



---

UW Biostatistics Working Paper Series

---

11-20-2009

# Is Survival the Only or Even the Right Outcome for Evaluating Treatments for Out-of-Hospital Cardiac Arrest? A Proposed Test Based on Both an Intermediate and Ultimate Outcome.

Al Hallstrom

*Univ of Washington, aph@u.washington.edu*

---

## Suggested Citation

Hallstrom, Al , "Is Survival the Only or Even the Right Outcome for Evaluating Treatments for Out-of-Hospital Cardiac Arrest? A Proposed Test Based on Both an Intermediate and Ultimate Outcome." (November 2009). *UW Biostatistics Working Paper Series*. Working Paper 352.

<http://biostats.bepress.com/uwbiostat/paper352>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Is survival the only or even the right outcome for evaluating treatments for out-of-hospital cardiac arrest? A proposed test based on both an intermediate and ultimate outcome.

By Alfred P. Hallstrom, Ph.D. Biostatistics Department, University of Washington

Funding for this work was provided by the University of Washington

Key words: intermediate endpoint, sample size, ROSC, survival, cardiac arrest, clinical trial, surrogate, combination test

Abstract word count 206

Word count 3249

Correspondence

Al Hallstrom

1107 N.E. 45<sup>th</sup>, Room 505

Seattle, WA, 98102

206 685 1302

[aph@u.washington.edu](mailto:aph@u.washington.edu)



## *Abstract*

It is generally agreed that the goal of resuscitation is survival with neurological and physiological status similar to that preceding the cardiac arrest. Previously I have argued that the lack of improvement in outcome from resuscitation over the past 3 to 4 decades, as compared to the substantial progress made in treatment of ischemic heart disease, is a consequence of the absence of randomized clinical trials of new interventions and the use of intermediate endpoints such as return of spontaneous circulation or admittance to hospital. Proponents of these intermediate endpoints have argued that those involved in the resuscitation have no control over what care is undertaken in the hospital and hence hospital mortality only adds noise, at best, thus making survival a less sensitive and less relevant endpoint for evaluation of resuscitation interventions. Recent reports of improvement in hospital survival have caused me to consider that their argument may have more validity than I had supposed. In this note I propose a test that gives weight both to the intermediate endpoint and survival. The test is responsive to the primary goal of testing survival with limited loss of power compared to a test based only on the intermediate endpoint. The test is illustrated with several examples.



## Introduction

It is generally agreed that the goal of resuscitation is survival with neurological and physiological status similar to that preceding the cardiac arrest [2008 AHA endpoint conference] Previously I have argued that the lack of improvement in outcome from resuscitation over the past 3 to 4 decades, as compared to the substantial progress made in treatment of ischemic heart disease, is a consequence of the absence of randomized clinical trials of new interventions and the use of intermediate endpoints such as return of spontaneous circulation (ROSC) or admittance to hospital. Proponents of these intermediate endpoints have argued that those involved in the resuscitation have no control over what care is undertaken in the hospital and hence hospital mortality only adds noise making survival a less sensitive and less relevant endpoint for evaluation of resuscitation interventions. Recent reports of improvement in hospital survival post resuscitation have caused me to consider that their argument may have more validity than I had supposed. [1, 2] Proponents also point out (figure 1) that sample sizes can be substantially smaller with the use of intermediate endpoints. The justification for using such intermediate outcomes as the primary endpoint is the belief that they represent a reasonable surrogate for survival. That is, believe that conditional upon achieving the intermediate endpoint, subsequent survival should not depend upon an ongoing effect from the interventions administered prior to achieving the intermediate endpoint.[3] A second reason for use of an endpoint such as ROSC is logistic; such information can be obtained directly from the agency involved in the resuscitation whereas survival requires follow-up with community hospitals which adds cost and complexity for its acquisition. However it should be noted that, in this context, the survival outcome does not require lengthy follow-up, a common argument for considering an intermediate endpoint. Previously I proposed using the bivariate outcome (intermediate, ultimate) e.g. (hospital admittance, survival) and applying a multivariate test such as Hotelling's  $T^2$ . [4] Under the assumption of surrogacy such an approach also reduces required sample sizes substantially. However, the test is likely to be significant if there is a large intermediate effect but no net effect and thus this approach suffers from many of the issues that relate to surrogate endpoints.

The perils of surrogate endpoints are well recognized.[5, 6] Indeed, it can be argued that to establish an intermediate endpoint as a surrogate for a specific condition and a specific intervention would require a trial of such size as to make the use of the intermediate unnecessary.[7] Instead, believe in an intermediate endpoint as a surrogate develops over time as it is found to fit the criteria over a number of interventions and through basic research that identifies causal pathways.[8-11] In a trial (TeleCPR) of bystander administered cardio-pulmonary resuscitation (CPR) comparing chest compression only versus ventilation and chest compression based on instructions administered over the telephone by EMS dispatchers, admittance to the hospital appeared to be a reasonable surrogate for survival.[12] However, at least two trials of interventions in resuscitation question the validity of hospital admittance as a surrogate for survival (ARREST, a trial of amiodarone (an antiarrhythmia drug) administered at the scene, which found increased hospital admittance but no effect on survival; and ASPIRE, a trial of mechanical CPR, which found no effect on hospital admittance but decreased survival).[13, 14] Absence of surrogacy could be that survival, conditional on reaching the positive intermediate outcome, was either reduced or increased for the intervention compared to control. I will refer to these as sub- (reverse- if all apparent gain from the intermediate is lost) and super-surrogate conditions, as compared with surrogacy in which the effect of the intervention is directionally similar and proportionately comparable between the intermediate endpoint and final outcome.

