Determining the equivalence of currently most used methods for evaluating cardiopulmonary resuscitation performance

Fabian Dhondt[1], Sandra Verelst[1], Jorg Roosen[2], Jan De Flander[2], Didier Desruelles[1], Philippe Dewolf[1]

Corresponding author: Dr Philippe Dewolf philippe.dewolf@uzleuven.be
Institution: 1. Department of Emergency Medicine, University Hospitals of Leuven, 2. KULeuven - University, Faculty of Medicine, Steps Skills Centre
Categories: Students/Trainees, Teaching and Learning, Clinical Skills, Research in Health Professions Education, Simulation and Virtual Reality

Received: 04/09/2020
Published: 12/01/2021

Abstract

Background

High quality cardiopulmonary resuscitation (CPR) is imperative to obtain a better outcome after a sudden cardiac arrest. However, a gold standard for training and evaluating CPR performance is lacking. In our institution, we recently changed from an observer only method (OOM) to a combined observer and electronic feedback method (EFM).

Objective

To determine whether an OOM and an EFM for evaluating CPR performance are equivalent and therefore can be used interchangeably.

Design

We performed a retrospective analysis on test results of CPR performance of medical students. In 2016, an OOM was used to evaluate CPR performance whereas in 2017, an EFM was added. The evaluation of hand placement, rate and depth of compressions, adequacy of ventilations, overall compression and ventilation score by both methods was compared.

Results
A total of 852 evaluation results using the OOM were compared with 713 results using the EFM. A significant discrepancy between the two methods was found on rate and depth of compressions, as well as on the volume of ventilations.

Conclusion

Based on these results, equivalence cannot be assumed between the two evaluation methods and they cannot be used interchangeably. Further research is needed to determine a gold standard for evaluating CPR performance.

Keywords: CPR; BLS; performance; evaluation; observer only method; electronic feedback method

Introduction

In the 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science, the evidence was reviewed for a series of actions that contributed to a better outcome after sudden cardiac arrest. Together, these actions were called the chain of survival. A crucial part of the chain of survival is high quality cardiopulmonary resuscitation (CPR) in terms of basic life support (BLS), consisting of the initiation of high-quality chest compressions and ventilations. Furthermore, recommendations were made for a specific algorithm and for the characteristics of good quality CPR. These different characteristics of CPR on which evidence based recommendations were made are: rate of compressions, depth of compressions, hand position, the minimization of pauses, and the compression-ventilation ratio (Perkins et al., 2015).

Educational institutions are responsible for the teaching of CPR. A previously conducted review by Yeung et al. found a benefit when using electronic feedback devices in CPR training (Yeung et al., 2009). In recent years, the use of electronic feedback systems has become widely spread (García-Suárez et al., 2019). However, at the moment a gold standard for evaluating CPR performance is lacking. Indeed, in a study by Oermann et al. in 2010 evaluating different training methods for CPR, participants were evaluated using an electronic feedback method (EFM) (Oermann et al., 2010). Other authors have also used electronic feedback systems to evaluate CPR proficiency (Partiprajak and Thongpo, 2016; Anderson et al., 2019; Cheng et al., 2018). On the other hand, a study published in January 2020 by Schmitz et al. relied on an observer only method (OOM) to evaluate the quality of CPR performance by emergency physicians (Schmitz et al., 2020).

Traditionally at our educational institution, an OOM was used that was standardised using a check list of the different critical steps of BLS. In 2017, we switched to an EFM via training manikins with the necessary sensors to measure the different parameters of BLS performance. The question arose whether, if both are available, the two methods should still be used interchangeably. Therefore, the aim of the present study was to determine whether the two methods are equivalent when it comes to evaluating CPR performance.

Methods

All students in their third and sixth year of medicine, who underwent the evaluation in 2016 and 2017, were included in this retrospective analysis. In both years, BLS training itself occurred without the use of an electronic feedback system for the student. If for any student no data were available due to a no show at the time of evaluation or due to a technical error, the results for this student were excluded from the analysis. We compared data obtained during the evaluation in 2016 with those obtained in 2017.
In 2016, an OOM was used to evaluate BLS performance. In 2017, an electronic feedback system was added. In both years, a standardised scenario was played out where the student was required to perform BLS. The evaluation consisted of the observer awarding a set amount of points for the correct execution of several actions. In 2016, there were a total number of 25 actions on which a score was given. Of these 25 actions, 17 related to either the scenario (for example being mindful of one's own safety, checking for consciousness) or the correct use of an automated external defibrillator. Five actions related to the performance of chest compressions and three related to ventilations. In 2017, the same list was used, but the eight actions relating to chest compressions and ventilations were replaced by a score produced using electronic feedback.

