Design and Development of an Objective Evaluation System for a Web-Based Simulator for Trauma Management

Blanca Larraga-García¹ Luis Castañeda López¹ Fernando Monforte-Escobar² Rubén Quintero Mínguez³ Manuel Quintana-Díaz⁴ Álvaro Gutiérrez¹

¹Escuela Técnica Superior de Ingenieros de Telecomunicación,

Universidad Politécnica de Madrid, Madrid, Spain

²SAMUR – Protección Civil, SAMUR-PC, Madrid, Spain

³Emergency Services SUMMA 112, Madrid, Spain

⁴Servicio de Medicina Intensiva, Hospital La Paz Institute for Health Research, IdiPAZ, Madrid, Spain

Appl Clin Inform 2023;14:714-724.

Address for correspondence Blanca Larraga-García, PhD, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Avenida de Complutense, 30, 28040 Madrid, Spain (e-mail: blanca.larraga@upm.es).

Abstract **Background** Trauma injuries are one of the main leading causes of death in the world. Training with guidelines and protocols is adequate to provide a fast and efficient treatment to patients that suffer a trauma injury. **Objectives** This study aimed to evaluate deviations from a set protocol, a new set of metrics has been proposed and tested in a pilot study. **Methods** The participants were final-year students from the Universidad Autónoma de Madrid and first-year medical residents from the Hospital Universitario La Paz. They were asked to train four trauma scenarios with a web-based simulator for 2 weeks. A test was performed pre-training and another one post-training to evaluate the evolution of the treatment to those four trauma scenarios considering a predefined trauma protocol and based on the new set of metrics. The scenarios were pelvic and lower limb traumas in a hospital and in a prehospital setting, which allow them to learn and assess different trauma protocols. **Keywords Results** The results show that, in general, there is an improvement of the new metrics training after training with the simulator. **Conclusion** These new metrics provide comprehensive information for both trainers educational program simulation and trainees. For trainers, the evaluation of the simulation is automated and contains all relevant information to assess the performance of the trainee. And for trainees, it guidelines and provides valuable real-time information that could support the trauma management protocols emergency learning process.

Background and Significance

Trauma is an unpredictable medical emergency that requires a rapid response. Reducing the time to treatment to reduce the number of preventable deaths that occur during the first peak of trauma deaths remains a challenge.^{1,2} To provide this

received April 25, 2023 accepted after revision June 15, 2023 treatment, trauma specialists must respond immediately without hesitation. Therefore, proper training is required.^{3–6}

As noted by Wurmb et al,⁷ severely injured patients are those who need treatment that best avoids preventable errors because they have a direct impact on mortality and

© 2023. Thieme. All rights reserved. Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany DOI https://doi.org/ 10.1055/s-0043-1771396. ISSN 1869-0327. morbidity. Therefore, a standardized educational program is required for these patients. Advanced Trauma Training and Life Support (ATLS), developed by the American College of Surgeons Commission on Trauma, is recognized and practiced worldwide.⁸ Additionally, there is evidence that this training provides techniques and skills for the rapid and efficient treatment of trauma patients, but the principles need to be translated into local contexts.^{9–11}

Given the potential for unpredictable situations, there is a debate about the adequacy of trauma management guidelines and protocols.^{12–15} Nevertheless, there are studies that show clear improvements in trauma management after following trauma guidelines, clinical pathways, or protocols.^{16–23} These are three levels of standardized procedures, from more general guidelines first, to more detailed clinical pathways, then to protocols.^{12,24–28} However, what has not been discussed is the ability to study deviations from defined protocols to improve training, build flexibility in protocols, and constantly evaluate whether changes in analyzed protocols have been considered.^{29–36} How to develop and implement these standardized procedures is an important aspect that should be addressed^{37–44}; nevertheless, only a few studies have focused on their evaluation.^{27,45–49}

Objectives

Since trauma management includes simulation as an important part of training, the inclusion of an objective assessment system in simulation is a clear need.⁵⁰ The use of simulators to learn and more specifically practice standard trauma procedures is already a reality. New technologies make it possible by incorporating new modalities, as the current simulation environment is quite diverse. Therefore, different simulation modalities would allow different medical aspects to be learned and practiced.⁵¹⁻⁵⁶ A web-based trauma simulator shows flexibility and a good possibility of implementing standard procedures that can be practiced by a large number of participants simultaneously, and which would allow the incorporation of an objective assessment system, which is a clear need in simulation.^{57–59} To do this, it is important to consider all possible treatments for one patient and keep in mind that some actions may be equivalent, similar, or completely different.

Current assessment methods available within simulation consider written assessment checklists in which simulation information is not automatically linked. Thus, the objective of this article is to develop and implement an objective evaluation system within a web-based trauma simulator to obtain objective information about how the simulation is performed in real time.