A desirable test would reject the intervention if survival was reduced, would optimally weight both the intermediate and survival if there was evidence of super-surrogacy, would rely on the intermediate if there was no evidence against surrogacy, and would give relatively more weight to survival if there was evidence of sub or reverse-surrogacy. In this note I propose such a test, investigate its properties through simulation, and discuss what such a test might have meant for the TeleCPR, ARREST and ASPIRE trials and for a potential trial of mild hypothermia (mild hypothermia initiated at the scene of the arrest is an intervention that might be supposed to be a super-surrogate in that cooling might help achieve ROSC (if given prior to it) and might improve survival conditional on ROSC through its impact on the brain and other vital organs. [2, 15])

### *The Proposed Test W*

#### *Notation*

Let  $I$  represent the intermediate endpoint and  $S$  the ultimate endpoint of survival. Let intervention,  $T$ , and control,  $C$ , represent the therapies to be compared. Suppose  $2N$  patients are to be randomized,  $N$  to treatment and  $N$  to control. Let  $P_C$  and  $P_T = P_C + \Delta$  be the anticipated rates for  $I$  under control  $C$  and treatment  $T$ . Similar let  $Q_C$  and  $Q_T = Q_C + \delta$  be the anticipated conditional survival,  $SI$ , rates. Let  $p_T, p_C, q_T, q_C$  be the corresponding values for the observed data. Let  $Z(Z_S, Z_I, Z_{SI})$  be the normal approximations for the tests comparing proportions.

#### *When Survival is Reduced ( $Z_S < 0$ )*

If  $Z_S < 0$  (i.e. evidence of harm) set  $W$  to  $Z_S$  or arbitrarily low, say  $W = 0$ .

#### *When there is Potential Super-surrogacy ( $Z_S \geq 0$ and $Z_{SI} > 0$ )*

If  $Z_S \geq 0$  (i.e. no evidence of worsened survival) and  $Z_{SI} > 0$  (i.e. potential of super-surrogacy) set

$$W = \frac{\text{sign}(Z_I)Z_I^2 + Z_{SI}^2}{\sqrt{Z_I^2 + Z_{SI}^2}}. \quad \text{This is an approximately optimal test of the combined outcomes}$$

(appendix 1).

#### *When there is Evidence of Sub or Reverse Surrogacy ( $Z_S \geq 0$ and $Z_{SI} < C_L < 0$ )*

If  $Z_S \geq 0$  and  $Z_{SI} < C_L = 0.6\hat{E}_{RS}$  where  $C_L$  is negative (evidence of sub or reverse-surrogacy) define  $W = Z_I + 3Z_{SI}$  which gives 3 times as much weight to survival as to the intermediate. Here

$$E_{RS} = E(Z_{SI} | \delta = -Q_C \Delta / P_T) = \frac{-Q_C \Delta / P_T}{\sigma_{SI}} \sqrt{N}$$

is the expected value of  $Z_{SI}$  under  $\delta = -Q_C \Delta / P_T$  which corresponds to the null for  $S$  (i.e. reverse-surrogacy). The choice of the factor 0.6 and the relative weight of 3 was based on simulations with the object of protecting against accepting the intervention when survival was actually worsened while maintaining the power of the intermediate outcome.

When there is no Evidence of Ultimate Harm and little or no Evidence against Surrogacy  
 $(Z_s \geq 0 \text{ and } 0 \geq Z_{sl} \geq C_L)$

If  $Z_s \geq 0$  and  $0 \geq Z_{sl} \geq C_L$  define  $W = Z_I$  thus taking advantage of the greater power of the intermediate endpoint when surrogacy is likely.

### Properties of $W$

Under the compound null and across a wide range of control rates,  $0.1 \leq P_C \leq 0.9$  and  $0.1 \leq Q_C \leq 0.9$ , the 95%tile of  $W$  is given by (appendix 2),

$$W_{.95} = 1.956 + 0.09672Q_C - 0.05067P_C$$

and the 97.5%tile is given by

$$W_{.975} = 2.273 + 0.07608Q_C - 0.037P_C$$

Indeed for  $.8 < x < .98$ , to a very good approximation

$$W_x = \text{qnorm}(x) + (-2.213) + 5.2 * x + (-2.676 * x ** 2) + (-.4245 - .3949 * x) * P_C + (.8346 - .7791 * x) * Q_C .$$

The power (appendix 3) of  $W$ ,  $Z_I$  and  $Z_s$  for a 1 sided .05 level test is displayed in Table 1 for some typical intermediate and conditional control outcome rates:  $(P_C, Q_C) = (0.3, 0.48)$ ,  $(0.4, 0.36)$  and  $(0.6, 0.24)$ , all yielding the same control survival rate of 0.144, under alternatives for the intermediate of  $P_T = P_C$  and  $1.4P_C$  (respectively null and substantially better intermediate outcome under treatment  $T$  than control  $C$ ); and alternatives for the conditional outcome of  $Q_T = 1.2 Q_C$ ,  $Q_C$ ,  $P_C Q_C / P_T$ , and  $0.8 P_C Q_C / P_T$  (respectively super-surrogacy, surrogacy, reverse-surrogacy, and moderately worsened survival). The sample sizes were chosen to correspond to a power of 0.9 for  $W$  under the substantially better intermediate outcome ( $P_T = 1.4P_C$ ) and surrogacy ( $Q_T = Q_C$ ).

When the null holds for the intermediate outcome,  $W$  and  $Z_s$  perform similarly (rows 1, 2, and 3).

