2008

A New Paradigm for Human Resuscitation Research Using Intelligent Devices

Charles F. Babbs
Purdue University, babbs@purdue.edu

Andre E. Kemeny

Weilun Quan

Gary Freeman

A New Paradigm for Human Resuscitation Research Using Intelligent Devices

Charles F. Babbs [a,b,c] Andre E. Kemeny[c], Weilun Quan[d], Gary Freeman[d]

[a] Department of Basic Medical Sciences, Purdue University; 1246 Lynn Hall, West Lafayette, IN 47907-1246, USA

[b] Indiana University School of Medicine, Indianapolis, IN, USA

[c] Weldon School of Biomedical Engineering, Purdue University; 206 S. Intramural Drive West Lafayette, IN 47907-2032, USA

[d] Zoll Medical Corporation, Chelmsford MA, USA

(Resuscitation Volume 77, Issue 3, June 2008, Pages 306–315)

Abstract

Objectives: To develop new methods for studying correlations between the performance and outcome of resuscitation efforts in real-world clinical settings using data recorded by automatic devices such as automatic external defibrillators (AEDs), and to explore effects of shock timing and chest compression depth in the field.

Methods: In 695 records of AED use in the pre-hospital setting, continuous compression data were recorded using AEDs capable of measuring sternal motion during compressions, together with timing of delivered shocks and the electrocardiogram. In patients who received at least one shock, putative return of spontaneous circulation (P-ROSC) was defined as a regular, narrow complex electrical rhythm > 40 beats/min with no evidence of chest compressions at the end of the recorded data stream. Transient return of spontaneous circulation (t-ROSC) was defined as the presence of a post-shock organized rhythm > 40 beats/min within 60 seconds, and sustained ≥ 30 seconds. 2x2 contingency tables were constructed to examine the association between these outcomes and dichotomized time of shock delivery or chest compression depth, using the Mood median test for statistical significance.

Results: The probability of P-ROSC for first shocks delivered < 50 seconds (the median time) after the start of resuscitation was 23%, versus 11% for first shocks > 50 seconds ($p=0.028$, one tailed). Similarly, the probability of t-ROSC for shorter times to shock was 29%, compared to the 15% for delayed first shocks ($p=0.016$). For shocks occurring >3 minutes after initiation of rescue attempts, the probability of t-ROSC with pre-shock average compression depth > 5 cm was more than double that with compression depth < 5 cm (17.7% vs. 8.3%, $p=0.028$). For shocks >5 minutes the effect of deeper compressions increased (23.4% vs. 8.2%, $p=0.008$).

Conclusions: Much can be learned from analysis of performance data automatically recorded by modern resuscitation devices. Use of the Mood median test of association proved to be sensitive, valid, distribution independent, noise-resistant, and also resistant to biases introduced by the inclusion of hopeless cases. Efforts to shorten the time to delivery of the first shock and to encourage deeper chest compressions after the first shock are likely to improve resuscitation success. Such refinements can be effective even after an unknown period of preceding downtime.

Key words: Basic life support (BLS); Defibrillation; External chest compression; Guidelines; Methodology; Ventricular fibrillation

1. Introduction

Modern resuscitation devices, including automatic external defibrillators and chest or abdominal compression devices record a wealth of digital data during their use, including time of day, the electrocardiogram (ECG), device actions, and sometimes voice and chest acceleration. We propose that these data can be analyzed to produce a rather detailed picture of events during the resuscitation, including device performance and patient outcome. Analysis of the association between performance and outcome can provide important clues—at little marginal cost—to guide the improvement of resuscitation techniques and to advance the science of resuscitation. In particular, the association between successful outcomes and the timing of defibrillation shocks or the depth of chest compression can be efficiently studied in this way, using real-world, field data, despite a lack of experimental protocols, strict inclusion criteria, or randomization of subjects.

2. Methods

2.1 Sample records

We analyzed 695 digital records of the pre-hospital use of AEDs recorded by ZOLL AED Pro and ZOLL AED Plus devices and obtained from Zoll Medical Corporation in early 2007. These devices were deployed in various locations and clinical settings. The authors are blind to locations where data were collected and to identifiable health information of all the patients used in this study, including in particular any demographic characteristics of the patients. The electronic records included compression pad acceleration, digitized at 100 and 125 Hz, and the ECG[*], digitized at 250 Hz, which were recorded continuously after device activation. A text file for each patient record included calibration readings for acceleration, notations of device actions and device decisions such as "shock advised", "shock delivered", as well as the times of all shocks.

To exclude cases when the AED was used only for monitoring, our analysis was limited to patients who received at least one shock from the AED. Electronic records included codes summarizing outputs of automated ECG analysis algorithms "normal sinus rhythm like", "PEA-like", "probable VT", "coarse VF", "fine VF", "abnormal but not shockable". We used these codes to characterize initial rhythms of the patients in our data set without revealing identifiable health information. Rhythms that triggered shocks included VT, coarse VF, VF, and fine VF.

[*] Abbreviations: AED automatic external defibrillator, CPR cardiopulmonary resuscitation, ECG electrocardiogram, EMS emergency medical systems, PEA pulseless electrical activity, ROSC return of spontaneous circulation, VF ventricular fibrillation, VT ventricular tachycardia

2.2 Measures of outcome

Using ECG and chest acceleration data for patients who received at least one shock, we labeled as putative survivors cases with a regular, narrow complex electrical rhythm at > 40 beats/min with no evidence of chest compressions at the end of the recorded data stream. This pattern of events implies that rescuers on the scene were convinced that signs of circulation were present and acted rationally. (Otherwise, they would have continued chest compressions.) The rate criterion tends to exclude slower agonal rhythms such as pulseless electrical activity at the end of a failed resuscitation attempt. We call this result P-ROSC, or putative return of circulation, indicating probable overall success. On a shorter time scale we labeled the electrical conversion of shockable rhythms to perfusing rhythms as t-ROSC (transient return of spontaneous circulation) in a subset of 251 patients. t-ROSC was defined as the presence of post-shock organized rhythm > 40 beats/min within 60 seconds, and sustained for at least 30 seconds. These categories of outcome can be established from the digital data without knowledge of patient identity or access to traditional medical records.

2.3 Measures of chest compression

We determined chest compression depth from double integration of high pass filtered acceleration and velocity data, followed by peak detection and extraction of the peak height. This sequence provided the depth of each valid chest compression. Full details of the methodology are provided in Appendix 1, including calibration, numerical integration, peak detection, and artifact removal.

