-based responses</td>
<td>0.83</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile-based responses</td>
<td>0.93</td>
<td>N/A</td>
</tr>
<tr>
<td>Mobile-based responses with overwrites</td>
<td>0.79</td>
<td>N/A</td>
</tr>
<tr>
<td>Difference in Kappa scores between paper and mobile-based responses</td>
<td>0.0</td>
<td>(-0.11; 0.16); p=0.86</td>
</tr>
<tr>
<td>Difference between paper and mobile-based responses with overwrites</td>
<td>-0.08</td>
<td>(-0.03; 0.17); p=0.06</td>
</tr>
</tbody>
</table>

As demonstrated in the first three rows of Table 4, the participants demonstrate good-to-excellent median agreement between the correct triage responses and paper-based, device-based, and device-based with overwrite responses respectively, with median weighted Kappa around 0.8. At the same time, there is a clear variability observed between the participants as emphasized by the ranges of agreement spanning the values of Kappa from 0.51 (average agreement) to 0.96 (excellent agreement).

Using the paper-based triage tool resulted in agreement that the correct triage responses exhibited a trend to being a statistically significantly higher than using the device-based tool with overwrite (median difference in agreement 0.08 (95%CI: -0.03, 0.17; p=0.06)) and they were not statistically significantly different to using the device without overwrite (median difference in agreement 0 (95%CI: -0.11, 0.16; p=0.86)).

5. Efficiency

Efficiency was evaluated by testing two criteria of time taken for task accomplishment and user productivity. To measure the first criterion, the start time and the finish time of both experiments for all the participants were recorded and the total time was calculated for each test. The results showed that all the participants completed the paper-based test at an average time of 15 minutes whereas it took them an average of 24 minutes to complete the mobile-based triage experiment. This could be due to the lack of training and familiarity of participants with the iTriage system.

The second criterion was user productivity. One of the related questions investigated if there were any aspects of the system that slowed done the information
entry. Seven participants felt that at some point the device created confusion. One of the participants reported:

“I believe with the format [Simple user format] we suggested the app would be clearer and quick to use”.

Another question aimed to determine if the device provided useful triage categorization information. Only six participants agreed. The other question asked if the participants had problems using the device at any stage. Similarly, six participants responded that they had no problems using the device while five of the participants disagreed. Two open-ended questions were focused on the consistency of results to measure productivity. Seven participants answered that the device helped to produce consistent results by reducing human errors. One of the participants who disagreed reported that system information was not consistent with the SIEVE standards and suggested that the results did not take into account severity of the injuries.

The last question concentrated on determining whether the device had the ability to be used for triage training or education purposes. Ten participants agreed but mentioned that the system needs to be further improved, particularly with respect to the interface. The examples of the participants’ suggestions regarding enhancing the system for the training purposes are as follows.

“Needs to be much easier to use”.

“Make the device have less options for a speedy result”.

5.1. Satisfaction

The satisfaction domain considered four criteria of user confidence, DSS interface, and relevance and quality of information.

The user confidence criterion was selected to determine if conducting triage with the help of mobile device would increase the level of confidence of the participants. In both paper-based and mobile-based experiments, the response sheets had an extra column that allowed participants to enter a confidence level for their triage decisions. The analysis results showed that the median and mode in both experiments were 100 but the means for paper-based and mobile-based triage was 93 and 92 respectively. Although the difference was very small, it indicated that the participants had a lower level of confidence while performing triage on a mobile device. This could due to some cases where they felt the need to overwrite the system proposed decision, but they were not fully confident while overwriting. This result requires further investigation.

The DSS interface applicability was assessed by seven questions using a Likert scale. The results are as follows.

- The first question was designed to evaluate whether the size of the text was easy to read. 82% of the participants agreed/strongly agreed.
- The second question was formed to gather information about the system navigation. 54% of the participants agreed/strongly agreed that it was easy to navigate through the screens and information items in the system prototype. 27% of the participants neither agreed nor disagreed to it.
- The third question was asked to establish whether the participants found the device easy to use. 54% participants agreed/strongly agreed that the prototype was easy to use while 27% participants neither agreed nor disagreed.
• The fourth question focused on the language. 81% participants agreed/strongly agreed that the language of information used was easily understandable.
• The fifth question aimed to determine if the participants found the design of the interface appealing. 45% agreed/strongly agreed and 27% disagreed/strongly disagreed.
• The sixth question was about the number of steps were acceptable to reach the necessary information. 63% participants agreed/strongly agreed.
• The last question investigated the prototype usefulness. 54% participants agreed/strongly agreed. 18% participants neither agreed nor disagreed.

