ELSEVIER

Resuscitation 34 (1997) 35–41

RESUSCITATION

Evaluation of cardiopulmonary resuscitation skills of general practitioners using different scoring methods

Jacobus J.M. Jansen[a,*], Hubert J.J.M. Berden[b], Cees P.M. van der Vleuten[c], Richard P.T.M. Grol[a], JanJoost Rethans[a], Charles P.M. Verhoeff[d]

[a]*Department of General Practice, University of Limburg, PO Box 616, 6200 MD, Maastricht, The Netherlands*
[b]*St. Radboud Hospital, University of Nijmegen, PO Box 9101, 6500 HB, Nijmegen, The Netherlands*
[c]*Department for Educational Development and Research, University of Limburg, PO Box 616, 6200 MD, Maastricht, The Netherlands*
[d]*Department of General Practice, University of Nijmegen, PO box 9101, 6500 HB, Nijmegen, The Netherlands*

Received 20 December 1995; revised 10 May 1996; accepted 21 June 1996

Abstract

In this study we evaluated the practical performance of 70 general practitioners in cardiopulmonary resuscitation (CPR) before and after instruction and compared checklist-based scores to mechanical recording scores in order to investigate which scoring method is preferable.

Both checklist and recording strip-based scores showed significant improvement after instruction, but only 37% were judged proficient according to the American Heart Association standards (checklist scoring), and 47% according to the recording print-based scoring system, while raters judged 97% as satisfactory by general impression. Interrater reliability was highest for the recording print (0.97) and lower for the checklist (0.79), especially for CPR performance (0.56). Comparison of checklist and recording print showed that the checklist was specific but not very sensitive in identifying poor performance for cardiac compression rate, since observers overestimated performance. The correlation for CPR performance between checklist score and recording strip score was low (0.45), indicating that candidates were ranked differently. The correlation between diagnosis and performance score was low for checklist as well as recording print (0.22), indicating that the score on diagnosis was a poor predictor for the score on performance of CPR.

These results support the use of the recording manikin as compared with the use of a checklist for formative evaluation of basic life support skills. However, as proficiency in diagnosis and performance in CPR are poorly correlated, assessment of diagnosis using a checklist must be included. Therefore we strongly recommend the combination of assessment by observers using a checklist for diagnostic procedures and the recording strip of the manikin for performance of CPR, as employed in most evaluation schemes.

Keywords: Resuscitation; Basic cardiac life support; Educational measurement; General practitioners; Scoring methods

1. Introduction

Acute myocardial infarction is a frequent cause of death in the developed world, with approximately two-thirds of the deaths occurring outside hospital [1]. Research evidence suggests that rapid initiation as well as correct technique of cardiopulmonary resuscitation (CPR) are essential links in the 'chain of survival' [2,3]. Since the majority of sudden deaths occur in the community, many lives could possibly be saved if adequate CPR skills were present throughout the community. General practitioners are confronted each year with 5–10 patients suffering from acute myocardial infarction [4,5]. The reported risk of cardiac arrest before reaching hospital varies from approximately 5% [6] up

* Corresponding author. Tel.: +31 43 3882323/+31 13 4676958; fax: +31 43 3619344; e-mail: koos.jansen@hag.rulimburg.nl

0300-9572/97/$17.00
PII S0300-9572(96)01028-3

to 25% [7]. In a recent survey in the Netherlands, general practitioners reported a mean performance of 2.0 CPR attempts per year [8]. Various studies have shown considerable deterioration in CPR skills among physicians, who had successfully completed prior courses in CPR [9–12], indicating that proficiency in these skills is not maintained.

For evaluation of competence in basic life support (BLS), checklists covering criteria of adequate performance are used [13] as well as recording strips of manikins [14]. In most research a combination of these methods is used [12,15–17], with checklist-based scoring for diagnostic procedures and the recording strip of the manikin for compression and ventilation procedures. The use of recording manikins permits assessment of outcome criteria (e.g. breathing volume and thorax compression depth) and some aspects of process, while checklists tend to concentrate on process criteria (e.g. how the ventilation procedure is performed, and position of shoulders and hands of the resuscitator during thorax compression), which are considered to be relevant for outcome. Moreover checklists can be used for scoring of the diagnostic assessment of the victim, which cannot be assessed by the recording manikin.