2.4. Statistical methods

Pre-hospital field data often include records of many patients who have absolutely no chance of survival. Thus it is necessary to develop statistical methods that can gracefully handle the dilution of the database with such hopeless cases. In addition, the distributions of continuous variables that we studied are typically highly skewed, containing a substantial number of extreme values in long "tails". Figure 1 illustrates a distribution independent statistical approach that is compatible with skewed sample distributions and yet provides a sensitive measure of association between performance variables and outcome variables.

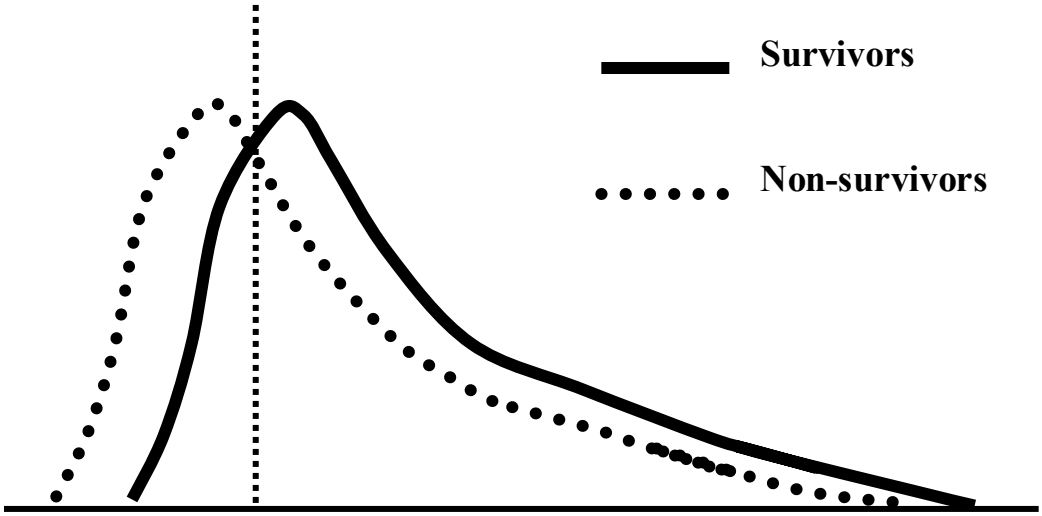

Figure 1. Mood median test of closely spaced distributions. Sketches of histograms showing statistical distributions of survivors and non-survivors as a function of performance are shown. Dotted vertical line represents the joint median for the two distributions. Survivors and non-survivors are counted as falling below or above the joint median. The proportion of survivors found below the median is one measure of the effect of the performance measure on outcome.

The figure shows distributions of survivors and non-survivors as a function of some measure of CPR performance, such as chest compression depth. The dotted vertical line represents joint median for the two distributions, or some other suitable cutoff used to classify performance into two categories indicating bad vs. good technique. The numbers of survivors and non-survivors falling below or above the cutoff are counted. Each patient then falls into one of the corresponding cells in a 2x2 contingency table, such as that reproduced in Table 1.

For example, group A could be all patients with less than 4 cm chest compression depth and group B could be all other patients. Performance measures on any scale and in any units can be dichotomized in this way. Also, inherently dichotomous variables such as urban vs. rural, or shock first vs. CPR first can be used to define test groups A and B. In this case the outcomes (survival vs. non-survival, or success vs. failure) are inherently dichotomous. The proportion of survivors found above versus below the cutoff is a measure of the influence of performance on survival.

Table 1. A generalized 2x2 contingency table. Variables a—d indicate counts of patients in the 4 possible classes. Variables m, n, r, and s, are marginal totals, and N is the grand total.

	Group A	Group B	Total
Non-survivors	a	c	m
Survivors	b	d	n
Total	r	s	N
Proportion of survivors	b/r	d/s	n/N

If the marginal totals, m, n, r, and s are known, it is only necessary to specify one of the counts a, b, c, or d to define the distribution of patients in the cells, since the other three may then be calculated from the marginal totals. Only one of these counts is used as the statistic for testing the null hypothesis. Conventionally, it is b.

Under the null hypothesis that the true probability of survival is the same (namely, n/N) for both group A and group B. The exact distribution of b is specified by a hypergeometirc distribution. For totals n and r greater than about 15, a close approximation using the normal distribution can be made. The approximation also makes it easier to visualize the distribution of b and to calculate p-values for hypothesis testing. In this approximation b is distributed as a normal distribution with mean $\mu(b) = nr/N$ and variance $\sigma^2(b) = nmrs/(N^2(N-1))$ [1]. In turn, the test statistic

$$Z = \frac{b - \mu(b)}{\sigma(b)}$$

is distributed as the standard normal distribution with a mean of zero and standard deviation of 1. If we adopt the convention that b represents the number of putative survivors who received "bad" CPR, then we expect $b < \mu(b)$ and $Z < 0$, if CPR performance is related to survival. Then, if the probability density function for the standard normal distribution is f(Z), the "one-tailed" probability of obtaining a result this negative or less under the null hypothesis, H_0, is simply

$$P = \int_{-\infty}^{Z} f(x)dx ,$$

the cumulative normal distribution, which is readily available in ubiquitous standard functions and tables. When the joint median performance is used as the cutoff for defining groups A and B, this test is known specifically as the Mood median test[1].

Analysis of hypothetical data in Table 2 shows how the test works.

Table 2. Hypothetical pre-hospital resuscitation data.

	Group A (< median CPR)	Group B (> median CPR)	Total
Non-survivors	30	25	55
Survivors	20	25	45
Total	50	50	100
Proportion of survivors	0.4	0.5	0.45

Here the proportions of survivors in sample groups A and B are different. We have b = 20, $\mu(b)$ under $H_0 = 0.45x50/100 = 22.5$, variance $\sigma^2(b) = 50^2$ x55x45/(100^2(99)) = 6.25, $\sigma(b) = 2.5$, Z = (20 – 22.5) / 2.5 = -1.0, and P = 0.16.

2.5 Dealing with hopeless cases and rescuer bias

There are two major biases in retrospective studies of pre-hospital resuscitation. The first, as emphasized by Gordon Ewy, is that most patients studied have absolutely no chance of surviving[2]. The second bias is the tendency of experienced rescuers to recognize such non-viable patients and apply less vigorous or even token resuscitation attempts. The first bias tends to dilute the success of good technique, decreasing the observed association between good technique and survival. The second bias tends to shift non-viable patients into the poor technique category, increasing the observed association between bad technique and non-survival. As is easily shown by trial and error, the Mood median test is relatively insensitive to dilution of the population with hopelessly dead patients, and, as shown in Appendix 2, the effects of hopeless cases and of rescuer bias tend to offset mathematically in the calculation of a p-value for the observed association. Accordingly, we used this approach to test the association or correlation between various CPR performance measures and P-ROSC or t-ROSC. We

suggest that it is especially appropriate for extracting "signal" from "noise" in the present application.