When there is a substantial effect on the intermediate outcome, there are clear differences between the tests. Consider the values in italics which correspond roughly to current control rates for ROSC and survival conditional on ROSC. Under super-surrogacy,  $W$  and  $Z_I$  have the same power, .97, 10% higher than  $Z_s$ . Under surrogacy  $Z_I$  has a power of .97 compared to .90 for  $W$  and .53 for  $Z_s$ . Under the null for survival  $W$  has a power of .38. Under worsened survival  $W$  will exceed the critical value 10% of the time. The corresponding  $N$ 's required to have a power of .9 for  $Z_I$  and  $Z_s$  are 170 and 725 respectively.

Results are similar across a wide range of control rates. For

$$0.3 \leq P_C \leq 0.6, 0.2 \leq Q_C \leq 0.6, \text{ and } P_C \leq P_T \leq (1 + P_C) / 2,$$

sample sizes for 80% and 90% power for  $W$ , assuming surrogacy, are approximated by

$$N_{.8} \approx \exp(.7799 - (1.994 \log(P_T - P_C) + .179 \log(Q_C) + .2421 \log(1 - Q_C)))$$

and

$$N_{.9} \approx \exp(.7042 + .2159 \log(P_T) - (1.94 \log(P_T - P_C) + 1.179 \log(Q_C) + .6551 \log(1 - Q_C))).$$

### *Dependence of the intermediate outcome and survival conditional on the intermediate outcome*

The above results assumed independence of the intermediate outcome and survival conditional on the intermediate outcome. Since the dependency structure will generally be unknown, the critical values for  $W$  must be determined under an independence model. However even a strong dependence (obtained by requiring a given case to have probabilities for each outcome with the same %tile in the respective distributions, i.e. those more likely to achieve the intermediate outcome are more likely to survive if the intermediate outcome is achieved and conversely those less likely to achieve the intermediate outcome are less likely to survive even if the intermediate outcome is achieved) had negligible impact on the test size (i.e. on the alpha level of the test). Indeed under this strong but reasonable dependence, the size of  $W$  is negligibly smaller than that of  $Z_S$  (based on 4000 simulations with  $N = 1000$  and over  $0.1 \leq P_C \leq 0.9$  and  $0.1 \leq Q_C \leq 0.9$  -appendix 4).

### *Discussion*

$$W = \begin{cases} 0 & \text{if } Z_S < 0 \\ \frac{\text{sign}(Z_I)Z_I^2 + Z_{SI}^2}{\sqrt{Z_I^2 + Z_{SI}^2}} & \text{if } Z_S \geq 0 \ \& \ Z_{SI} > 0 \\ Z_I & \text{if } Z_S \geq 0 \ \& \ C_L \leq Z_{SI} \leq 0 \\ Z_I + 3Z_{SI} & \text{if } Z_S \geq 0 \ \& \ Z_{SI} < C_L \end{cases}$$

$W$  attempts to capture the added efficiency when there is evidence for super-surrogacy, the usual efficiency of the intermediate outcome when there is evidence for surrogacy and attempts to avoid type I error when survival is actually worsened by giving substantial weight to the conditional survival comparison when there is evidence of substantial sub-surrogacy. Even the extreme (though directionally likely) violation of the assumption of independence of the intermediate and conditional survival outcomes has little effect on the size of  $W$ .  $W$  maintains power under super-surrogacy, has modest losses in power under surrogacy, and substantial protection against claiming an advantage when survival is worse. There were only modest effects on  $W$  when the constant 3 was varied between 2.5 and 3.5. Varying the constant 0.6 in the value for  $C_L$  essentially traded power under surrogacy for protection against surrogacy failure.

Note that  $W$  is essentially a one-sided test, which is appropriate for trials comparing a new therapy to a standard of care.

### *The TeleCPR Study*

The TeleCPR study evaluated the effect on survival of bystander CPR administered according to instructions from the EMS dispatcher.[12] The randomly assigned instructions either included steps for ventilation and chest compression (ABC-CPR), the standard of care, or steps for chest compression only (CC-CPR). The trial was designed as 1-sided, testing whether CC-CPR was superior to ABC-CPR as a previous 1-sided trial had demonstrated that CC-CPR was not inferior to ABC-CPR. The trial terminated with slightly over 500 of a planned 700 patients enrolled when funding ran out. The data is shown in table 2a. Although there is a slightly greater survival rate conditional on admittance to the hospital for the CC-CPR arm (36.1% versus 30.5%), this was not significant (one-sided  $p =$

0.25) so surrogacy could not be rejected. The one-sided p-values for  $T^2$ ,  $W$ ,  $Z_I$  and  $Z_S$  are 0.12, 0.088, 0.084, and 0.097 respectively. One of the problems with the bivariate outcome is defining a good and a bad outcome. However, in[4] we show, based on economic considerations, that essentially half of the alpha will result in declaring a good outcome under the null. Because of the slight super-surrogacy,  $W$  is almost as powerful as  $Z_I$  and  $Z_S$  loses relatively little power.