The electronic feedback system used in 2017 was the Laerdal SimPad®. The SimPad connects to the simulation manikin and measures a range of different parameters related to chest compressions and ventilations. The different aspects on which feedback is given are: compression rate, compression depth, incomplete release, hand position, number of compressions per cycle, ventilation volume, ventilation rate and flow time (which is the amount of time per cycle spent on compressions). From these parameters, except for flow time, an overall compression score and an overall ventilation score are calculated according to a predetermined algorithm by Laerdal. These two overall scores are then weighted together with the flow time to form a combined overall compression and ventilation score. All parameters are tracked in a non-binary way. In essence, each action performed within the limits of the guidelines is scored at 100%. When CPR performance deviates from the guidelines, the scores are reduced along S-curves outside of the thresholds. That means that small deviations create small score reductions, and larger deviations will generate substantial score reductions. The total score for compressions or ventilations is calculated by determining the mean of the scores of all the individual compressions or ventilations after which a combined overall score is generated. In this score, compressions are weighed double as compared to ventilations. Any interruption is penalized based on the chest compression fraction calculation (Laerdal, 2020).

Outcome measures

Both evaluation methods were assessed on which parameters could be compared. Ultimately, the following parameters were withheld: correct hand position on the chest, rate of compression, depth of compression, correct number and volume of ventilations per cycle, overall ventilation score, overall compression score, and the combined overall compression and ventilation score. Given the fact that the overall compression score and overall ventilation score in the OOM were scored in a binary way where points were either awarded or not, and in the electronic method were expressed as percentages, we determined a cut-off for these parameters that indicated the difference between pass or fail. Any score of 75% or higher was deemed sufficient and was awarded a pass, whereas any score under 75% was deemed insufficient and received a fail.

To determine an overall compression score in the OOM, we added the points for all elements related to compression but not to ventilation. The reverse was applied for the overall ventilation score. In this calculation, we included parameters that weren't measurable with the electronic feedback system such as the correct performance of the head-tilt/chin-lift procedure or the correct technique for the placing of the hands. For the combined overall compression and ventilation score in the OOM, we added all the points for all elements used in the overall compression and overall ventilation score.

A student received a pass score on the evaluation of an individual parameter when there was no loss of points on any of the elements the score was comprised of.

Statistical analysis
In order to determine the significance of the calculated difference in score for each parameter between the two observation methods, we used a two-proportion z-test. For each parameter, a z-score and corresponding p-value was calculated. We maintained a significance level (α) of 0.01 for each parameter.

**Results/Analysis**

In 2016, there were 853 recorded evaluation results using the OOM of which one was excluded due to a no show, leaving 852 test results to be included in the analysis. In 2017, there were 713 recorded evaluation results using the EFM. The number of students who achieved a pass in terms of correct hand position on the chest was 776/852 (91%) in the OOM versus 614/713 (86%) in the EFM (p = 0.002). The number of students who achieved a pass on the correct rate of compression was 714/852 (84%) in the OOM versus 343/713 (36%) in the EFM (p < 0.001). The number of students who achieved a pass on the correct depth of compression was 703/852 (83%) in the OOM versus 259/713 (36%) in the EFM (p < 0.001). The number of students who achieved a pass on the correct number and volume of ventilations per cycle was 652/852 (77%) in the OOM versus 266/713 (37%) in the EFM (p < 0.001). The number of students awarded a pass on the overall ventilation score was 584/852 (69%) in the OOM versus 461/713 (65%) in the EFM (p = 0.103). The number of students who achieved a pass on the overall compression score was 523/852 (61%) in the OOM versus 370/713 (52%) in the EFM (p < 0.001). The number of students who achieved a pass on the combined overall compression and ventilation score was 389/852 (46%) in the OOM versus 138/713 (19%) in the EFM (p < 0.001). (See Table 1 and Figure 1).

| Table 1: Pass rate for each element in the group evaluated using OOM as compared to the group using EFM |
|--------------------------------------------------------|----------------------------------------------------------|--------------------------|
| **Observer only method (n=852)** | **Electronic feedback method (n=713)** | **Z-score (p-value)** |
| The student's hands are placed at the correct position on the chest | 776 (91%) | 614 (86%) | 3.1 (0.002) |
| The rate of compressions is between 100-120/min | 714 (84%) | 343 (48%) | 15.02 (<0.001) |
| The depth of compressions is greater than 5cm but doesn’t exceed 6cm | 703 (83%) | 259 (36%) | 18.7 (<0.001) |
| Two ventilations of the correct volume are given per cycle | 652 (77%) | 266 (37%) | 15.69 (<0.001) |
| The overall ventilation score | 584 (69%) | 461 (65%) | 1.63 (0.103) |
| The overall compression score | 523 (61%) | 370 (52%) | 3.78 (<0.001) |
| The combined overall compression and ventilation score | 389 (46%) | 138 (19%) | 10.97 (<0.001) |

**Figure 1:** The percentage of students whose performance was deemed sufficient for each element in the group evaluated using the OOM (blue) as compared to the group using the EFM (red)
Discussion