3. Results

3.1 Time to first shock

There was a significant relationship between time to first shock and P-ROSC. In patients who received at least one shock the observed probability of survival when first shocks were delivered in less than the median time of 50 seconds was 23 percent, whereas the observed probability of survival when shocks were delivered after 50 seconds was only 11 percent (Figure 2(A), one-tailed $p = 0.028$). If the definition of P-ROSC was modified to include patients with regular, narrow complex rhythms at a rate > 50 beats/min (rather than > 40 beats/min) with no evidence of chest compressions at the end of the AED record, then the probability of survival when first shocks were delivered in less than the median time of 50 seconds was 23 percent, compared to 8 percent for shocks delivered after 50 seconds (one-tailed $p = 0.007$). The latter definition excluded two patients from the list of survivors who might have had pulseless electrical activity while ECG recording continued after chest compressions were abandoned.

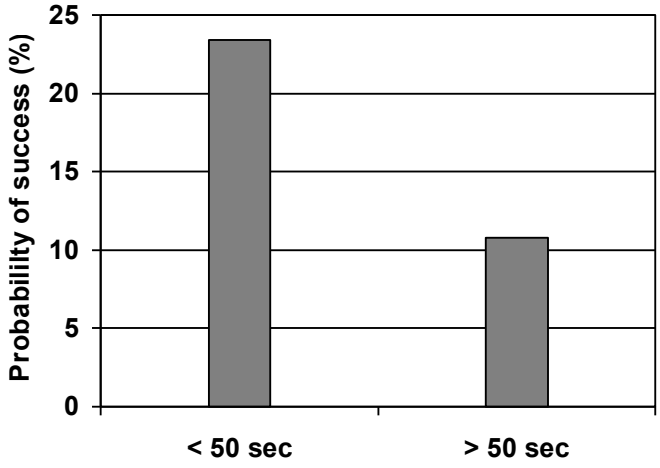

Figure 2(A). Association between time to first shock and P-ROSC using a 2x2 contingency table analysis and the Mood median test of the null hypothesis, for which p = 0.028. The rate criterion for a viable final rhythm was > 40 beats/min.

One possible interpretation of the results in Figure 2(A) is that early shocks are actually a marker for faster overall resuscitation efforts, perhaps performed by more experienced teams. We know that in prolonged resuscitations the probability of survival diminishes with time. It could be that more efficient and aggressive rescuers delivered shocks early, and so had more success. To examine this possibility we selected data for ventricular

fibrillation (VF) arrests lasting 5 minutes (300 sec) or longer. For these prolonged arrests the results are shown in Figure 2(B).

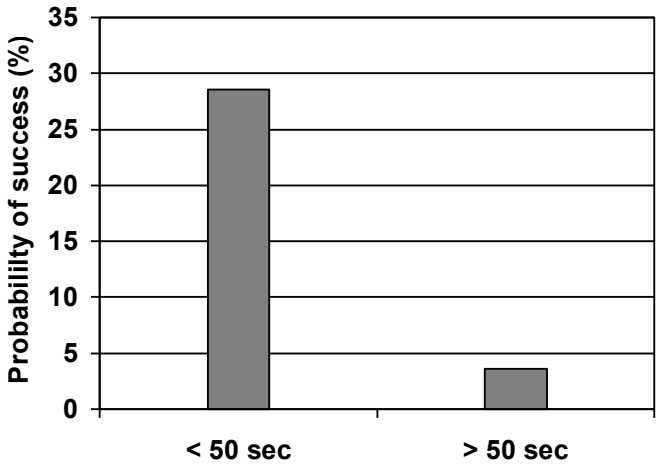

Figure 2(B). Association between time to first shock and P-ROSC for VF cardiac arrests lasting at least 5 minutes from device application. Here N = 56. Seven of 8 survivors received first shocks in < 50 sec. For the 2x2 table p = 0.006.

In this subgroup the arrests lasted at least 5 min after pad application, neglecting any preceding downtime. The median time to first shock was still 50 seconds in this population of prolonged VF arrests. However, the first shock did not result in prompt ROSC in these cases, since compressions were delivered intermittently for at least 250 additional seconds. Nevertheless, early first shocks were associated with better P-ROSC outcome in this subgroup. One reasonable interpretation is that early removal of VF itself is of benefit, even if the resulting rhythm is not a spontaneously perfusing rhythm, and additional CPR is required. The underlying physiology may be that VF causes increased myocardial oxygen demand, which is inimical to resuscitation success.

The importance of early shocks is supported as well in an analysis of successful vs. unsuccessful first shocks using the t-ROSC criterion, which does not require evidence of a perfusing rhythm at the end of the resuscitation, and so yields more cases. For these shock-based data, we noted the number of pre-shock compressions. If the number of pre-shock compressions was less than 5, we labeled the case as "shock first", i.e. no pre-shock CPR. If the number of pre-shock compressions was 5 or greater, we labeled the case as "CPR first", to compare the effectiveness of the shock first vs. CPR first paradigms. Forty of 55 survivors had a shock-first protocol. The observed probability of survival for the shock-first paradigm was 0.265 vs. 0.150 for the CPR-first paradigm. This difference was highly significant statistically, with $p = 0.016$ (Figure 3). These data show that early shocks are highly predictive of success in patients with VF/VT.

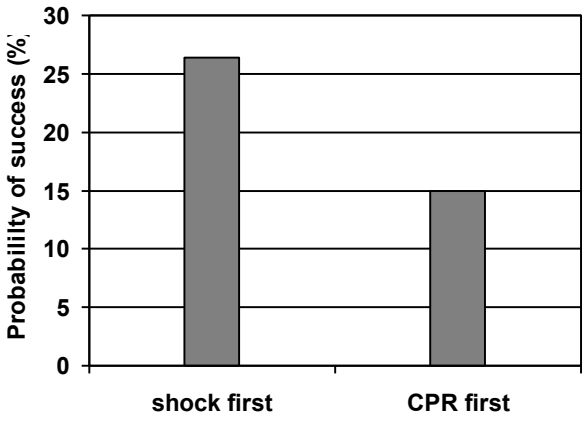

Figure 3. Association between first shock delivery paradigm and t-ROSC using a 2x2 contingency table analysis, for which p = 0.016. There were 151 shock first patients and 100 CPR first patients by this definition.