Methods

Several aspects should be considered when evaluating how a sequence of actions is performed in a simulation environment. On one hand, the number of correct and incorrect actions is important, but on the other hand, the order of performing different actions is important. Therefore, a dis-

tance metric should be defined to measure the distance between the sequence of actions performed by the trainee and the sequence of actions that should be performed according to the trauma protocol. There are several distance measures: edit distance, token-based distance, and seguence-based distance.⁶⁰ Edit compares the distance between two strings by counting the minimum number of operations required to convert one string to another. Tokenbased distance compares two strings by checking string units (tokens). The sequence-based distance of two strings compared them by examining different sequences of the strings. From these criteria, it was decided to focus on the Needleman–Wunsch algorithm.⁶⁰ This is because it is an edit distance algorithm that allows for different considerations: if there are matches, mismatches, or gaps between two sequences. This gives us the flexibility to tailor the algorithm to compare two trauma management sequences. In addition, two other metrics were established. One is the diagonal score (DS), which provides information about the correct sequence of actions, and the other is the subsequence score (SS), which provides information about the number of subsequent sequences of actions performed during the simulation.

Modified Needleman–Wunsch algorithm

Initially, the Needleman–Wunsch algorithm appeared as a dynamic programming algorithm that would provide a global solution to the problem of comparing two amino acid sequences.⁶¹ This global alignment (GA) would allow the distance between two sequences to be measured. This algorithm consists of three steps: the first is to initialize the score matrix, the second is to calculate the scores to construct the traceback matrix, and finally, the third step is to derive the best alignment from the traceback matrix.

Following this algorithm, a modification has been included considering that some actions may have the same impact on the patient's vital signs, some others may have a similar impact, some others may be completely opposite, and some others may just be different. All these options are included in the algorithm to create a traceback matrix that would then allow finding the best alignment. The way the score is calculated to obtain the traceback matrix, D(i, j), is as follows: (Eq. 1)

$$D (i - 1, j - 1) + s (x_b, y_i)$$
$$D (i - 1, j) + S_g$$
$$D (i, j-1) + S_g$$
$$D (i, j) = \max (1)$$

being s (x_i , y_i) different scores with respect to four different possibilities: actions that match, S_{match} , actions that are equivalent and therefore could be swapped, S_{swap} , actions that are opposed, $S_{contrary}$, or actions that are just different and, therefore, they are considered a mismatch, $S_{mismatch}$. The score provided to a gap is S_{g} .

Assuming that the sequence of actions to be performed is: "5786324196" and that a trainee performed the following



Fig. 1 Example of the sequence to accomplish: "5 7 8 6 3 2 4 1 9 6" and the sequence of actions performed by the trainee "7 4 1 3 2 8 6 9."

Traceback Matrix Best Alignment i 5 7 8 6 3 2 4 1 9 6 5 7 8 6 3 2 1 9 6 -4 -1 -2 -7 -9 -10 -2 -7 -9 -10 0 -3 -4 -5 -6 -8 0← -3 -5 -6 -8 9 3 4 7 -1 -2 8 7 6 5 4 2 1 -1 -2 3 2 1 7 -2 -3 8 6 5 4 15 14 13 12 -2 -3 8 7 6 5 4 15 14 13 12 4 7 4 5 3 14 25 24 23 -3 7 6 5 4 3 14 25 24 23 -3 7 6 4 -4 1 -4 1 -4 -5 6 5 4 15 14 13 24 23 22 -4 -5 6 5 4 15 14 13 24 23 22 3 3 5 3 14 25 24 23 22 21 -5 -6 4 3 14 25 24 23 22 21 2 -5 -6 5 4 2 -7 4 15 14 13 24 23 22 21 20 -7 4 15 14 13 24 23 22 21 20 -6 8 -6 8 3 14 25 24 23 22 21 20 31 3 14 25 24 23 22 21 20 -7 -8 -7 6 -8 6 2 13 24 23 22 21 20 31 3⁰ 2 13 24 23 22 21 20 31 30 -9 -8 -9 -8 g 9

Fig. 2 Example of the traceback matrix calculation and the best alignment obtained for two sequences: "5 7 8 6 3 2 4 1 9 6," which is the sequence that should be done and "7 4 1 3 2 8 6 9," which is the sequence of actions performed by the trainee.



Fig. 3 Best sequence alignment obtained after applying the modified Needleman–Wunsch algorithm for two sequences: "5786324196," the ideal one and "7 4 1 3 2 8 6 9," the one performed by the trainee.

sequence: "7 4 1 3 2 8 6 9" as shown in **► Fig. 1**; the traceback matrix is built (**► Fig. 2**) considering the values of the scores mentioned above and the best alignment obtained is shown in **► Fig. 3**.

Therefore, the best sequence alignment is **Fig. 3**. The blue color in **Fig. 3** means that a gap has been introduced in the sequence, which means that nothing has been done or should not be done to better fulfill the sequence of actions that the student should take. The green color in **Fig. 3** means that the actions between the two sequences match, and the dark red color means that there is a mismatch. In the best alignment path shown in **Fig. 2**,

the vertical arrows indicate gaps entered in the upper level of the matrix, diagonal arrows indicate that there is a match or mismatch, and horizontal arrows indicate gaps introduced in the sequence on the left side of the matrix. Then in this example, the best sequence alignment has four matches, one mismatch, and eight gaps. The maximum score is given to the match and then the swap. Later, if there is a gap, the score provided is higher than the contrary score and the mismatch score as when dealing with trauma patients, doing something that should not be done is worse than doing nothing. Ultimately, the worstcase scenario is taking an action that has the opposite impact to what the patient should have. A total of 15 trauma experts were asked to define the four categories of actions above. For example, actions that are considered equivalent are "oxygenate the patient with an oxygen mask," "oxygenate with a self-inflating bag," and "use an oropharyngeal airway."