The information relevance criterion was measured with a question that asked if participants could find the relevant screens with the desired information. Eight participants entered ‘Yes’ and one participant reported:

“*Yes as it is very user friendly and doesn’t leave room for personal interpretation.*”

Only three participants disagreed with the question and suggested that the information and its order should exactly follow the standard SIEVE triage guidelines that they use at work.

Quality of information was measured using different types of questions. An open-ended question aimed to determine whether any information was missing and needed to be added. Ten participants responded that there was no missing information. However, two of the participants disagreed and reported the following comments:

*Needed a lot more info about patient to make accurate triage*.

*Age of patient is relevant based on a person’s heart rate and respiratory rate*.

Another question aimed to determine whether the participants believed that the information to be entered on the device was up-to-date in accordance with triage standards. The majority of the participants believed that the information followed the standards and only four disagreed. Furthermore, a Likert scale style question was used to investigate whether the system provided complete information for decision making. 63% participants agreed/strongly agreed that all the information required to complete the triage process was available.

Regarding the satisfaction domain, further information was gathered. One of the participants responded that when using the mobile iTriage system they did not have to remember the triage labelling details, such as value ranges for heart rate or pulse rate of a patient. Another participant made the following comment:

*The future ability to gather information in one place and ability to be loaded on to the server. The ability to develop a ‘scoreboard’.*

Another question was about suggesting any improvements and enhancements that could enhance the mobile decision support system. Five participants reported that they would like the application to present more information about the patients, such as patient age and weight. Three participants mentioned that the interface and the navigation could be improved to save some time.

5.2. Use

The use domain included the criterion of ease of use. A question was used that focused on the best feature of the device and the reason for selecting it. Eight participants felt that the device (i.e. smart phone) was easy to use and convenient to carry. Two
participants also said that the device had great potential to be used in emergency departments. Two participants reported that there was no need to carry pen and paper when using the mobile device. The participants reported these comments:

“easy access, ease of use”.

“Good for a non-technical person to use.”

Another question focused on studying whether the paramedics would consider the mobile device as a hindrance, when carried along with any other triage equipment. Seven participants responded that they do not consider carrying the mobile device a problem in their job. However, one of the participants felt that there was a risk that the device could be lost or broken and there could also be a risk of information loss.

6. Discussion and Conclusion

In this paper we considered a possibility of supporting emergency management decisions with a mobile CDSS. The iTriage system has been developed to implement the Australian triage SIEVE method for out of hospital primary medical triage [20]. Taking the feedback from the previous evaluations, the system has been re-developed on a mobile phone and tested for its accuracy by developers with very promising results [12]. In this paper we have described and compared the perspectives of the practitioners on the potential of such system to assist them in their day-to-day triage decision-making. Although the number of participants was very small but is was a good representative of paramedics’ population in terms of mix of gender, age and experience level. We demonstrated how multiple-constituency and multi-objective DSS evaluation framework proposed by Maynard et al [11] can be successfully adapted for this purpose. The framework provided a uniform set of criteria and evaluation measures to solicit and compare perspectives of the developers and practitioners.

The results, although somewhat unexpected, clearly illustrate that high accuracy of any DSS cannot guarantee a positive view of the system performance when tested with the user-decision-maker. The results confirm that a mobile triage DSS can be treated as a decision support rather than a decision-making tool. The findings also reveal that the less experienced the decision-maker is, the more chance exists that they would not trust the decision made by the system, and if allowed, would actually overwrite it, and results can negatively impact the patient outcomes. To address this deficiency, it is important to allow sufficient training time with the system, and also make sure that the interface and usability of the system is fully tested according to the practitioners’ standard and routine decision-making settings and environment. The iTriage system was performing well in finding the “correct” outcome, but failed in supporting the process of decision-making to the satisfaction of the users. Hence, we posit, that as part of the introduction of modern technological solutions for CDSS development a thorough study of the decision-making processes during the system design and development stages is vital for the overall success of the proposed DSS solution. As informed by the multiple-constituency framework, all the stakeholder views have to be considered accordingly. The development of CDSS requires close participation of developers and practitioners. This will be the next step for our future research.
Acknowledgements

This research is partly supported by Australian Research Council funding (LP0453745). Support of the Ambulance Victoria staff is gratefully acknowledged.

References