3.2. Chest compression depth

In the total data set, including both long and short duration resuscitations, there was no statistically significant association between positive outcome (P-ROSC or t-ROSC) and chest compression depth. However, since many survivors in a shock first protocol would have had few chest compressions, we wondered if CPR is important only in longer resuscitations when many compressions are given. To test this hypothesis we examined the t-ROSC data set for patients who received more than one shock. These are patients who failed to respond to the first shock, and so required at least some CPR. The results are shown in Table 3.

Table 3. t-ROSC success vs. failure and average pre-shock chest compression depth* in multiple shock resuscitations.

	< 4.78 cm	>=4.78 cm	Total
Failure (no t-ROSC)	93	84	177
Success (t-ROSC)	8	17	25
Total	101	101	202
Proportion of success	0.079	0.168	0.124

* The average value for the compression set immediately preceding the shock

The one-tailed p-value for Table 3 is 0.028, indicating a statistically significant association between greater pre-shock chest compression depth and survival in those patients who required multiple shocks. Similar results are obtained for the subset of t-ROSC data for shocks delivered 3 or more minutes after pad application.

By either criterion, compression depth seems to matter for prolonged resuscitations, which upon reflection, is quite reasonable. It should take time for differences in blood flow to have an effect. To test this hypothesis further we examined t-ROSC data for even longer resuscitations, in particular the set for shock times of 5 minutes or greater. In this subset the proportion of survivors for mean compression depth less than 5 cm was 8 percent, and the proportion of survivors for mean compression depth greater than 5 cm was 24 percent. The p-value was 0.008 (Figure 4). These results are consistent with the hypothesis that practical differences in CPR technique can influence myocardial perfusion, which in turn influences myocardial metabolism.

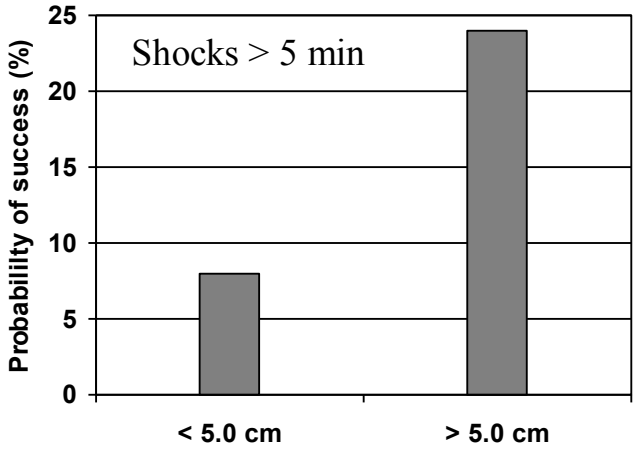

Figure 4. Association between pre-shock compression depth and t-ROSC for shocks delivered 5 min or more after pad application. Here for the Mood median test p = 0.008. N is 126.

Thus we can augment the story of successful shocks as follows. Early shock with a shock-first paradigm produces superior results. However, if the first shock fails, then more vigorous CPR with deeper pre-shock chest compressions sets the stage for greater success.

4. Discussion

Here we describe a method for retrospective analysis of data recorded by resuscitation devices. The method works for all types of continuous data, including times to shock, computed chest compression depth, chest compression frequency, etc. Statistical inference using the Mood median test is valid for skewed distributions characteristic of these data. It can also be applied to variables that are intrinsically dichotomous, such as shock-first vs. CPR-first. The results are not adversely influenced by the admixture of large numbers of hopeless cases (long-down time arrests) and a realistic degree of rescuer bias, as shown in Appendix 2. Although the conclusions derived from such retrospective analysis are more tentative than those derived from prospective, randomized clinical trials, the method is far less costly to implement and has a real-world validity that exceeds in some ways that of clinical trials at a few exceptional centers. Often there is evidence in the data set that can confirm or deny alternative hypotheses. As previously pointed out by Sunde and coworkers[3], a wide variety of issues in resuscitation can be studied retrospectively using such methods.

Here we focus on two important variables for which our own preliminary data are interesting. These are the timing of shock delivery and the depth of chest compression in conventional CPR for cases of VF/VT. These issues of shock timing[4, 5] and shock-first

vs. CPR-first protocols are debated within the resuscitation research community[6, 7]. Based upon several hundred resuscitations using ZOLL AED and AED-Plus devices, the strongest and most significant predictor of P-ROSC and t-ROSC success for shockable rhythms was time to first shock. A biologically meaningful interpretation of this correlation is that early shocks remove VF, permit perfusing rhythms, and improve the balance of myocardial oxygen supply and demand, making the heart easier to resuscitate.

Chest compression depth is a major but still underappreciated variable governing blood flow during CPR[8-10], and the trade-offs between the risk of rib fracture, liver laceration, or other injury vs. the benefits of enhanced blood flow remain unresolved[11]. Chest compression depth in the field or hospital tends to be too shallow compared to guideline recommendations[12-14] and to decay rapidly over time owing to rescuer fatigue[15]. Interestingly, Edelson and coworkers[16] found that successful removal of ventricular fibrillation adult out-of-hospital cardiac arrests was associated with higher mean compression depth during the 30 seconds of CPR preceding the shock. We found no convincing evidence in these early data that successful outcomes were related to the mechanics of CPR when all patients in either the P-ROSC or t-ROSC data sets were analyzed together. However, when we examined patients in the t-ROSC data set that received more than one shock, the probability of survival for those receiving deeper pre-shock compressions (> 4.8 cm) was 17 percent, compared to 8 percent for those receiving shallower pre-shock compressions. Similar results were obtained for the subset of shocks delivered 5 or more minutes after pad application.

This finding, upon reflection, is quite reasonable. It is consistent with Becker's hypothesis that that the pathophysiology of ventricular fibrillation cardiac arrest consists of three time-sensitive phases: electrical, circulatory, and metabolic, each requiring a different therapeutic emphasis for best results[17]. According to this concept cardiac arrest due to ventricular fibrillation (VF) has three successive phases in time, named electrical, circulatory, and metabolic. During the first phase the problem is mostly electrical. The myocardium is viable and capable of beating if the electrical abnormality is removed. The treatment of choice is prompt defibrillation. During the second phase the ability of the heart to recover is limited also by lack of oxygen. The treatment of choice is artificial circulation and then shock. During the third phase the metabolism of cardiomyocytes is further deranged (e.g. by mitochondrial dysfunction), such that aerobic metabolism does not resume immediately even when oxygen is restored. The treatments of choice include drug cocktails or hypothermia to heal broken metabolic pathways.

Figure 5 illustrates the relation of our findings to this conceptual model.