Only limited research has addressed comparison of checklist and recording strip as evaluation methods for CPR. Two authors reported comparisons between checklist-based scores and mechanical recording-based scores [17,18], and concluded that checklist-based scores overestimated competence.

In this study we evaluated the practical performance of general practitioners in cardiopulmonary resuscitation before and after instruction and compared checklist-based scores to mechanical recording scores to investigate which scoring method is preferable.

2. Materials and methods

Seventy-one general practitioners participated in a continuing medical education course with basic CPR as one of the topics. An account of this course has been published elsewhere [19]. The training time for CPR was 1 h and training was given in small groups (8–12 participants) by two experienced CPR trainers. Participants were randomly divided into two groups, one was evaluated before instruction and one was evaluated after instruction.

A checklist [20] was used for evaluation based on the guidelines of the Dutch Heart Association [21], comparable to the guidelines of the American Heart Association (AHA) [22,23] and the European Resuscitation Council (ERC) [24], except for the sequence used to initiate CPR. This checklist contained 16 items and included criteria for diagnostic assessment of unresponsiveness, circulation and airway in the correct sequence and speed (six items) and correct sequence and performance of CPR procedures (10 items) (see Table 1). Criteria for cardiac compression included correct placement of hands and position during cardiac compression, compression rate (80–100 per min) and ratio for compression and ventilation (15:2). Criteria for ventilation included correct head tilt-chin lift manoeuvre, prevention of air escape during ventilation and observing for chest rise and fall. Scoring of the checklist criteria allowed for marking adequate or inadequate performance. After scoring the separate criteria, raters were also requested to provide a general impression of CPR proficiency on a 10-point rating scale.

The performance of CPR procedures was also assessed by the structured use of the recording strip of a resuscitation manikin (Laerdal Recording ResusciAnne type 20.00.10) as described by Berden et al. [14]. This scoring system includes criteria for placement of the hands, compression rate and depth, compression/relaxation ratio, and breathing volume and interval (see Table 2).

After receiving standardized instruction, participants were rated while performing single-rescuer CPR during 2 min on a ResusciAnne recording manikin. Then feedback on performance was provided by the rater, based on the checklist rating and the recording printout strip. One third of the encounters was double-rated to determine interrater reliability of the checklist-based score and the general impression rating. Raters were general practitioners recruited from the staff of two university departments of general practice and had no specific experience as CPR trainers (the two CPR trainers were not included as raters). Two weeks before the course the raters received 1 h of instruction to practice scoring and discuss interrater differences, the aim being to achieve consensus. The recording strips were scored after the course by the first and second author, and half of the strips were double-rated to determine interrater reliability.

2.1. Data management and statistical analysis

Complete scores on checklist and recording strip were available for 70 participants, as from one candidate no recording strip was available due to malfunctioning of the manikin. A general impression rating was available for 64 participants, and was missing for six participants. Raw scores on the checklist (maximum score 16 points), general impression rating (maximum 10 points) and recording strip (maximum 95 penalty points) were converted into a percentage score, after penalty scores on the recording strip were reversed to bonus scores. The Mann-Whitney two-tailed test was used to compare mean scores before and after training. Pass–not yet passed decisions were based on the perfor-

Table 1
Checklist for CPR (according to the guidelines of the Dutch Heart Association) [20,21]

	Not or incorrectly performed	Correctly performed
Diagnostic procedures		
1. Assessment of unresponsiveness	☐	☐
-Calls loudly on the victim	☐	☐
-Gives strong pain stimulus	☐	☐
2. Assessment of circulation		
-Checks unilaterally for carotid pulse for more than 4 s	☐	☐
3. Assessment of airway		
-Checks if airway is free and assesses breathing	☐	☐
4. Diagnostic procedures are performed in correct sequence	☐	☐
5. Concludes diagnosis within 30 s	☐	☐
Performance of CPR		
6. Starts with chest compressions	☐	☐
7. Proper compression position		
-Correct position of shoulders	☐	☐
-Proper hand placement	☐	☐
8. Adequate compression technique		
-Maintains chest compression rate of 80-100 per min	☐	☐
-Performs cycles of 15 compressions and two ventilations	☐	☐
9. Adequate ventilation technique		
-Performs correct head tilt chin lift manoeuvre	☐	☐
-Mouth fully covers mouth of patient	☐	☐
-Prevents air escaping from nose of patient	☐	☐
-Watches for chest movements during insufflation	☐	☐
10. Chest rises and falls during ventilation	☐	☐
General impression (1–10): ----		