Therefore, all scores set to fulfill the traceback matrix shown in **-Fig. 2** follow the criteria: $S_{match} > S_{swap} > S_{gap} > S_{contrary} > S_{mismatch}$.

Then, once the traceback matrix is created and the best sequence for comparison is obtained, a score is provided for that comparison. This score is called the GA score and is calculated as follows: (Eq. 2)

Eq. 2:

 $GA = n_{match} S_{match} + n_{swap} S_{swap} + n_{gap} S_{gap} + n_{mismatch} S_{mismatch} + n_{contrary} S_{contrary} (2)$

where n_{match} , n_{swap} , n_{gap} , n_{mismatch} , and n_{contrary} are the number of matches, swaps, gaps, mismatches, and contrary actions that exist within the best alignment. Given the maximum and the minimum possible punctuation values, this score is normalized to GA ε [-1, 1]. A negative value of this score means that the two sequences are different, and a positive value means that they are similar. The higher the value, the more different and the more alike are the two sequences. Therefore, this score provides information on how the trainee performed the simulation, considering all the options above.

Diagonal Score

This score is created to provide information about the correct actions taken during the simulation. If an action is executed at the correct time, it will receive more points than if it is executed at any other time during the simulation. Therefore, a score matrix, S (*i*, *j*), is built representing the actions performed by the trainee as rows and the actions to be performed as columns. Then, if the two actions match, a score of 1 is entered into the matrix, otherwise zero is entered as shown in Eq. 3.

1, when
$$seq1(i) = seq2(j)$$

0, when $seq1(i) \neq seq2(j)$

seq1 is the sequence of actions performed by the trainee and *seq2* is the sequence of actions that should have been performed. \rightarrow Fig. 4 shows an example of how to construct a score matrix. The sequence "7 4 1 3 2 8 6 9" is performed by the trainee and the sequence "5 7 8 6 3 2 4 1 9 6" is the one that should have been performed (\rightarrow Fig. 4).

To calculate the DS, the values in the score matrix called individual scores are summed along the diagonal of the matrix, $diag_i$, being *i* the number of the diagonals. Following this same example, the values of the diagonals are shown in **Fig. 5** with the subsequent values: $diag_{-7}=0$, $diag_{-6}=0$, $diag_{-5}=0$, $diag_{-4}=0$, $diag_{-3}=2$, $diag_{-2}=0$, $diag_{-1}=0$, $diag_{0}=0$, $diag_{-1}=0$

(5	7	8	6	3	2	4	1	9	6)
7	0	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0
8	0	0	1	0	0	0	0	0	0	0
6	0	0	0	1	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0)

Fig. 4 Example of a score matrix for two sequences: "578632419 6" and "74132869," in which *S* (*i*, *j*) is 1 when the two sequences are identical and 0 when they are not.



Fig. 5 Diagonals used to sum the individual values of the score matrix, *S* (*i*, *j*), for two sequences: "5 7 8 6 3 2 4 1 9 6" and "7 4 1 3 2 8 6 9."

=0, $diag_1$ =3, $diag_2$ =0, $diag_3$ =1, $diag_4$ =0, $diag_5$ =2, $diag_6$ =0, $diag_7$ =0, $diag_8$ =0 $diag_9$ =0 (**>Fig. 5**).

Then, they are all squared and summed up together according to Eq. 4 to obtain the DS:

(Eq. 4)
$$DS = \sum_{i=0}^{n} diag_{i}^{2} (4)$$

being *i* the number of diagonals from 0 to *n*. Considering that the maximum DS is obtained when the sequences are identical, this score is normalized. The values of this score: DS ε [0, 1]. As the values provided are always positive, the maximum value obtained when two sequences are identical is 1 and the value 0 means that they are completely different.

Subsequences Score

This score focuses on identifying the correct subsequences performed by the trainee and the length of each subsequence. As mentioned for DS, a matrix S(i, j) is built with both sequences containing a 1 when two actions match and a 0 when they do not match. This SS identifies the number of



Fig. 6 The sequence of actions performed by the trainee: "7 4 1 3 2 8 6 9" finds a first subsequence comparing it with the actions that should have been performed: "5 7 8 6 3 2 4 1 8 6" when he or she performs the actions [4,1], then a second one [3,2] and finally, the actions [8,6].

actions that are executed in order and join them together until an action that should not be executed is found. This is done by comparing the value of S(i, j) and the value of S(i + 1, j + 1) according to Eq. 5.

If
$$S(i, j) = 1$$
 and $S(i + 1, j + 1) = 1 \rightarrow$
seq1 (i) = seq2 (j) and seq1 (i + 1) = seq2 (j + 1)
If $S(i, j) = 1$ and $S(i + 1, j + 1) = 0 \rightarrow$
seq1 (i) = seq2 (j) and seq1 (i + 1) \neq seq2 (j + 1)

(Eq. 5)

The actions that fulfill the first condition are included into a subsequence vector until the value of S(i + 1, j + 1) = 0, which means that the values of two sequences are not the same. When this is the case, the subsequence ends, and the algorithm tries to find other subsequences that might appear. By applying this algorithm in the example presented, the subsequences identified are: [4, 1], [3, 2], and [8, 6]. Therefore, as shown in **– Fig. 6**, there are three subsequences with a length of two (**– Fig. 6**).