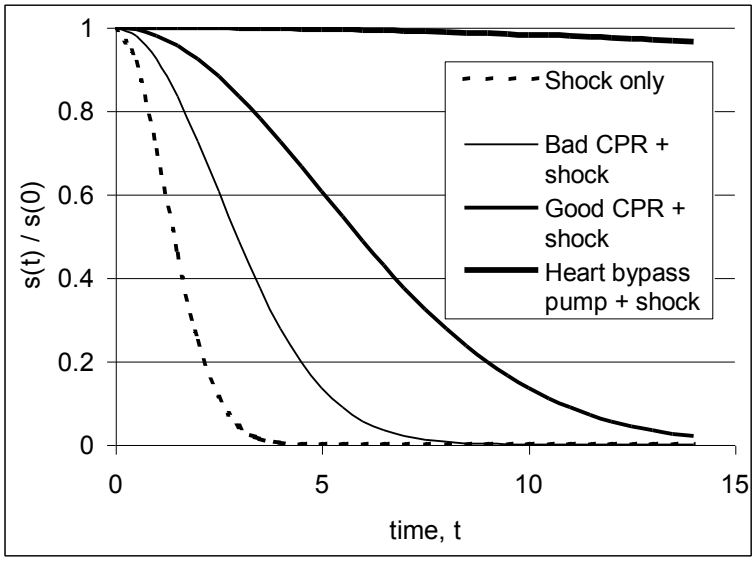

Figure 5. Hypothetical decay of the effectiveness of various therapies for VF as a function of time after onset of cardiac arrest.

Let s(t) indicate the number of saves or successes as a function of time, t, after onset of VF. The relative effectiveness of various treatments as a function of time is indicated by the ratio s(t)/s(0). Success with all treatments decays as time in VF increases. The fastest decay is for shock alone. The slowest decay is for cardiopulmonary bypass during open heart surgery, during which myocardial perfusion is nearly normal. These extremes bracket a family of curves representing increasingly normal perfusion. The hypothesis is that progressively better blood flow during VF improves the chances of successful electrical ventricular defibrillation with ROSC. As blood flow during VF improves, the percent success curves are pushed to the right. At times > 3 min one would observe a greater difference between poor blood flow and good blood flow during CPR.

It makes sense that during the earlier electrical phase the best strategy is to shock first. Better CPR mechanics would be expected to have an effect during the circulatory phase, which occurs later. Gilmore's recent study of witnessed cardiac arrests[17], suggests that the electrical phase lasts for up to 5 minutes and the circulatory phase lasts from 5 to 15 minutes. Our AED data set, however, includes unwitnessed arrests with an unknown amount of downtime. Time zero for our clock is the time of pad application, which occurs after time zero for witnessed arrests. We found evidence for a transition from the electrical to the circulatory phase at 3 to 5 minutes after pad application. There was a cluster of successful shocks at 8 to 11 minutes in the t-ROSC data set and only one successful shock after 15 minutes for any quality of CPR. Hence the end of the

circulatory phase on our clock might occur at somewhat less than 15 minutes. This is because for unwitnessed arrests, the AED clock starts sometime after the actual onset of VF.

The curves of Figure 5 also imply that the ideal strategy remains a combination of the best possible artificial circulation from the onset of VF and the earliest possible shock. Neglecting for the moment a "blind shock" strategy, there needs to be an electrical diagnosis of VF, and current technology requires a pause in chest compressions for human or robotic monitoring of the ECG, recognition of VF, and charging of an automatic defibrillator after VF detection. During this pause chest compression amplitude is zero, which is quite different from the best possible perfusion. However, new technology is being developed that allows electrical diagnosis of VF and charging of an automatic defibrillator during chest compressions, as well as continuation of CPR if the detected rhythm is non-shockable (Q. Tan, G. Freeman, and F. Geheb, Zoll Medical Corp., personal communication). This technology uses adaptive filtering to predict the CPR artifact in the electrocardiogram, based upon the sensed acceleration signal, and to remove it from the ECG channel. In the future it may be possible to deliver both faster shocks and stronger chest compressions for VF arrests using smarter, more capable automatic external defibrillators.

5. Limitations

Our data come exclusively from patients who were treated with at least one shock from automatic external defibrillators in the field. If one is interested in the population of all patients in cardiac arrest, the limitation of our study to those who were treated with an AED may be undesirable. However, for patients with shockable rhythms we can still offer the tentative conclusions that (1) early shocks are preferable to later shocks, and (2) if these shocks fail to establish ROSC, adequate chest compression depth is an important predictor of success. Further, the population of patients shocked by AEDs is itself an important subpopulation because the chances of success are higher than for non-VF arrest, and so effort to improve CPR in this population may especially fruitful.

Our data set does not include important factors known to influence the outcome from cardiac arrest such as downtime, EMS response time, or bystander CPR. Because the major goal of the methodology we propose is to compare effects upon outcome of different resuscitation techniques, unspecified downtime would only be a problem if downtime were correlated with resuscitation technique. Otherwise, downtime is just another uncontrolled random variable that is part of the statistical noise in the problem. Our conclusions could be erroneous, however, if short downtime were systematically correlated with an observed variable, such as time to first shock, that seemed to predict better survival. In this case differences in downtime, rather than shock time, might actually cause the observed differences in survival. This limitation is characteristic of all retrospective, observational studies. If factors A and B are correlated, then A could cause B, perhaps B could cause A, or perhaps a third factor C could cause both A and B. In this case one needs to offer and evaluate alternative hypotheses.

We observed that early shocks (A) are correlated with better survival (B). We offer the hypothesis that removal of VF reduces myocardial oxygen consumption, permits a perfusing rhythm, and promotes resuscitation (A causes B). But suppose, for example, that patients with long downtime got CPR first rather than shock first, because they looked to EMS personnel like they needed CPR. Then the correlation of success with a shock-first strategy could be actually have happened because long downtime (C) prompted rescuers to delay shock (A) and also caused worse survival (B). In a retrospective study this is a legitimate possibility.

However, one can still deal with this possibility using a rich observational database and the methodology that we propose. For example, if patients who got CPR-first were the long downtime patients and therefore hopeless cases, then when one removes the CPR-first cases, the correlation should disappear. Thus one can ask whether among apparently viable patients treated only with a shock-first protocol, whether improved success is associated with shorter time to first shock. If so, then the myocardial metabolism hypothesis is supported. If not, then the downtime hypothesis is supported.

We did this additional analysis within a Microsoft Excel spreadsheet in about 15 minutes, and the results are as follows. In the t-ROSC data set for patients with no pre-shock compressions (shock first patients) the median time to first shock was 49 seconds. The proportion of successful first shocks with less than median time to first shock was 37 percent. The proportion of successful first shocks with greater than median time to first shock was 18 percent. The p-value for the Mood median test is 0.005. Hence the alternative hypothesis is not supported. It is not always necessary to propose exceedingly difficult randomized clinical trials to dispose of alternative hypotheses. With a sufficiently rich database, one can address them by asking additional questions of the data.