With regard to the evaluation of BLS quality, there is evidence in the literature for the use of both the OOM and an EFM (Oermann et al., 2010; Partiprajak and Thongpo, 2016; Anderson et al., 2019; Cheng et al., 2018; Schmitz et al., 2020). We did not find a justification for the use of one over the other. Therefore, the question arises whether both methods can be used interchangeably. In practice, this means that if both methods are applied to a large group of students with the same level of training and practical exposure, they should highlight roughly the same strengths and weaknesses. Furthermore, one would expect the overall test result to be in the same range. The use of an EFM requires an added financial investment which would be hard to justify if both methods are equivalent. However, our analysis on the comparison of both evaluation methods showed a highly significant discrepancy between the two methods on the pass rate on especially the rate and depth of compressions, as well as on the volume of ventilations. In the OOM, these parameters are evaluated according to the judgement of the observer. In the EFM, the depth of each individual compression and the volume of ventilations are measured by a sensor in the manikin. The software assesses the time interval between every two subsequent compressions and subsequently calculates the average rate and the percentage of compressions achieved with an adequate time interval. The large difference in the test results between the two evaluation methods on these parameters might be explained by the fact that it is quite difficult to estimate these measurements by sight. Therefore, an objective measurement of these parameters is likely more reliable.

As for the overall ventilation score, no significant difference between the two evaluation methods was found. However, the overall ventilation score refers to an aggregated outcome measure that consists of different parameters in both evaluation methods.

Furthermore, it is possible that a significant proportion of students who received a pass result in one method would
fail in the other and vice versa, but that the overall percentage of students receiving a pass is similar between the two. (See Figure 2) In order to exclude this possibility, both evaluation methods should be compared synchronously in the same student population with the student blinded to the evaluation method.

**Figure 2:** Venn diagram illustrating the possibility of similar percentages but different populations that receive a pass result between the two different methods.

The fact that results differ significantly between the two evaluation methods on several crucial aspects of CPR could have wider implications. Since it does not seem possible to visually assess the quality of CPR adequately, there might be an added value with the use of a feedback device during real life CPR in an advanced life support setting. Although such a feedback device has been proposed in the past, it is not yet widely used (Sahyoun, Siliciano and Kessler, 2018). Indeed, it was shown that the quality of the providers’ CPR technique improved. To date, however, it has not been proven that this translates into an improved patient outcome in case of a cardiac arrest (Rapid and Service, 2015; An, Kim and Cho, 2019). Therefore, future research should focus on the combined approach of teaching and evaluating CPR with an electronic feedback system. Finally, real life CPR could be performed using an objective live electronic feedback system which could improve patient outcome following a sudden cardiac arrest.

**Conclusion**

In this retrospective analysis, where we compared two evaluation methods for the assessment of BLS performance by medical students, we found highly significant differences in test scores on rate of compressions, depth of compressions and volume of ventilations. No significant difference between the observer only method and the electronic feedback method was found for the overall ventilation score which may have been due to the fact that this involved an aggregated score built up of different parameters in each method. Based on these results, equivalence cannot be assumed between the two evaluation methods and they cannot be used interchangeably. In our opinion further research should focus on the combined approach of teaching and evaluating CPR with an electronic feedback system with the final aim of improving patient outcome following a sudden cardiac arrest.
Take Home Messages

- A reliable way of evaluating CPR performance is necessary to determine the adequacy of training and to evaluate the effect of interventions in research papers concerning CPR training and performance.
- A gold standard for evaluating CPR performance is lacking.
- Two groups of medical students with comparable backgrounds showed scores that differed significantly upon evaluation using an observer only method compared to an electronic feedback method.
- An observer only method for evaluating CPR was found not to be equivalent to an electronic feedback method. Interchangeable use is discouraged.
- Further research to determine the optimal way to evaluate CPR performance is necessary.

Notes On Contributors

**Fabian Dhondt** is an Emergency medicine trainee in his sixth post graduate year, scheduled to complete his training at the university hospital of Leuven in 2020.

**Sandra Verelst** is the head of the emergency department at the university hospital of Leuven. She obtained her PhD in the Department of Public Health and Primary Care in 2014. Besides a supervising position on various topics, her research interests include emergency department crowding and sepsis.

**Jorg Roosen** is a medical doctor at the KU Leuven. Currently full-time active in the medical simulation center STEPS developing and assisting a vast variety of simulation courses. Aside from simulation he has an extra interest in anatomical and orthopaedic research.

**Jan De Flander** CCRN, MHS. Worked for 25 years on an IC unit at the university hospital of Leuven before trading this environment for the medical skills lab of the faculty 6 years ago. There he introduced the electronic feedback system in CPR for educational, training and assessment purposes.

**Didier Desruelles** is an Emergency Physician at the university hospital of Leuven. His areas of interest are emergency medicine, mobile emergency medical assistance, urgent and intensive care for the critically ill patient, clinical toxicology, disaster medicine and management.

**Philippe Dewolf** is an Emergency Physician at the university hospital of Leuven and a Ph.D student in the Department of Public Health and Primary Care. His research interests include resuscitation and simulation. He is working on a project on the impact of a mixed reality platform for simulation in an advanced life support setting.

Acknowledgements

Figure 1-2: source the authors.

Bibliography/References


Appendices

None.
Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons "CC BY 4.0" (https://creativecommons.org/licenses/by-sa/4.0/)

Ethics Statement

Master Paper 015304 approved by the Research Ethics Committee Universitair Ziekenhuis/Katholieke Universiteit Leuven.

External Funding

This article has not had any External Funding