Then, SS is calculated as follows: (Eq. 6)

$$SS = \left(\frac{\sum_{i=0}^{n} \text{Length of individual subsequences}}{\text{Lenght of the total sequence}}\right) \times \frac{1}{n}, (6)$$

being *i* the number of subsequences from 0 to *n*. The value of this score is also normalized to: SS ε [0, 1]. As for the DS, considering the values provided to the score matrix and that the length of the sequences is always positive, a 1 means that there is a unique subsequence which matches entirely with the sequence of actions that should have been done, and 0 means that there is not a single subsequence in the actions performed by the trainee.

Pilot Study

Once these scores are set, a pilot study is accomplished using the web-based simulator developed in Larraga-García et al.⁶² This pilot study was accomplished with 24 participants: 14 final-year medical students from Universidad Autónoma de Madrid and 10 first-year residents from Hospital Universitario La Paz. On the first day, they were asked to manage four different trauma scenarios. Then, for the next 2 weeks, the participants trained with the simulator at least one simulation per day. After each simulation, the trainee could download a pdf document with all the steps taken as well as their impact along the simulation. Finally, the first four different trauma scenarios were repeated. Within this pilot study, 91 simulations were performed before training with the webbased simulator and 66 simulations were repeated after the training. This difference is due to the fact that not all the trainees performed all the scenarios after the training period. Then, the scores previously defined are obtained for the simulations pre- and post-training. To compare the evolution, a Wilcoxon's rank-sum test was used, and statistical significance is obtained when the *p*-value is lower than 0.05.⁶³

Results

Once the pilot study was accomplished, all the data were gathered and analyzed. The four trauma scenarios trained were: a prehospital pelvic trauma, a hospital pelvic trauma, a prehospital lower limb trauma, and a hospital lower limb trauma. All the results are shown comparing the pre-training simulation, which is shown in blue, and the post-training simulation, which is in orange. From the 91 simulations performed pre-training, 26 are prehospital pelvic, 22 are hospital pelvic, 23 are prehospital lower limb, and 20 are hospital lower limb scenarios. From the 66 simulations performed post-training, 17 are prehospital pelvic, 17 are hospital pelvic, 16 are prehospital lower limb, and 16 are hospital lower limb scenarios.

Modified Needleman–Wunsch Algorithm

The evolution of the GA score obtained with the modified Needleman–Wunsch algorithm, pre- and post-training, is shown in \succ Fig. 7.

This score remains practically the same in - Fig. 7(a, c), having the same median value in the first case and a slightly lower value in the second one. In - Fig. 7a, the median values are zero, whereas in - Fig. 7c, the median values are negative. Nevertheless, in - Fig. 7b, a drop is shown in the GA score being in both, pre- and post-training, a negative value. Finally, in - Fig. 7d, a clear improvement is shown pre- and post-training showing statistical significance with a *p*-value of 0.025. All the relevant values are shown in - Table 1, in which the median values, the interquartile ranges, the statistics, and the *p*-values are presented.

Diagonal Score

The DS improves in all scenarios except in the prehospital lower limb scenario in which it slightly drops. In **~ Fig. 8(a**, b, d), the DS improves showing statistical significance only in the hospital lower limb scenario (**~ Fig. 8d**), with a *p*-value of 0.03.



Fig. 7 Global alignment scores pre- and post-training for different trauma scenarios: (A) prehospital pelvic trauma scenario, (B) hospital pelvic trauma scenario, (C) prehospital lower limb trauma scenario, and (D) hospital lower limb trauma scenario.

	Median	Q1	Q3	Stats	p-Value	Pre/post
Prehospital pelvic	0.0	-0.06578	0.10045	0.36343	0.14660	Pretraining
	0.0	-0.13544	0.04574			Posttraining
Hospital pelvic	-0.02542	-0.15477	0.10541	1.61158	0.20426	Pretraining
	-0.18758	-0.20147	-0.01478			Posttraining
Prehospital lower limb	-0.10445	-0.14978	0.01445	1.88409	0.69861	Pretraining
	-0.11254	-0.15007	-0.03014			Posttraining
Hospital lower limb	-0.17585	-0.20144	-0.02504	2.88435	0.02500	Pretraining
	-0.09887	-0.11578	0.02355			Posttraining

Table 1 Main results of the global alignment scores for all the trauma scenarios

In the case in which the DS decreases the median value, it drops only from 0.0468 to 0.0443. All the relevant values are shown in **~Table 2**, in which the median values, the interquartile ranges, the statistics, and the *p*-values are presented.