Of course, the forgoing strategy will fail if there is a hidden alternative explanation that is not evident to the investigators. However, we do address perhaps the most likely cause of spurious correlations, namely rescuer bias, in which rescuers are able to discern the look and feel of a potential survivor and so perform better than average quality CPR. Such victims, for example, might have a higher probability of getting rapid shocks or stronger chest compressions. However, as shown in Appendix 2, the Mood median test is relatively insensitive mathematically to moderate amounts of rescuer bias.

Future studies can be done to validate the assumptions leading to the definitions of P-ROSC and t-ROSC and the overall validity of this approach. Accumulation of larger numbers of cases in the electronic database from multiple centers also increases the validity of the approach, since a very large number of cases would tend to approximate the "real world", as long as a large subgroup of unidentified survivors were not systematically excluded from the database.

6. Conclusions

The research methodology we, and formerly Sunde and coworkers[3], have proposed and demonstrated lacks many expected features of modern clinical research. There is no informed consent. The identities of the patients are not known, and so their demographic characteristic and clinical diagnoses cannot be listed in a table. The studies are retrospective and observational, rather than prospective and randomized. It will be left to the others to debate the merits of this approach. Perhaps authorities will conclude that in this paradigm there is no real need for informed consent, that there is little chance of deliberate selection bias by investigators, and that as the electronic database becomes larger and larger, the retrospective nature of the analysis and lack of randomization become less important. One might regard the growing volume of electronic data from resuscitation devices as a registry of information that can be interrogated for the purpose of quality control. Trends discovered in the registry can suggest working hypotheses that can be further tested prospectively in randomized clinical trials or future observational studies in different geographic regions.

The process of data mining in this registry is rather like interpreting the data gathered by a remote spacecraft. Indeed, one can think of an automated, intelligent resuscitation device as being like a spacecraft that collects meaningful scientific and technical data in a hostile, chaotic environment. With such technology resuscitation research can be done as an observational science like astronomy or geology or ecology, as compared to a laboratory science like chemistry or molecular biology. Observational sciences test hypotheses by developing new instruments, exploring new spectral regions, and making insightful new observations. They rely on natural experiments rather than controlled laboratory experiments. In space exploration expensive manned missions remain a gold standard and ultimate goal, and yet much can be learned from robotic craft equipped with multiple sensors. In resuscitation research expensive multi-center, randomized clinical trials remain a gold standard and ultimate source of truth, and yet much can be learned from robotic devices deployed in the field and equipped with multiple sensors.

Appendix 1: Data processing to obtain compression depth from acceleration

Interpretation of acceleration data

Accelerometers work by measuring the bending of a weighted cantilever beam inside a box, which acts as a movable frame of reference. When the box is at rest on the Earth's surface the beam is bent downward, registering +1g. If the box is turned over, the beam is bent in the opposite direction, registering –1g. These calibration values are provided with each ZOLL AED device. When a rescuer presses on the sternum through the AED compression pad, the box moves. The change in acceleration felt by the cantilever beam is similar to that felt by a man in a moving elevator. The baseline signal from the device before chest compression is +1g. During a downward push on the box the apparent gravity inside the box first becomes less, typically by about 1g, as our records show. Then, as the speed of downward compression slows and reverses, the box becomes a rising elevator. The weight at the end of the cantilever becomes heavier, and the signal from the accelerometer is greater than the +1g calibration value.

Figure A1.1 shows the theoretical patterns of displacement, velocity, and acceleration expected for a single chest compression of 5 cm depth. Here a sinusoidal pattern, ranging from 0 to , y_{max} = +5 cm is assumed. Downward compression, y, velocity toward the floor, y', and acceleration, y'', are given as functions of frequency, f, and time, t, by the expressions

$$y = \frac{y_{max}}{2}\left(1 - \cos(2\pi f t)\right) \qquad (1a)$$

$$y' = y_{max}\,\pi f \sin(2\pi f t) \qquad (1b)$$

and

$$y'' = 2\,y_{max}\,\pi^2 f^2 \cos(2\pi f t)\,. \qquad (1c)$$

In Figure A1.1 these waveforms are rescaled to arbitrary units for plotting on the same vertical axis. When downward compression toward the floor is taken in a positive sense, the classical pattern of acceleration is M-like in shape. The extreme value of acceleration at mid-compression is negative in direction, indicating acceleration away from the floor, opposite gravity.

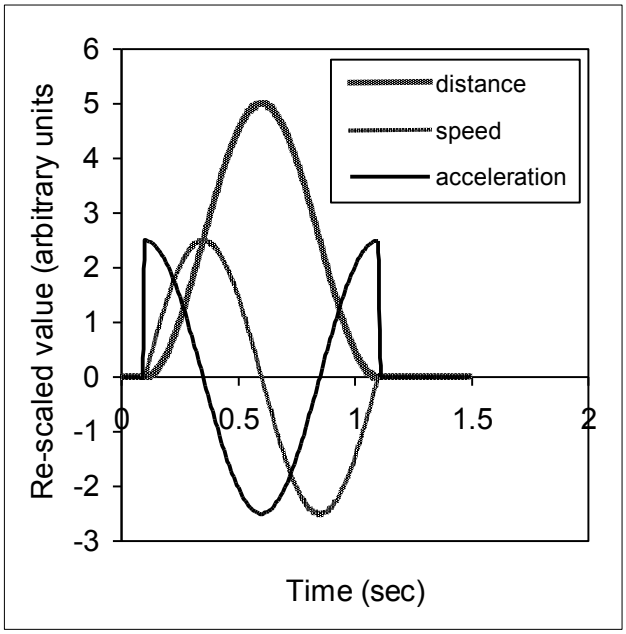

Figure A1.1. Waveforms of displacement distance, speed, and acceleration for one chest compression.

When sensed by a practical accelerometer, calibrated to +1g and to –1g, this acceleration away from the floor is felt by the cantilevered mass as being greater than +1g. The raw accelerometer data show a positive, mid-compression spike. Accordingly, to obtain the net downward acceleration of the chest compression pad in units of cm/sec^2, we rescaled raw acceleration data as follows:

a_cm_per_sec2 = -2*(981/(cal_plus1g – cal_minus1g))*(a_ADU – cal_plus1g), (2)

where the value of +1g = 981 cm/sec^2, and a_ADU is raw acceleration in analog-to-digital units.