mance of CPR using the standard of the AHA [22] for the checklist (i.e. no errors allowed) and the standard set by Berden [14] for the recording strip, allowing a maximum of 15 penalty points. For the general impression rating a score of 6 or more was considered a pass score. The methods were compared with regard to reliability using intraclass correlation coefficients [25]. Accuracy of observer assessment based on checklist criteria also covered by the recording strip was measured calculating sensitivity and specificity indexes [26], with the recording strip serving as gold standard. Consistency in ranking between the different methods was measured with Spearman's rank correlation coefficient.

3. Results

3.1. Scores

In Table 3 the mean scores (S.D.) are given for the checklist, rating scale and recording print for the group before ($n = 32$) and after ($n = 38$) instruction. Mean scores were lowest for diagnosis and showed no significant improvement after instruction. Scores on performance were higher for the checklist compared with the recording strip, and showed improvement on both scoring methods. However, this difference was not statistically significant for checklist-based ventilation. Finally, total checklist-based score and rating scale showed difference in score before and after instruction. Applying the standard of the AHA to the checklist resulted in 5/32 (15%) participants with adequate CPR performance before instruction and 14/38 (37%) after instruction. For the scoring system based on the recording print the figures rose from 6/32 (18%) before to 18/38 (47%) after instruction. Based on the general impression rating, pass scores were 17/28 (61%) before and 35/36 (97%) after instruction.

3.2. Interrater reliability

The interrater reliability values for the overall scores on the different assessment methods and checklist subscores are shown in Table 4. The reliability of the score based on the recording strip was highest, while the general impression rating showed the lowest reliability. For the checklist, interrater reliability for the total score and diagnosis was much higher than for performance.

3.3. Observer accuracy

It was possible to evaluate accuracy of observers on two criteria, compression rate and ventilation volume, which were covered both by the checklist and recording strip. The recording strip scores were dichotomized according to checklist criteria: for compression rate,

Table 2
Recording strip and scoring system for CPR [14]

Parameter	Value	Penalty points
Placement of hands	Right	0
	Wrong	5
Compression rate (per min)	80–100	0
	100–120	5
	120–140 or 60–80	10
	>140 or 40–60	15
	<40	20
Compression depth (mm)	38–52	0
	30–38 or 52–60	5
	22–30 or >60	10
	<22	20
Compression/relaxation ratio	0.6–1.4	0
	<0.6 or >1.4	10
Breathing volume (l)	0.8–1.2	0
	1.2–1.5	5
	0.4–0.8 or 1.5–2.0	10
	>2.0	15
	<0.4	20
Breathing interval (s)	<4	0
	4–6	5
	6–8	10
	8–10	15
	>10	20

cutoff points were 80 and 100 compressions per min, while for the ventilation volume the cutoff point was 0.8 l, considered equivalent to the minimum volume necessary to make the chest rise [23]. As shown in Table 5, observers judged a higher number of participants as performing adequately for compression rate and ventilation volume, compared with the recording results. The sensitivity and specificity indexes reveal that observers were specific but not very sensitive in identifying poor performance on the two criteria.

3.4. Consistency in ranking

The Spearman rank correlation coefficients between the different scores are shown in Table 6. The correlation between checklist score and general impression was moderate (0.67). The correlation for CPR performance between the checklist score and recording strip score was low (0.45), indicating that the two methods apparently ranked participants quite differently. Also the correlation between the diagnosis score and performance score was low (0.22) for the checklist as well as for the recording strip.