Subsequences Score

For the subsequences score, the median value improves in all cases in the post-training simulation. However, **► Fig. 9a** shows, in some simulations, that the subsequences score values are better in the pre-training simulation than in the

from 0 to 0.125 in the post-training simulation. All the relevant values are shown in **—Table 3**, in which the median values, the interquartile ranges, the statistics, and the *p*-values are presented.

post-training one. Nevertheless, the median value increases

Discussion

The obtained results show that the simulations improve after training with the web-based simulator. Nevertheless,



Fig. 8 Diagonal scores pre- and post-training for different trauma scenarios: (A) prehospital pelvic trauma scenario, (B) hospital pelvic trauma scenario, (C) prehospital lower limb trauma scenario, and (D) hospital lower limb trauma scenario.

	Median	Q1	Q3	Stats	p-Value	Pre/post
Prehospital pelvic	0.04687	0.03472	0.05769	0.36343	0.54660	Pretraining
	0.05228	0.04273	0.06923			Posttraining
Hospital pelvic	0.04769	0.04127	0.06201	0.01747	0.89483	Pretraining
	0.05128	0.03921	0.06070			Posttraining
Prehospital lower limb	0.04687	0.03125	0.05555	0.02078	0.88537	Pretraining
	0.04437	0.03375	0.05236			Posttraining
Hospital lower limb	0.03703	0.03047	0.04887	4.69810	0.03019	Pretraining
	0.05428	0.03928	0.07164			Posttraining

 Table 2
 Main results of the diagonal scores for all the trauma scenarios

improvements are clearly seen in the diagonal and SSs, whereas in the GA scores these improvements are not so clear. This may be because the diagonal and SSs consider only a unique aspect. The DS analyzes the correct sequence of actions, and the SS examines the number of correct sequences and their length. However, the GA score takes into account several aspects. This score provides additional information that can be provided as real-time feedback to trainees such as when and in what order to perform the actions. The web-based simulator used for this pilot study did not provide this realtime feedback to trainees during the simulation, which may explain the poor improvement achieved for this score.

Hospital Pelvic Trauma Scenario

Interestingly, for hospital pelvic trauma, the DS increases, meaning more actions are performed correctly at the right time, but the GA score decreases. This is because even if the right actions and the right moments are improved, there is still a significant number of mismatches, gaps, or contrary actions. It is important to emphasize that mismatches, gaps,



Fig. 9 Subsequences scores pre- and posttraining for different trauma scenarios: (A) prehospital pelvic trauma scenario, (B) hospital pelvic trauma scenario, (C) prehospital lower limb trauma scenario, and (D) hospital lower limb trauma scenario.

	Median	Q1	Q3	Stats	p-Value	Pre/post
Prehospital pelvic	0.0	0.0	0.25	0.36343	0.54660	Pretraining
	0.125	0.0	0.13333			Posttraining
Hospital pelvic	0.0	0.0	0.09090	1.99754	0.15755	Pretraining
	0.125	0.0	0.16666			Posttraining
Prehospital lower limb	0.0	0.0	0.0	1.17371	0.27863	Pretraining
	0.0	0.0	0.1357			Posttraining
Hospital lower limb	0.039	0.035	0.0498	0.0	1	Pretraining
	0.04	0.0	0.1357			Posttraining

Table 3 Main results of the subsequences scores for all the trauma scenarios

or contrary actions have negative score values in the GA score. Thus, obtaining a negative GA score provides information about the balance between positive actions, similar and equivalent actions taken, and negative actions, opposite actions, mismatches, and gaps. Indeed, the Needleman-Wunsch literature provides different score values for match, mismatch, and gap penalties, but there are no guidelines that provide recommendations on which criteria should be used. In general, a negative value for the GA score means that the two sequences are different, and a positive value means that they are similar. In addition, the inclusion of different scores for matches, swaps, gaps, contrary, and mismatches in the GA score provides more comprehensive information for trauma management. Furthermore, it is important to include this information in web-based simulators. This provides realtime information about student performance, highlights errors, and suggests what to do next. Integrating this information into the simulator supports the learning rate of the trauma protocol. The DS and SS provide information not included in the GA score, providing a more comprehensive assessment of the simulation performance according to predefined clinical trauma guidelines.

Prehospital Pelvic Trauma Scenario

Considering the prehospital pelvic trauma scenario, there are improvements on the diagonal and the SSs, indicating better trainee performance in the post-training simulations. However, no statistical significance is obtained. Furthermore, the GA score has not improved. This improvement is less clear compared with the hospital pelvic trauma scenario, as there is a small improvement in the DS and the GA score worsens. Nevertheless, there is an improvement on the SS, but it does not show statistical significance. Then, an analysis of these simulations was performed with trauma specialists considering that there was an improvement on how well the trainees performed after the training. Nevertheless, incorporating all the information provided by these new metrics would support the trauma training process. This allows us to identify areas that need further reinforcement to enhance learning of trauma management protocols.

Prehospital Lower Limb Trauma Scenario

The prehospital lower limb trauma scenario decreases on the GA and DSs, but slightly improves SSs. This contrasts with the apparent improvement in hospital lower limb trauma scenario.

Hospital Lower Limb Trauma Scenario

In this scenario, the improvement is statistically significant for global and diagonal alignment scores, with *p*-values of 0.014 and 0.03, respectively. This is the only trauma scenario where a statistical difference is observed. This means that hospital management for lower limb injuries may be better learned whereas the prehospital setting must be improved.