Numerical integration of acceleration

To characterize the depth of chest compressions we developed a double integration scheme for the acceleration data that is based on an image processing approach. The underlying assumption is that the mean value of acceleration during a series of chest compressions is zero and that the mean value of sternal velocity of the patient during a series of chest compressions is also zero, despite noise or minor errors in offset or calibration of the device. Otherwise the patient would start moving somewhere. Further, we define the chest compression depth as the maximum minus the minimum downward position of the chest compression pad during a compression cycle.

Calibrated acceleration data were high-pass filtered by subtracting the average value of acceleration for a centered, 3-second wide moving window from each time-indexed value of acceleration. In this way the local mean of the filtered acceleration data points remained nearly zero. To obtain velocity of chest compression the filtered acceleration data were integrated using the trapezoidal rule. Velocity data were then high-pass filtered in a similar manner and integrated a second time to obtain displacement. Finally, the displacement results were high pass filtered, as before, to eliminate unrealistic drift. This double integration scheme was validated using test data generated from expression (1c), which produced results expected from expression (1a) after integration with less than 1 percent error.

Peak detection

We applied a moving peak detection operator to the resulting displacement data to find relative displacement during actual compressions. The output of the operator was positive if the indexed displacement value was 1 cm or greater than displacement values 100 msec on either side in the time domain. In this case the indexed point was labeled as a potential peak. Each contiguous sub-domain of points so labeled as potential peaks was then searched point-by-point to find the exact local maximum. This exact maximum and its sample time were logged as a compression peak.

In many records electrical defibrillation and pad removal produced large spikes in acceleration that were obviously different from genuine chest compressions to a human observer (Figure A1.2). Double integration of these spurious peaks generally yielded chest compression amplitudes greater than 12 cm in depth. Accordingly, we accepted compression peaks as genuine only if their amplitudes were greater than 1 cm and less than 12 cm. This selection process eliminated pad application and removal artifacts.

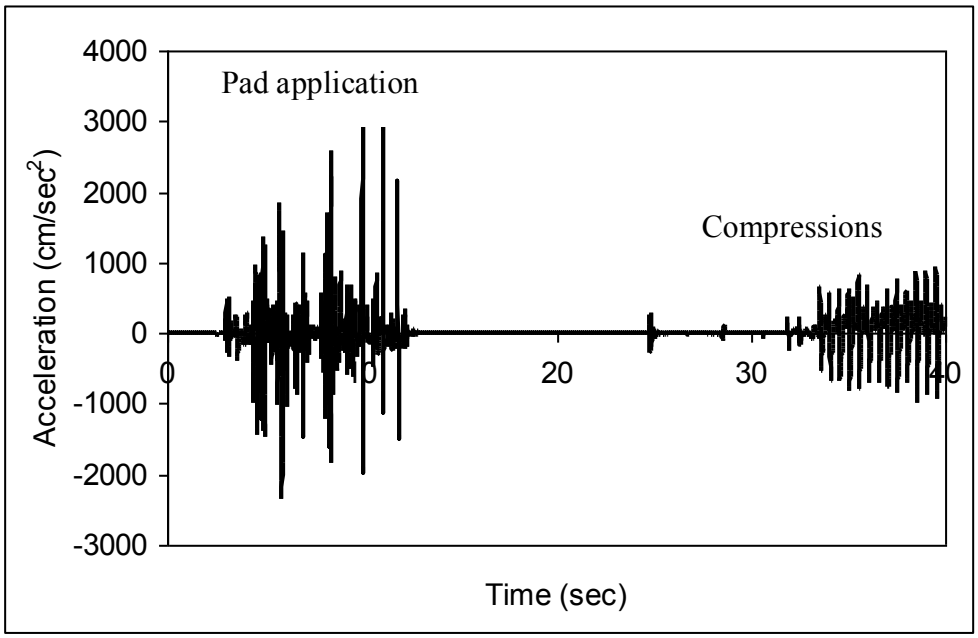

Figure A1.2. Chest acceleration data from a field resuscitation scaled to units of cm/sec^2. Positive values indicate acceleration of the instrumented chest compression pad toward the floor.

Appendix 2: Bias resistance of the Mood median test for pre-hospital CPR data

Suppose that we could reasonably estimate the fraction, θ, of hopeless cases in the data set, and that we could also estimate the fraction of these hopeless cases that are shifted by rescuer bias from the vigorous "good" CPR group to the less enthusiastic "bad" CPR group. Suppose we add θN hopeless cases to the study. All will be non-survivors. If these additions are split at random between treatment classes A and B (denoting bad vs. good CPR by some measure), then in the 2x2 table count a will be replaced by $a + (r/N)\theta N = a + r\theta$, and count c will be replaced by $c + (s/N)\theta N = c + s\theta$. However, if rescuer bias causes a difference in allocation of the hopeless cases of εθN cases from the group B to group A, that is if "bad" CPR is given to εθN more hopeless cases and "good" CPR is given to εθN fewer cases, for $\varepsilon < 1$, then count a will be replaced by $a + r\theta + \varepsilon\theta N$, and count c will be replaced by $c + s\theta - \varepsilon\theta N$. The augmented 2x2 table will look like this:

	Group A	**Group B**	**Total**
Non-survivors	$a' = a + r\theta + \varepsilon\theta N$	$c' = c + s\theta - \varepsilon\theta N$	$m' = m + \theta N$
Survivors	b	d	n
Total	$r' = r + r\theta + \varepsilon\theta N$	$s' = s + s\theta - \varepsilon\theta N$	$N' = N + \theta N$

Now let us analyze the 2x2 table including hopeless cases and bias just as before. Under null hypothesis we would calculate that now the observed distribution of b will have mean $\mu'(b) = nr'/N'$ and variance $\sigma'^2(b) = nm'r's'/(N'^2(N'-1))$. That is,

$$\mu'(b) = \frac{nr'}{N'} = \frac{nr(1+\theta+\varepsilon\theta N/r)}{N(1+\theta)} = \mu(b)\left(1+\frac{\varepsilon\theta N/r}{1+\theta}\right).$$

Similarly, for larger N's (with $N \approx N-1$) we would have

$$\sigma'^2(b) = \frac{nm'r's'}{N'^3} = \frac{nmrs}{N^3}\cdot\frac{(1+\theta N/m)(1+\theta+\varepsilon\theta N/r)(1+\theta-\varepsilon\theta N/s)}{(1+\theta)^3} \approx \frac{nmrs}{N^3}\cdot\frac{(1+\theta N/m)}{(1+\theta)}$$

In turn, we can estimate that the standard deviation of count, b, would be

$$\sigma'(b) \approx \sqrt{\frac{nmrs}{N^3}\cdot\frac{(1+\theta N/m)}{(1+\theta)}} = \sqrt{\frac{nmrs}{N^3}\cdot\frac{(1+\theta+\theta N/m-\theta)}{(1+\theta)}} = \sigma(b)\sqrt{1+\frac{\theta}{1+\theta}\left(\frac{N}{m}-1\right)}.$$