### *The ASPIRE Study*

The ASPIRE study was intended to compare manual CPR (the standard of care) with mechanical CPR provided by a load distributing band device (the AutoPulse).[13] To avoid the idiosyncrasies of criteria for hospital admission in difference cities, the study considered the endpoint of alive a 4 hours post the time of call reporting an out-of-hospital cardiac arrest. Investigators estimated that control rates for this outcome would be 0.178 and anticipated a 35% relative improvement to 0.24 under the intervention. The estimated rate of survival conditional on being alive at 4 hours was about .4. It was unclear whether any additional increment should be expected for survival conditional on being alive at 4 hours in the intervention arm. The sponsor (a startup company funded by venture capital) and the investigators wanted to do a survival trial, although it was felt that a neutral impact on survival would be a positive result as the device might substantially reduce manpower requirements. Ignoring issues such as sequential monitoring, the sample size required ( $2\alpha = 0.05$ , power = 0.9) for the intermediate endpoint was  $2N \sim 1804$  and for survival, assuming surrogacy, was  $2N \sim 5320$ . It was consider impossible to raise enough money to do a survival study. The sample size that would have been required for  $W$  and  $T^2$  would have been  $2N \sim 2224$  and 2150 respectively. The study was terminated early due to concerns about safety. The data is shown in Table 2b. The slight trend toward an improvement in 4 hour survival was not significant. The one-sided p-values for  $T^2$ ,  $Z_I$  and  $Z_S$  are 0.01, 0.32, and 0.024 [note the reduction in survival (the secondary endpoint) was significant at  $p = 0.03$  after adjustment for baseline covariates]. The point estimates for the significant bivariate and survival outcomes are indicative of harm. The results of the study are being largely ignored by the manufacturer on several grounds including that survival was not the primary endpoint and the difference was not significant ( $p = 0.06$  when considered as a two-sided test). Whether or not the sponsor could/would have raised the capitol to finance the study if the outcome was tested with  $W$  is unclear. The study would undoubtedly have been terminated in the same time frame and the value of  $W$  would have been 0 as  $Z_S < 0$ .

### *The ARREST Study*

The ARREST study evaluated the use of amiodarone in patients with out-of-hospital cardiac arrest who had not been resuscitated after 3 or more precordial shocks.[14] This single center study used hospital admittance as the primary outcome measure with survival as a secondary outcome. This choice was again made on the practical grounds of financing and feasibility of conducting the study. Assuming a control rate for hospital admission of 0.41 and a 30% relative improvement with use of amiodarone, and assuming a survival rate conditional on admittance of approximately .38 and surrogacy, the sample sizes needed for a  $2\alpha = 0.05$  and power = 0.8 design using  $T^2$ ,  $W$ ,  $Z_I$  and  $Z_S$  would be  $2N = 630, 680, 500,$  and 2090 respectively. The data are shown in Table 2c. The improvement in admittance rates was significant (one-sided  $p = 0.019$ ). Survival was not worse (one-



sided  $p = 0.52$ ). It is likely the site might have elected to set the sample size at 680 if they were convinced that the test  $W$  were a valid means of investigating survival. The site chose not to implement amiodarone as part of standard protocol even though the trial, as designed, had a significantly positive outcome. However, many other EMS systems did include amiodarone in their resuscitation protocols. For the data collected,  $C_L = -1.027$ ,  $Z_{SU} = -1.133$ ,  $W_{.975} = 2.284$ , and  $W = -1.234$ . In this case  $W$  would have provided protection against a positive result for an intermediate endpoint where there was evidence of reverse-surrogacy. The one-sided  $p$ -value for  $T^2$  was 0.025, providing an example of how the bivariate outcome can suffer the same problems as the intermediate.

### *A proposed trial of Mild Hypothermia*

A novel means of cooling through the nose has been demonstrated, in animal studies, to produce a large increase in ROSC long before any real cooling of the brain or vital organs could occur, [16] but also provides rapid core body cooling suggesting the possibility of additional improvement in hospital survival as demonstrated in several trials. [2, 15] The control rates based on current outcomes at a number of sites planning to participate in a randomized trial of this novel approach and anticipated relative improvements are shown in table 3.

Simulation results for  $Z_I$ ,  $Z_S$ , and  $W$  are shown in Table 4 based on sample sizes which yield a power for  $Z_I$ ,  $Z_S$  and  $W$  of 0.9 for a 1 sided 0.05 level. If the trial were sized for  $W$  and the super-surrogacy assumption holds  $W$  would preserve most of the power of the intermediate, but would relinquish about 7 % power compared to the intermediate. However, under a modest negative effect on survival  $W$  would indicate a significant positive effect 26% of the time. At first glance this appears to be unacceptable, but is surely better than using  $Z_I$  which would have indicated a positive effect 98% of the time. Also recall that for  $W$  to be significantly positive  $Z_S$  has to be positive and thus the data from the trial would not indicate any worsening of survival. This is similar to the ARREST situation, except that the effect of mild hypothermia on the intermediate outcome would have been much greater and would probably result in adoption of the therapy by many EMS agencies. Had the trial been sized for  $Z_S$  ( $2N = 1900$ ) but tested with  $W$  this undesirable result would have occurred 14% of the time, while the power under surrogacy would be almost 100%. Under the surrogacy assumption, the sample size for the bivariate outcome would be  $2N = 526$ .

### *Would the use of $W$ change the current assessment of the outcomes of clinical trials in EMS?*

Ignoring regulatory/approval concerns, the EMS community, being a group of clinician/scientists, would consider all evidence including both the intermediate and the survival outcomes when evaluating the results of a trial, whether the trial was sized for and tested with  $T^2$ ,  $Z_I$ ,  $Z_S$  or  $W$ . What might the collective judgment be in the case of concern where the unknown truth is that the intervention has a large positive effect on the intermediate outcome but a modest negative effect on survival and  $W$  yielded a significant  $p$ -value? Figure 2 shows the distribution of  $Z_I$  and  $Z_S$  for such a situation (defined by the 4<sup>th</sup> row of Table 4). In all cases a large effect on the intermediate and a trend (often substantial) toward a positive effect on survival would be seen. I suspect the EMS

community would be inclined to adopt the intervention whether the trial was based on  $Z_S$  or  $W$ . If based on  $Z_S$  adoption would usually be without the “stamp of approval, i.e.,  $p < 0.05$ ”.