4. Discussion

General practitioners showed considerable deficien-

Table 3
Checklist and recording print scores for CPR: results before and after training[a]

	Checklist (mean (S.D.))		Recording print (mean (S.D.))	
	Before	After	Before	After
Diagnosis	52.6 (27.5)	61.0 (21.7)		
Performance	75.0 (15.5)	87.1 (11.4)[c]	68.1 (11.6)	78.0 (10.6)[c]
Cardiac compression	64.4 (22.0)	86.3 (18.7)[c]	73.3 (13.5)	82.8 (13.8)[b]
Ventilation	83.1 (19.7)	90.0 (11.2)	60.9 (21.7)	72.4 (18.6)[b]
Total score	65.8 (12.2)	78.0 (12.3)[c]		
General impression (rating scale)	60.0 (11.8)	75.4 (10.8)[c]		

[a]Scores are presented as percentage of maximum scores.
[b]Mann-Whitney $P<0.05$ for difference before–after.
[c]Mann-Whitney $P<0.001$ for difference before–after.

Table 4
Interrater reliability for general impression, checklist and recording strip CPR

	Interrater reliability
General impression	0.70
Total score checklist	0.79
Subscore diagnosis (item 1–5)	0.77
Subscore performance (item 7–10)	0.56
Recording strip	0.97

cies in basic CPR skills. This confirms results of earlier studies among different health professionals [9–12]. A 1-h refresher course improved scores but was not enough for all participants to acquire an adequate level of performance according to the scoring system based on the recording strip or criteria of the AHA. However, the general impression of the raters was much more favourable. As raters were general practitioners, they could have been reluctant to judge their peers as performing unsatisfactorily. On the other hand, pass–not yet passed decisions based on the standard format of the AHA or the recording strip may be unnecessarily stringent concerning CPR performance procedures. For early activation of the emergency medical service and rapid initiation of BLS, the effect on outcome is well demonstrated [2,3,27–29], while evidence for the effect on survival of variability in performance of cardiac compression or ventilation is not substantial. Lund [2] demonstrated a negative effect on survival of gross omissions in CPR technique (e.g. performing ventilation without cardiac compression). In other investigations, no relation was found between level of CPR skills and patient outcome [30]. The high standards, as used in this study, perhaps have more significance as an educational goal of excellence and are not necessarily critical for survival. The formative value of CPR assessment, which allows providing of immediate detailed feedback to trainees, should therefore be emphasized rather than its summative value, in order to avoid possible discouraging effects on motivation to perform CPR [31].

The comparison of a checklist-based and a recording strip-based scoring system revealed considerable differences between these methods. The interrater reliability for the checklist was comparable to those reported in the literature for technical clinical skills [32,33], but was lower compared with the recording strip, as recording strip scoring allowed less observer error. Nevertheless, interrater reliability for the diagnostic procedures was very acceptable, indicating that observers agreed strongly about scoring in this part of the checklist. This provides support for the use of a checklist for scoring of the diagnostic procedures.

The interrater reliability was considerably lower for the performance of CPR (cardiac compression and ventilation), indicating that perhaps observation criteria for behaviour during cardiac compression and ventilation were less clear or procedures themselves were more difficult to observe. Moreover, accuracy of checklist scoring for compression rate, using the recording strip as gold standard, was low. Although raters were specific in identifying poor performance, they were not very sensitive, as they tended to overestimate correct performance. Others have reported similar results [17,18]. For ventilation volume, the difference between checklist scoring and recording strip may be a consequence of criteria used, because apparently volumes lower than 0.8 l will also make the chest of the recording manikin rise [35]. Recently, stronger emphasis on observation of chest rise as criterium for adequate ventilation has been recommended [36], so recording strip criteria used in this study were perhaps less valid compared to checklist criteria. Finally, the correlation between checklist score and recording strip score was rather low, indicating that candidates were ranked differently according to their scores in the two methods, as has been reported earlier [17]. These results support the superiority of the recording manikin print as compared with the checklist to evaluate performance of cardiac compression, while the study does not allow conclusions concerning preferable method for ventilation volume.