Conclusion

The goal of this article was to develop and implement an objective evaluation system obtaining real-time objective information. To do so, new metrics have been developed. This supports the need to include, in clinical simulation, an objective evaluation system and, additionally, to provide valuable and objective information about how the simulation has been performed.

These metrics have been successfully built and implemented in a web-based simulator to evaluate trauma protocols to treat two specific trauma injuries: pelvic and lower limb lesions and they could be easily adapted for other types of traumas.

Additionally, these metrics could be used to provide realtime information to trainees during the learning process, providing information with respect to the actions that are equivalents or the ones that are not done in order, amongst other. This would make the simulation a self-explanatory learning tool that could be adapted to different levels of expertise. Moreover, these metrics are a powerful tool for trainers as they could objectively evaluate a simulation with all the data coming directly from the simulator.

Nonetheless, this pilot study should be extended to a larger community and to other trauma injuries to continue analyzing the evolution of the trauma management learning process. Additionally, with a larger sample of studies, a deeper statistical analysis should be performed.

Clinical Relevance Statement

Trauma management training is of key importance considering that trauma is one of the leading death causes worldwide. Trauma deaths have followed a classical trimodal distribution and even though the epidemiology of these deaths has changed since the year 2000 toward a bimodal distribution, the number of immediate deaths is still quite high. Therefore, trauma training remains to be a necessary task. There are several trauma trainings that have been created since the ATLS started in 1978 which use different simulation modalities to achieve the goal to train on trauma management skills. Considering that not so many studies in the literature were found incorporating an objective evaluation system, this research presents metrics to evaluate trauma management learning. This allows us to, objectively and automatically, evaluate knowledge acquisition. Therefore, incorporating this information in real time would help encourage trainees and help them acquire the skills more efficiently.

Multiple Choice Questions

- 1. Which is the third step of the Needleman–Wunsch algorithm?
 - a. The initialization of the score matrix
 - b. The calculation of the individual scores
 - c. The traceback matrix
 - d. To find the best alignment

Correct Answer: The correct answer is option d. The calculation of the individual scores can be changed by together with the traceback matrix is the second step as to di one, implies to accomplish the other one. Therefore, the third step is to find the best alignment.

- 2. Which type of action evaluates if the trainee did not perform an action that should have been performed?
 - a. A mismatch action
 - b. A contrary action
 - c. A gap action
 - d. A swap action

Correct Answer: The correct answer is option c. A gap means that either nothing has been done in the sequence of actions performed by the trainee when he or she should be doing something according to the trauma management protocol to evaluate.

3. Which is the score that provides information about how well the actions are accomplished with respect to the correct timing in which they should be done?

- a. Subsequence score
- b. Diagonal score
- c. Match score
- d. Swap score

Correct Answer: The correct answer is option b. The diagonal score is built to provide information with respect to the correct actions accomplished along the simulation considering that, if they are done at the correct timing, the score will be higher than if they are done in a different moment.

Protection of Human and Animal Subjects

This study did not involve human and/or animal subjects as the research objective.

Authors' Contributions

All authors whose names appear on the submission, (1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work; (2) drafted the work or revised it critically for important intellectual content; (3) approved the version to be published; and (4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-forprofit sectors.

Conflicts of Interest

None declared.

References

- 1 Ruchholtz S, Nast-Kolb D, Waydhas C, Betz P, Schweiberer L. Early mortality in polytrauma. A critical analysis of preventable errors [in German]. Unfallchirurg 1994;97(06):285–291
- 2 Ruchholtz S, Waydhas C, Aufmkolk M, et al. Interdisciplinary quality management in the treatment of severely injured patients. Validation of a QM system for the diagnostic and therapeutic process in early clinical management [in German]. Unfallchirurg 2001;104(10):927–937
- 3 Bouillon B, Kanz KG, Lackner CK, Mutschler W, Sturm J. The importance of Advanced Trauma Life Support (ATLS) in the emergency room [in German]. Unfallchirurg 2004;107(10): 844–850
- 4 Bernhard M, Becker TK, Nowe T, et al. Introduction of a treatment algorithm can improve the early management of emergency patients in the resuscitation room. Resuscitation 2007;73(03): 362–373
- 5 Hilbert P, zur Nieden K, Hofmann GO, Hoeller I, Koch R, Stuttmann R. New aspects in the emergency room management of critically injured patients: a multi-slice CT-oriented care algorithm. Injury 2007;38(05):552–558
- 6 Kanz KG, Körner M, Linsenmaier U, et al. Priority-oriented shock trauma room management with the integration of multiple-view spiral computed tomography [in German]. Unfallchirurg 2004; 107(10):937–944