The new test statistic would be

$$Z' = \frac{b - \mu'(b)}{\sigma'(b)} \text{ and the new probability } P' = \int_{-\infty}^{Z'} f(x)dx .$$

For a positive association between group B, "better" CPR and survival we would have $b < \mu(b)$ and $Z < 0$. Then, with the addition of hopeless cases and bias we would have

$$Z' \approx \frac{b - \mu(b)\left(1 + \frac{\varepsilon\theta}{1+\theta} \cdot \frac{N}{r}\right)}{\sigma(b)\sqrt{1 + \frac{\theta}{1+\theta}\left(\frac{N}{m} - 1\right)}},$$

with offsetting errors! The numerator is increased in absolute value by selection bias and the denominator is increased by the addition of hopeless cases. The amount by which standard normal variable, Z is changed can be appreciated by re-writing Z' as

$$Z' \approx \frac{b - \mu(b) - \mu(b)\left(\frac{\varepsilon\theta}{1+\theta} \cdot \frac{N}{r}\right)}{\sigma(b)\sqrt{1 + \frac{\theta}{1+\theta}\left(\frac{N}{m} - 1\right)}} = Z\frac{1 + \frac{\varepsilon\theta}{1+\theta} \cdot \frac{N}{r} \cdot \frac{\mu(b)}{b - \mu(b)}}{\sqrt{1 + \frac{\theta}{1+\theta}\left(\frac{N}{m} - 1\right)}}$$

or

$$\frac{Z'}{Z} \approx \frac{1 + \frac{\varepsilon\theta}{1+\theta} \cdot \frac{N}{r} \cdot \frac{\mu(b)}{b - \mu(b)}}{\sqrt{1 + \frac{\theta}{1+\theta}\left(\frac{N}{m} - 1\right)}}.$$

Now it is easy to appreciate the offsetting errors. They mean that with the 2x2 table test the values of test statistic Z and the associated P-values are not changed much by the problems that plague non-randomized pre-hospital resuscitation data. This is a remarkable and lucky result.

Consider the following example, which gives an educated guess as to the situation in the real world. Suppose $\theta = 1$, indicating that half the patients in pre-hospital series are hopeless cases. Suppose $\varepsilon = 0.1$, indicating that there is a 60% / 40% split in the hopeless cases, favoring bad (sloppier or lackadaisical technique). Suppose that $N/r = 2$, that is the median is used for a cutoff. Then we would have

$$\frac{Z'}{Z} \approx \frac{1 + 0.1 \cdot \dfrac{\mu(b)}{b - \mu(b)}}{\sqrt{1 + 0.5}}.$$

For results approaching statistical significance the observed count of survivors receiving group A, CPR would be less than that expected under H_0, and we might guess for a typical case that $\mu(b)/(\mu(b) - b) \sim 2$. In this case we would have

$$\frac{Z'}{Z} \approx \frac{1.2}{\sqrt{1.5}} = 0.98.$$

There is very little difference in the calculated value of Z with added hopeless cases and bias and the calculated value of Z if these confounding factors were absent. In this sense the 2x2 table analysis provides a robust way to evaluate retrospective field data in CPR and resuscitation. Hopeless cases tend to make the results worse, but rescuer bias tends to make the results better. If there is a true association or correlation between survival and a performance measure, the statistical method will be able to separate it from the effects of random sampling variation anyway.

References

1. Cooper BE. Statistics for Experimentalists. Oxford: Permagon Press Ltd., 1969:336 pages.
2. Ewy GA. Cardiac arrest--guideline changes urgently needed. Lancet 2007; 369:882-4.
3. Sunde K, Eftestol T, Askenberg C, Steen PA. Quality assessment of defribrillation and advanced life support using data from the medical control module of the defibrillator. Resuscitation 1999; 41:237-47.
4. Eilevstjonn J, Kramer-Johansen J, Eftestol T, Stavland M, Myklebust H, Steen PA. Reducing no flow times during automated external defibrillation. Resuscitation 2005; 67:95-101.
5. Cammarata G, Weil MH, Csapoczi P, Sun S, Tang W. Challenging the rationale of three sequential shocks for defibrillation. Resuscitation 2006; 69:23-7.
6. Yakaitis RW, Ewy GA, Otto CW, Taren DL, Moon TE. Influence of time and therapy on ventricular defibrillation in dogs. Critical Care Medicine 1980; 8:157-163.
7. Cobb LA, Fahrenbruch CE, Walsh TR, et al. Influence of cardiopulmonary resuscitation prior to defibrillation in patients with out-of-hospital ventricular fibrillation. Jama 1999; 281:1182-8.
8. Babbs CF, Voorhees WD, Fitzgerald KR, Holmes HR, Geddes LA. Relationship of artificial cardiac output to chest compression amplitude--evidence for an effective compression threshold. Annals of Emergency Medicine 1983; 12:527-532.
9. Babbs CF. Relative effectiveness of interposed abdominal compression CPR: Sensitivity analysis and recommended compression rates. Resuscitation 2005.
10. Ristagno G, Tang W, Chang YT, et al. The quality of chest compressions during cardiopulmonary resuscitation overrides importance of timing of defibrillation. Chest 2007; 132:70-5.
11. Odegaard S, Kramer-Johansen J, Bromley A, et al. Chest compressions by ambulance personnel on chests with variable stiffness: abilities and attitudes. Resuscitation 2007; 74:127-34.
12. Greingor JL. Quality of cardiac massage with ratio compression-ventilation 5/1 and 15/2. Resuscitation 2002; 55:263-7.
13. Wik L, Kramer-Johansen J, Myklebust H, et al. Quality of cardiopulmonary resuscitation during out-of-hospital cardiac arrest. Jama 2005; 293:299-304.
14. Abella BS, Alvarado JP, Myklebust H, et al. Quality of cardiopulmonary resuscitation during in-hospital cardiac arrest. Jama 2005; 293:305-10.
15. Hightower D, Thomas SH, Stone CK, Dunn K, March JA. Decay in quality of closed-chest compressions over time. Ann Emerg Med 1995; 26:300-3.
16. Edelson DP, Abella BS, Kramer-Johansen J, et al. Effects of compression depth and pre-shock pauses predict defibrillation failure during cardiac arrest. Resuscitation 2006; 71:137-45.
17. Gilmore CM, Rea TD, Becker LJ, Eisenberg MS. Three-phase model of cardiac arrest: time-dependent benefit of bystander cardiopulmonary resuscitation. Am J Cardiol 2006; 98:497-9.