The correlation between rating scale and checklist was moderate, and higher than that between rating

Table 5
Accuracy of checklist marking for compression rate and ventilation volume[a]

	Compression rate (80–100/min)		Ventilation volume (>0.8 l)	
	Yes	No	Yes	No
Recording strip[a]	32	38	52	18
Checklist[a]	42	28	62	8
Agreement[a]	26	22	51	7
Sensitivity	0.81	0.58	0.98	0.39
Specificity	0.58	0.81	0.39	0.98

[a]Recording strip as gold standard.

Table 6
Correlations[a] between different methods

Methods	General impression	Checklist		
		Total score	Diagnosis (item 1–5)	Performance (item 7–10)
Recording strip total score	0.45	0.45	0.22	0.46
Checklist performance score	0.67	0.71	0.22	
Checklist diagnosis score	0.25	0.79		
Checklist total score	0.61			

[a]Spearman's rank correlation coefficient.

scale and recording strip. This may have been caused by a 'halo effect' on the checklist (i.e. the raters' general impression of performance influenced the scoring of the separate criteria) [34]. The rather low correlation between the checklist and recording strip score for performance of CPR indicates that 'process'-oriented and 'outcome'-oriented assessment ranked resuscitators differently. Therefore, if feasible, a recording strip should be used to evaluate performance of CPR. Within the checklist, correlation between diagnosis and performance score was low, as well as correlation between diagnosis and recording strip, indicating that the score on diagnosis is a poor predictor for the score in performance of CPR and vice versa. This has important implications for assessment of proficiency in CPR, because proficiency in diagnostic procedures should not apparently be taken for granted in individuals who demonstrate proficiency in performance of CPR.

Therefore, we strongly recommend the combination of assessment by raters using a checklist for diagnostic procedures and the recording strip of the manikin for performance of CPR, as employed in most evaluation schemes.

Acknowledgements

The authors thank the SVUH (National Institute for Evaluation of Vocational Training) for permission to use the checklist, Jeroen Pielage for statistical support, and members of the Skillslab of the University of Limburg (head of department: Albert Scherpbier) and the Clinical Training Centre in Nijmegen (head of department: Jaap Metz) for their contribution. Gillian Hull of the General Practitioner Writers Association is thanked for polishing the English. This study was financially supported by a grant from the Dutch Ministry of Health.

References

[1] Gillum RF, Folsom A, Luepker RW et al. Sudden deaths and acute myocardial infarction in a metropolitan area, 1979–1980. N Engl J Med 1983; 316: 1353–1357.
[2] Lund I, Skulberg A. Cardiopulmonary resuscitation by lay people. Lancet 1967; ii: 702–704.
[3] Eisenberg MS, Bergner L, Hallstrom A. Cardiac resuscitation in the community. JAMA 1979; 241: 1905–1907.
[4] Hodgkin K. Towards earlier diagnosis in primary care. Edinburgh: Churchill Livingstone, 1978: 252.
[5] Lamberts H, Brouwer H, Mohrs J. Reason for encounter-, episode- and process-oriented standard output from the Transition Project. Amsterdam: University of Amsterdam, 1991.
[6] Pai GR, Haites NE, Rawles LM. One thousand heart attacks in Grampian: the place of cardioplumonary resuscitation in general practice. Br Med J 1987; 294: 352–354.
[7] American Heart Association, Emergency Cardiac Care Committee and Subcommittees. Guidelines for cardiopulmonary resuscitation and emergency cardiac care. I: Introduction. JAMA 1992; 268: 2172–2183.
[8] Dogger CA, Teijink JAW, den Blanken MM, Patka P, Haarman HJT. De rol van de huisarts in de preklinische spoedeisende hulpverlening (The role of the general practitioner in prehospital emergency care). Med Contact 1992; 47: 205–207.
[9] Stross JK. Maintaining competency in advanced cardiac life support skills. JAMA 1983; 249: 3339–3341.
[10] Kaye W, Mancini ME. Retention of cardiopulmonary resuscitation skills by physicians, registered nurses, and the general public. Crit Care Med 1986; 14: 620–622.
[11] Berden HJJM, Willems FF, Ten Have FTM et al. De primaire reanimatievaardigheid van de huisarts (Basic life support skills of general practitioners) Ned Tijdschr Geneeskd 1988; 132: 1797–1801.
[12] Seraj MA, Naguib M. Cardiopulmonary resuscitation skills of medical professionals. Resuscitation 1990; 20: 31–39.
[13] Lowenstein SR, Hansborough JF, Libby LS et al. Cardiopulmonary resuscitation by medical and surgical house-officers. Lancet 1981; ii: 679–681.
[14] Berden HJJM, Pijls NHJ, Willems FF et al. A scoring system for basic cardiac life support skills in training situations. Resuscitation 1992; 23: 21–31.
[15] Ramirez AG, Weaver FJ, Raizner AE et al. The efficacy of lay CPR instruction: an evaluation. Am J Publ Health 1977; 67: 1093–1095.
[16] Fossel M, Kiskaddon RT, Sternbach GL. Retention of cardiopulmonary resuscitation skills by medical students. J Med Educ 1983; 58: 568–575.
[17] Kaye W, Mancini ME, Rallis SF et al. Can better basic and advanced cardiac life support improve outcome from cardiac arrest? Crit Care Med 1985; 13: 916–920.
[18] Van Kalmthout PM, Speth PAJ, Rutten JR et al. Evaluation of lay skills in cardiopulmonary resuscitation. Br Heart J 1985; 53: 562–566.