- 7 Wurmb TE, Frühwald P, Knuepffer J, et al. Application of standard operating procedures accelerates the process of trauma care in patients with multiple injuries. Eur J Emerg Med 2008;15(06): 311–317
- 8 American College of Surgeons Committee on Trauma. Advanced Trauma Life Support. 10th ed., Chicago, IL: American College of Surgeons; 2018
- 9 Sturm JA, Lackner CK, Bouillon B, Seekamp A, Mutschler WE. Advanced Trauma Life Support (ATLS) [in German]. Unfallchirurg 2002;105(11):1027–1032
- 10 van Olden GD, Meeuwis JD, Bolhuis HW, Boxma H, Goris RJ. Advanced trauma life support study: quality of diagnostic and therapeutic procedures. J Trauma 2004;57(02):381–384
- 11 van Olden GD, Meeuwis JD, Bolhuis HW, Boxma H, Goris RJ. Clinical impact of advanced trauma life support. Am J Emerg Med 2004;22(07):522–525
- 12 Spain DA, McIlvoy LH, Fix SE, et al. Effect of a clinical pathway for severe traumatic brain injury on resource utilization. J Trauma 1998;45(01):101–104, discussion 104–105
- 13 Pritts TA, Nussbaum MS, Flesch LV, Fegelman EJ, Parikh AA, Fischer JE. Implementation of a clinical pathway decreases length of stay and cost for bowel resection. Ann Surg 1999;230(05):728–733
- 14 Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. BMJ 1998;316(7125):133–137
- 15 Campion FX, Rosenblatt MS. Quality assurance and medical outcomes in the era of cost containment. Surg Clin North Am 1996;76 (01):139–159
- 16 Curtis K, Asha SE, Unsworth A, et al. ChIP: an early activation protocol for isolated blunt chest injury improves outcomes, a retrospective cohort study. Australas Emerg Nurs J 2016;19(03):127–132
- 17 Carrie C, Stecken L, Cayrol E, et al. Bundle of care for blunt chest trauma patients improves analgesia but increases rates of intensive care unit admission: a retrospective case-control study. Anaesth Crit Care Pain Med 2018;37(03):211–215
- 18 Todd SR, McNally MM, Holcomb JB, et al. A multidisciplinary clinical pathway decreases rib fracture-associated infectious morbidity and mortality in high-risk trauma patients. Am J Surg 2006;192(06):806–811
- 19 Nyland BA, Spilman SK, Halub ME, et al. A preventative respiratory protocol to identify trauma subjects at risk for respiratory compromise on a general in-patient ward. Respir Care 2016;61(12): 1580–1587
- 20 Sahr SM, Webb ML, Renner CH, Sokol RK, Swegle JR. Implementation of a rib fracture triage protocol in elderly trauma patients. J Trauma Nurs 2013;20(04):172–175, quiz 176–177
- 21 Menditto VG, Gabrielli B, Marcosignori M, et al. A management of blunt thoracic trauma in an emergency department observation unit: pre-post observational study. J Trauma Acute Care Surg 2012;72(01):222–228
- 22 Morrison CA, Lee TC, Wall MJ Jr, Carrick MM. Use of a trauma service clinical pathway to improve patient outcomes for retained traumatic hemothorax. World J Surg 2009;33(09):1851–1856
- 23 Frederickson TA, Renner CH, Swegle JR, Sahr SM. The cumulative effect of multiple critical care protocols on length of stay in a geriatric trauma population. J Intensive Care Med 2013;28(01): 58–66
- 24 Dunham CM, Bosse MJ, Clancy TV, et al; EAST Practice Management Guidelines Work Group. Practice management guidelines for the optimal timing of long-bone fracture stabilization in polytrauma patients: the EAST Practice Management Guidelines Work Group. J Trauma 2001;50(05):958–967
- 25 Jacobs DG, Plaisier BR, Barie PS, et al; EAST Practice Management Guidelines Work Group. Practice management guidelines for geriatric trauma: the EAST Practice Management Guidelines Work Group. J Trauma 2003;54(02):391–416
- 26 Latini EE. Trauma critical pathways: a care delivery system that works. Crit Care Nurs Q 1996;19(01):83–87