There are ethical and practical arguments for and against using  $W$  in a clinical trial. Some would argue that only survival matters and only  $Z_S$  should be used. Others would argue that if there is no net effect on survival then improvement in the ROSC rate is good because more patients may survive with improving hospital care. Still others would argue that if there is no net effect on survival then improvement in the ROSC rate is bad because it results in consumption of unnecessary and expensive health resources. Others would argue that too large of sample sizes would have a negative effect on efforts to find and test potentially beneficial interventions. Even if therapies are developed and suggest potential in phase II trials using an intermediate endpoint, survival trials would probably have to be funded by NIH. Current funding for the Resuscitation Outcomes Consortium (<http://www.nih.gov/news/pr/mar2006/nhlbiv1-24.htm>) constitutes only a small fraction of what would be needed to conduct all the needed trials. Although each scenario would need to be evaluated, in my opinion the proposed test,  $W$ , is a compromise that may provide for robust and cost-effective evaluation.

## References

1. Sunde, K., et al., *Implementation of a standardized treatment protocol for post resuscitation care after out-of-hospital cardiac arrest*. Resuscitation, 2007. **73**(1): p. 29-39.
2. *Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest*. N Engl J Med, 2002. **346**(8): p. 549-56.
3. Prentice, R.L., *Surrogate endpoints in clinical trials: definition and operational criteria*. Stat Med, 1989. **8**(4): p. 431-40.
4. Hallstrom, A.P., *What is the appropriate outcome for studies of treatments for out-of-hospital cardiac arrest?* Resuscitation, 2006. **71**(2): p. 194-203.
5. Fleming, T.R. and D.L. DeMets, *Surrogate end points in clinical trials: are we being misled?* Ann Intern Med, 1996. **125**(7): p. 605-13.
6. Fleming, T.R., *Surrogate endpoints and FDA's accelerated approval process*. Health Aff (Millwood), 2005. **24**(1): p. 67-78.
7. Prentice, R.L., *Surrogate and mediating endpoints: current status and future directions*. J Natl Cancer Inst, 2009. **101**(4): p. 216-7.
8. Yothers, G., *Toward progression-free survival as a primary end point in advanced colorectal cancer*. J Clin Oncol, 2007. **25**(33): p. 5153-4.
9. Daniels, M.J. and M.D. Hughes, *Meta-analysis for the evaluation of potential surrogate markers*. Stat Med, 1997. **16**(17): p. 1965-82.
10. Lipicky, R.J. and M. Packer, *Role of surrogate end points in the evaluation of drugs for heart failure*. J Am Coll Cardiol, 1993. **22**(4 Suppl A): p. 179A-184A.
11. Petrylak, D.P., et al., *Evaluation of prostate-specific antigen declines for surrogacy in patients treated on SWOG 99-16*. J Natl Cancer Inst, 2006. **98**(8): p. 516-21.
12. Hallstrom, A., et al., *Cardiopulmonary resuscitation by chest compression alone or with mouth-to-mouth ventilation*. N Engl J Med, 2000. **342**(21): p. 1546-53.
13. Hallstrom, A., et al., *Manual chest compression vs use of an automated chest compression device during resuscitation following out-of-hospital cardiac arrest: a randomized trial*. Jama, 2006. **295**(22): p. 2620-8.

14. Kudenchuk, P.J., et al., *Amiodarone for resuscitation after out-of-hospital cardiac arrest due to ventricular fibrillation*. N Engl J Med, 1999. **341**(12): p. 871-8.
15. Bernard, S.A., et al., *Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia.[comment]*. New England Journal of Medicine., 2002. **346**(8): p. 557-63.
16. Wolfson MR, M.D., Wu J, Hoffman J, Rozenberg A, Shaffer TH, Barbut D, *Intranasal Perfluorochemical Spray for Preferential Brain Cooling in Sheep*. . Neurocritical Care 2008. **8**(3): p. 437-447.



Table 1

Power of  $Z_I$ ,  $Z_S$  and  $W$  for a 1 sided .05 level test

$I$	$SI$	$P_C, Q_C = 0.3, 0.48$			$P_C, Q_C = 0.4, 0.36$			$P_C, Q_C = 0.6, 0.24$		
		N=348			N=250			N=203		
		$Z_I$	$Z_S$	$W$	$Z_I$	$Z_S$	$W$	$Z_I$	$Z_S$	$W$
Null	Super-surrogacy	0.051	0.272	0.281	0.051	0.226	0.197	0.048	0.199	0.157
	Surrogacy (Null for $SI$ )	0.051	0.051	0.051	0.052	0.052	0.052	0.048	0.050	0.050
	Worsened survival	0.051	0.003	0.013	0.052	0.005	0.014	0.050	0.006	0.014
Substantial effect	Super-surrogacy	0.955	0.953	0.967	0.973	0.873	0.972	1.000	0.810	0.988
	Surrogacy (Null for $SI$ )	0.953	0.648	<b>0.901</b>	0.975	0.531	<b>0.900</b>	1.000	0.460	<b>0.900</b>
	Reverse surrogacy (Null survival)	0.954	0.050	0.352	0.974	0.050	0.376	1.000	0.050	0.414
	Worsened survival	0.953	0.003	0.062	0.975	0.005	0.100	1.000	0.006	0.152