[19] Jansen JJM, Scherpbier AJJA, Metz JCM et al. Construct validity of performance-based assessment in continuing medical education for general practitioners. Med Educ 1996; 30: in press.
[20] Werkgroep Vaardigheden SVUH. Reanimatie zonder hulp(middelen) (Single-rescuer resuscitation). Utrecht: SVUH, 1993.
[21] Van Drenth J. Wanneer elke seconde telt (When every second counts). Den Haag: Nederlandse Hartstichting, 1989.
[22] Instructor's Manual for Basic Life Support. Dallas, TX: American Heart Association, 1985.
[23] American Heart Association, Emergency Cardiac Care Committee and Subcommittees. Guidelines for cardiopulmonary resuscitation and emergency cardiac care. II: Adult basic life support. JAMA 1992; 268: 2184–2198.
[24] Basic Life Support Working Party of the European Reuscitation Council. Guidelines for basic life support. Resuscitation 1992; 24: 103–110.
[25] Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. Clin Pharmacol Ther 1981; 29: 111–123.
[26] Inglefinger JA, Mosteller F, Thibodeau LA et al. Biostatistics in clinical medicine. New York: Macmillan, 1983: 4–10.
[27] Copley DP, Mantle JA, Rogers WJ et al. Improved outcome for prehospital cardiopulmonary collapse with resuscitation by bystanders. Circulation 1977; 56: 901–905.
[28] Cummins RO, Eisenberg MS, Hallstrom AP et al. Survival of out-of-hospital cardiac arrest with early initiation of cardiopulmonary resuscitation. Am J Emerg Med 1985; 3: 114–119.
[29] Hallstrom AP, Cobb LA, Swain M et al. Predictors of hospital mortality after out-of-hospital cardiopulmonary resuscitation. Crit Care Med 1985; 13: 927–929.
[30] Tweed WA, Wilson E. Is CPR on the right track? Can Med Assoc J 1984; 131: 429–433.
[31] Sigsbee M, Geden EA. Effects of anxiety on family members of patients with cardiac disease learning cardiopulmonary resuscitation. Heart Lung 1990; 19: 662–665.
[32] Van der Vleuten CPM and Swanson DB. Assessment of clinical skills with standardized patients; state of the art. Teach Learn Med 1990; 2: 58–76.
[33] Jansen JJM, Tan LHC, Van der Vleuten CPM et al. Assessment of competence in technical clinical skills of general practitioners. Med Educ 1995; 29: 247–253.
[34] Streiner DL. Global rating scales. In: Neufeld VR, Norman GR, editors. Assessing clinical competence. New York: Springer, 1985: 119–141.
[35] Laerdal Company; personal communication.
[36] van Vliet J. Consensus basale reanimatie van de Nederlandse Reanimatie Raad (Consensus basic life support of the Dutch Resuscitation Council) Ned Tijdschr Geneeskd 1996; 140: 596–599.