- 27 Wilson S, Bin J, Sesperez J, Seger M, Sugrue M. Clinical pathwayscan they be used in trauma care. An analysis of their ability to fit the patient. Injury 2001;32(07):525–532
- 28 Friesdorf W, Konichezky S, Gross-Alltag F, Geva D, Nathe M, Schraag S. Decision making in high dependency environmentscan we learn from modern industrial management models? Int J Clin Monit Comput 1994;11(01):11–17
- 29 Baker E, Woolley A, Xyrichis A, Norton C, Hopkins P, Lee G. How does the implementation of a patient pathway-based intervention in the acute care of blunt thoracic injury impact on patient outcomes? A systematic review of the literature. Injury 2020;51 (08):1733–1743
- 30 Unsworth A, Curtis K, Asha SE. Treatments for blunt chest trauma and their impact on patient outcomes and health service delivery. Scand J Trauma Resusc Emerg Med 2015;23:17
- 31 Kourouche S, Buckley T, Munroe B, Curtis K. Development of a blunt chest injury care bundle: an integrative review. Injury 2018; 49(06):1008–1023
- 32 Roberts HC, Pickering RM, Onslow E, et al. The effectiveness of implementing a care pathway for femoral neck fracture in older people: a prospective controlled before and after study. Age Ageing 2004;33(02):178–184
- 33 Kwan J, Hand P, Dennis M, Sandercock P. Effects of introducing an integrated care pathway in an acute stroke unit. Age Ageing 2004; 33(04):362–367
- 34 Vanhaecht K, De Witte K, Panella M, Sermeus W. Do pathways lead to better organized care processes? J Eval Clin Pract 2009;15 (05):782–788
- 35 Evans-Lacko S, Jarrett M, McCrone P, Thornicroft G. Facilitators and barriers to implementing clinical care pathways. BMC Health Serv Res 2010;10:182
- 36 Hipp R, Abel E, Weber RJ. A primer on clinical pathways. Hosp Pharm 2016;51(05):416–421
- 37 Helbig K, Rmer M, Mellouli T. A clinical pathway mining approach to enable scheduling of hospital relocations and treatment services. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Switzerland: Springer; 2015:9253:242–250
- 38 Michalowski W, Wilk S, Thijssen A, Li M. Using a Bayesian belief network model to categorize length of stay for radical prostatectomy patients. Health Care Manage Sci 2006;9(04):341–348
- 39 Konrad R, Tulu B, Lawley M. Monitoring adherence to evidencebased practices: a method to utilize HL7 messages from hospital information systems. Appl Clin Inform 2013;4(01):126–143
- 40 Huang Z, Dong W, Bath P, Ji L, Duan H. On mining latent treatment patterns from electronic medical records. Data Min Knowl Discov 2015;29(04):914–949
- 41 Bahou N, Fenwick C, Anderson G, van der Meer R, Vassalos T. Modeling the critical care pathway for cardiothoracic surgery. Health Care Manage Sci 2018;21(02):192–203
- 42 Mans R, Schonenberg H, Leonardi G, et al. Process mining techniques: an application to stroke care. Stud Health Technol Inform 2008;136:573–578
- 43 Ben Othman S, Zgaya H, Hammadi S, Quilliot A, Martinot A, Renard J-M. Agents endowed with uncertainty management behaviors to solve a multiskill healthcare task scheduling. J Biomed Inform 2016;64:25–43
- 44 Joranger P, Nesbakken A, Hoff G, Sorbye H, Oshaug A, Aas E. Modeling and validating the cost and clinical pathway of colorectal cancer. Med Decis Making 2015;35(02):255–265

- 45 Sesperez J, Wilson S, Jalaludin B, Seger M, Sugrue M. Trauma case management and clinical pathways: prospective evaluation of their effect on selected patient outcomes in five key trauma conditions. J Trauma 2001;50(04):643–649
- 46 Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015; 350:h1258
- 47 Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013;50(05): 587–592
- 48 O'Cathain A, Croot L, Duncan E, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019;9(08):e029954
- 49 Boyko EJ. Observational research-opportunities and limitations. J Diabetes Complications 2013;27(06):642–648
- 50 Aspland E, Gartner D, Harper P. Clinical pathway modelling: a literature review. Health Syst (Basingstoke) 2019;10(01):1–23
- 51 Abelsson A, Rystedt I, Suserud BO, Lindwall L. Learning by simulation in prehospital emergency care an integrative literature review. Scand J Caring Sci 2016;30(02):234–240
- 52 Knudson MM, Khaw L, Bullard MK, et al. Trauma training in simulation: translating skills from SIM time to real time. J Trauma 2008;64(02):255–263, discussion 263–264
- 53 Patel D, Hawkins J, Chehab LZ, et al. Developing virtual reality trauma training experiences using 360-degree video: tutorial. J Med Internet Res 2020;22(12):e22420
- 54 Cohen D, Sevdalis N, Taylor D, et al. Emergency preparedness in the 21st century: training and preparation modules in virtual environments. Resuscitation 2013;84(01):78–84
- 55 Fleiszer D, Hoover ML, Posel N, Razek T, Bergman S. Development and validation of a tool to evaluate the evolution of clinical reasoning in trauma using virtual patients. J Surg Educ 2018;75 (03):779–786
- ⁵⁶ Taylor D, Patel V, Cohen D, et al. Single and multi-user virtual patient design in the virtual world. Stud Health Technol Inform 2011;163(01):650–652
- 57 Wise EM, McIvor WR, Mangione MP. Assessing student usage, perception, and the utility of a Web-based simulation in a third-year medical school clerkship. J Clin Anesth 2016;33: 5–13
- 58 Murray D, Boulet J, Ziv A, Woodhouse J, Kras J, McAllister J. An acute care skills evaluation for graduating medical students: a pilot study using clinical simulation. Med Educ 2002;36(09): 833–841
- 59 Hayes KA, Lehmann CU. The interactive patient: a multimedia interactive educational tool on the World Wide Web. MD Comput 1996;13(04):330–334
- 60 Aspland E, Harper PR, Gartner D, Webb P, Barrett-Lee P. Modified Needleman-Wunsch algorithm for clinical pathway clustering. J Biomed Inform 2021;115:103668
- 61 Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48(03):443–453
- 62 Larraga-García B, Castañeda López L, Rubio Bolívar FJ, Quintana-Díaz M, Gutiérrez Á Design and development of an interactive web-based simulator for trauma training: a pilot study. J Med Syst 2021;45(11):96
- 63 Lam F, Longnecker M. A modified Wilcoxon rank sum test for paired data. Biometrika 1983;70(02):510–513