Table 2				
Table 2a				
Results of the TeleCPR Trial				
	N	Admitted to Hosp	Survived Admitted	Survived
ABC-CPR	278	95 (34.2%)	29 (30.1%)	29 (10.4%)
CC-CPR	240	97 (40.4%)	35 (36.1%)	35 (14.6%)
Table 2b				
Results of the ASPIRE Trial				
	N	Alive at 4 hours	Survived Alive4hrs	Survived
Manual CPR	373	92 (24.7%)	37 (40.2%)	37 (9.9%)
AutoPulse	394	104 (26.4%)	23 (22.1%)	23 (5.8%)
Table 2c				
Results of the ARREST Trial				
	N	Admitted	Survived Admitted	Survived
Control	258	89 (34.5%)	34 (38.2%)	34(13.2%)
Amiodarone	246	108(43.9%)	33 (30.6%)	33 (13.4%)



Table 3			
Expected control and intervention rates for the intermediate outcome (ROSC) and survival conditional on ROSC in a trial of mild hypothermia treatment for Out-of-Hospital Cardiac Arrest.			
	Control Rates	Expected Rates Under Intervention	Assumed Improvement
<i>I</i> (ROSC)	40%	54%	14%
<i>SI</i> (Survival conditional on ROSC)	36%	41.4%	5.4%



Table 4

Power of  $Z_I$ ,  $Z_S$  and  $W$  for a 1 sided .05 level test with  $P_C, Q_C = 0.4, 0.36$

$I$	$SI$		2N=422			2N=1900			2N=650		
			$Z_I$	$Z_S$	$W$	$Z_I$	$Z_S$	$W$	$Z_I$	$Z_S$	$W$
$\Delta = .14$	$\delta = .054$	Super-surrogacy	0.900	0.684	0.873	1.000	0.998	1.000	0.976	0.838	0.966
Large	$\delta = 0$	Surrogacy (Null for $SI$ )	<b>0.901</b>	0.397	0.774	1.000	<b>0.902</b>	0.994	0.975	0.529	<b>0.898</b>
Effect	$\delta = -.0933$	Reverse surrogacy (Null for $S$ )	0.902	0.050	0.370	1.000	0.051	0.308	0.977	0.050	0.377
on $I$	$\delta = -.1102$	Worsened survival	0.899	0.028	0.281	1.000	0.013	0.141	0.977	0.023	0.258



Figure 1

## Endpoints and Sample Size in the Setting of Out-of-Hospital Cardiac Arrest (25% improvement in outcome)

$$\text{Sample size } 2N \sim 4(Z_{\alpha/2} + Z_{\beta})^2 \frac{2p(1-p)}{(\lambda p)^2} \sim \frac{4 * 9 * 2}{.25^2} \frac{1-p}{p} = 1152 \frac{1-p}{p}$$

**ROSC**  $p \sim .6$   $2N \sim 768$

**Admit to ED**  $p \sim .4$   $2N \sim 1728$

**Admit to hospital**  $p \sim .3$   $2N \sim 2688$

**Discharge alive from hospital**  $p \sim .07$   $2N \sim 15305$

**Discharge alive from hospital neurologically intact**  $p \sim .05$   $2N \sim 21888$

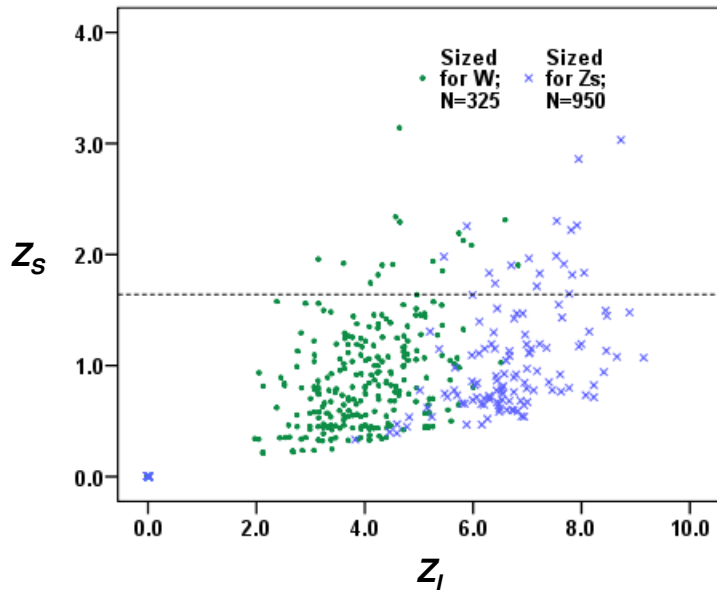
$2\alpha = .05$   
 $\beta = .1$   
 $\lambda = .25 (25\% \uparrow)$





Figure 2

Values of  $Z_I$  by  $Z_S$  when there is a large positive effect on the intermediate outcome, a modest negative effect on survival, and  $W$  is significant.



An approximately optimal statistic when  $Z_S \geq 0$  (i.e. no evidence of worsened survival) and  $Z_{SI} > 0$  (i.e. potential of super-surrogacy).

Under the null of no treatment effect on  $I$ ,  $Z_I = \sqrt{N}(p_T - p_C)/\sigma_I \sim N(0, 1)$  with  $\sigma_I^2 = 2\bar{P}\tilde{P}$  where  $\bar{P} = (P_T + P_C)/2$  and  $\tilde{P} = 1 - \bar{P}$ . Similarly under the null of no treatment effect conditional on  $I$ ,  $Z_{SI} = \sqrt{N}(q_T - q_C)/\sigma_{SI} \sim N(0, 1)$  where  $\sigma_{SI}^2 = \bar{Q}\tilde{Q}\left(\frac{1}{P_T} + \frac{1}{P_C}\right) = 2\bar{Q}\tilde{Q}\frac{\bar{P}}{P_T P_C}$ . Thus under the composite null,  $Z = Z(a, b) = aZ_I + bZ_{SI} \sim N(0, 1)$  provided  $a^2 + b^2 = 1$ . Under the alternatives,  $P_T = P_C + \Delta$  and  $Q_T = Q_C + \delta$ ,  $Z_I \sim N(\sqrt{N}\Delta/\sigma_I, 1)$  where  $\sigma_I \sim P_T\tilde{P}_T + P_C\tilde{P}_C$  and  $Z_{SI} \sim N(\sqrt{N}\delta/\sigma_{SI}, 1)$  where  $\sigma_{SI}^2 \sim \frac{Q_T\tilde{Q}_T}{P_T} + \frac{Q_C\tilde{Q}_C}{P_C}$  so, if the conditional outcome is independent of the intermediate outcome,  $Z \sim N\left(\left(a\Delta/\sigma_I + b\delta/\sigma_{SI}\right)\sqrt{N}, 1\right)$ . The null is rejected in favor of the alternative if  $Z > Z_{1-\alpha/2}$  so the power is  $\Phi(-Z_\beta) = \Phi\left(\left(a\Delta/\sigma_I + b\delta/\sigma_{SI}\right)\sqrt{N} - Z_{1-\alpha/2}\right)$ . This is maximized when  $Z_{SI} > 0$  with the ratio  $\frac{a}{b} = \frac{\Delta/\sigma_I}{\delta/\sigma_{SI}}$ . Estimating from the data,

$$\frac{a}{b} = \frac{\hat{\Delta}/\hat{\sigma}_I}{\hat{\delta}/\hat{\sigma}_{SI}} = \frac{\sqrt{N} \hat{\Delta}/\hat{\sigma}_I}{\sqrt{N} \hat{\delta}/\hat{\sigma}_{SI}} = \frac{Z_I}{Z_{SI}}$$

and substituting would give an approximately optimal statistic  $Z = aZ_I + bZ_{SI} \approx \frac{Z_I^2 + Z_{SI}^2}{\sqrt{Z_I^2 + Z_{SI}^2}}$ . Thus if  $Z_S \geq 0$  (i.e. no evidence of worsened survival) and

$Z_{SI} > 0$  (i.e. potential of super-surrogacy) an approximately optimal statistic is given by

$$W = \frac{\text{sign}(Z_I)Z_I^2 + Z_{SI}^2}{\sqrt{Z_I^2 + Z_{SI}^2}}.$$



## Appendix 2

S-plus program for the value for  $W$ ,  $W_c$ , corresponding to the  $C_{th}$  %tile across  
A wide range of intermediate (.1-.9) and conditional (.1-.9) outcome rates

```
nsim_100000
N_1000
C_.95
wc_vector(length=0)
for (Pc in .1*(1:9)){
  Pc_Pc
  for (Qc in .1*(1:9)){
    Qt_Qc
    x1_rbinom(nsim,N,Pc)
    x2_rbinom(nsim,N,Pc)
    mx2_x2/N
    mx1_x1/N
    rx_(x1+x2)/(2*N)
    sigx_(rx*(1-rx)/N)**.5
    y1_rbinom(nsim,x1,Qc)
    y2_rbinom(nsim,x2,Qt)
    my2_y2/x2
    my1_y1/x1
    ry_(y1+y2)/(x1+x2)
    sigy_(ry*(1-ry)*2/(x1+x2))**.5
    u_y1-y2
    x_((x1/N)-(x2/N))/(2**.5*sigx)
    y_((y1/x1)-(y2/x2))/(2**.5*sigy)
    delnull_mx2*my2/(mx1)-my2#estimates the conditional survival delta corresponding to null
    #survival effect
    EZrs_delnull*(((mx1+mx2)/2)*N)**.5/(((my1+my2)/2)*(1-((my1+my2)/2))**.5)
    CL_.6*EZrs
    w_(u>=0)*((y>=0)*(sign(x)*x**2+y**2)/((x**2+y**2+.00001)**.5)+(y<0)*(y>=CL)*x+
    (y<CL)*(x+3*y)))+(u<0)*0#computes W
    hist(w,nclass=200)
    wo_w[order(w)]
    wc_rbind(wc,c(Pc,Qc,wo[C*nsim]))
  }
}
wcregression_lm(wc[,3]~wc[,1]+wc[,2])
```



Appendix 3

S-plus program for comparing power of  $Z_i$ ,  $Z_s$ , and  $W$  for specific control rates and alternatives for the intermediate of null and substantial improvement and alternatives for the conditional survival corresponding to super-surrogacy, surrogacy, reverse-surrogacy and worsened survival

```

nsim_100000
N_250 #number per group
Pc_.4 #Control rate for intermediate outcome
Qc_.36 #Control rate for conditional survival
power_vector(length=0)
for (Pt in c(Pc,1.4*Pc)){#null and substantial improvement for intermediate outcome
for (Qt in c(Qc*1.2,Qc,Pc*Qc/Pt,.8*Pc*Qc/Pt)){#super-surrogacy, surrogacy, reverse-
surrogacy, #worsened survival
x1_rbinom(nsim,N,Pt)
x2_rbinom(nsim,N,Pc)
mx2_x2/N
mx1_x1/N
rx_(x1+x2)/(2*N)
sigx_(rx*(1-rx)/N)**.5
y1_rbinom(nsim,x1,Qt)
y2_rbinom(nsim,x2,Qc)
my2_y2/x2
my1_y1/x1
ry_(y1+y2)/(x1+x2)
sigy_(ry*(1-ry)*2/(x1+x2))**.5
ru_(y1+y2)/(2*N)
u_((y1-y2)/N)/(ru*(1-ru)*2/N)**.5
x_((x1/N)-(x2/N))/(2**.5*sigx)
y_((y1/x1)-(y2/x2))/(2**.5*sigy)
w95_1.956 + (-0.05067*Pc) + (.09672*Qc)#critical value for W-see appendix 1
delnull_mx2*my2/(mx1)-my2#estimates the conditional survival delta corresponding to null
#survival effect
EZrs_delnull*(((mx1+mx2)/2)*N)**.5/(((my1+my2)/2)*(1-((my1+my2)/2))**.5
CL_.6*EZrs
w_(u>=0)*((y>=0)*(sign(x)*x**2+y**2)/((x**2+y**2+.00001)**.5)+(y<0)*(y>=CL)*x+
(y<CL)*(x+3*y)))+(u<0)*0#computes W
pw_sum(w>=w95)/nsim
px_sum(x>=1.64)/nsim
pu_sum(u>=1.64)/nsim
power_rbind(power,c(px,pu,pw))
}}

```



#### Appendix 4

S-plus program for creating dependent outcome data and evaluating size

```
#converting mean, m, and stdev, s, into alpha and beta parameter
#for beta dist (NOTE s**2<m-m**2)!!!!!!
betapara_function(m,s)
  {if (s**2>=m-m**2) return("warning")
   if (s**2<m-m**2)
   {b_(((m/s)**2)*(1-m)**2/m)-(1-m)
   a_(m/(1-m))*b
   return(c(a,b))}}
```

```
getdata_function(nsims,Pc,Pt,Qc,Qt,N){
#for control
pab_betapara(Pc,.05)#returns the parameters of a beta dist with mean Pc and stdev .05
pc_rbeta(nsims*N,pab[1],pab[2])#generates individual probabilities for intermediate outcome
qab_betapara(Qc,.05)
qc_qbeta(pbeta(pc,pab[1],pab[2]),qab[1],qab[2])#generates individual probabilities for
#survival conditional on intermediate outcome with same %tile as the individuals %tile
#for probability of the intermediate outcome
xc_rbinom(nsims*N,1,pc)#generates the intermediate outcome
qc_qc*xc#sets the conditional survival probability to zero if the intermediate outcome is
not #reached
yc_rbinom(nsims*N,1,qc)#generates the survival outcome
#for intervention
pab_betapara(Pt,.05)
pt_rbeta(nsims*N,pab[1],pab[2])
qab_betapara(Qt,.05)
qt_qbeta(pbeta(pt,pab[1],pab[2]),qab[1],qab[2])
xt_rbinom(nsims*N,1,pt)
qt_qt*xt
yt_rbinom(nsims*N,1,qt)
return(rbind(xc,yc,xt,yt))}
```

```
#driver
nsim_100
N_1000
k_40#total number of simulations is k*nsim
out_vector(length=0)
for (i in 1:k){
power_vector(length=0)
for (Pc in .1*(1:9)){
for (Qc in .1*(1:9)){
data_getdata(nsim,Pc,Pc,Qc,Qc,N)
xc_data[1,]
xt_data[3,]
yc_data[2,]
yt_data[4,]
xc_matrix(xc,N,nsim)
xt_matrix(xt,N,nsim)
yc_matrix(yc,N,nsim)
yt_matrix(yt,N,nsim)
x1_colSums(xc)
x2_colSums(xt)
y1_colSums(yc)
y2_colSums(yt)
mx2_x2/N
mx1_x1/N
rx_(x1+x2)/(2*N)
sigx_(rx*(1-rx)/N)**.5
my2_y2/x2
```

```

my1_y1/x1
ry_(y1+y2)/(x1+x2)
sigy_(ry*(1-ry)*2/(x1+x2))**.5
ru_(y1+y2)/(2*N)
u_((y1-y2)/N)/(ru*(1-ru)*2/N)**.5#computes Zs
x_((x1/N)-(x2/N))/(2**.5*sigx)#computes Zi
y_((y1/x1)-(y2/x2))/(2**.5*sigy)#computes Zs|i
w95_1.956 + (-0.05067*Pc) + (.09672*Qc)#critical value for W-see appendix 1
delnull_mx2*my2/(mx1)-my2#estimates the conditional survival delta corresponding to null
#survival effect
EZrs_delnull*(((mx1+mx2)/2)*N)**.5/(((my1+my2)/2)*(1-((my1+my2)/2))**.5
CL_.6*EZrs
w_(u>=0)*((y>=0)*(sign(x)*x**2+y**2)/((x**2+y**2+.00001)**.5)+(y<0)*(y>=CL)*x+
(y<CL)*(x+3*y)))+(u<0)*0#computes W
pw_sum(w>=w95)/nsim
px_sum(x>=1.64)/nsim
pu_sum(u>=1.64)/nsim
power_rbind(power,c(Pc,Qc,px,pu,pw))
}}
out_cbind(out,power)}
#averages across the k runs and reassembles matrix
pq_out[,1:2]
pw_out[,5*(1:k)-0]
pu_out[,5*(1:k)-1]
px_out[,5*(1:k)-2]
pwa_rowMeans(pw)
pua_rowMeans(pu)
pxa_rowMeans(px)
size_cbind(pq,pxa,pua,pwa)
wusize_lm(size[,5]-size[,4]~size[,1]+size[,2])
wusize

```

