

2003

Simplified Meta-analysis of Clinical Trials in Resuscitation

Charles F. Babbs

Purdue University, babbs@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/bmepubs>



Part of the [Biomedical Engineering and Bioengineering Commons](#)

Recommended Citation

Babbs, Charles F, "Simplified Meta-analysis of Clinical Trials in Resuscitation" (2003). *Weldon School of Biomedical Engineering Faculty Publications*. Paper 59.

<http://docs.lib.purdue.edu/bmepubs/59>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Simplified Meta-analysis of Clinical Trials in Resuscitation

Charles F. Babbs, MD, PhD

Purdue University, West Lafayette, Indiana, USA,

(Resuscitation, 57, 245-255, 2003)

Abstract

Objective: To present and demonstrate a new simplified method for synthesizing results of multiple clinical trials in resuscitation research. **Methods:** The mean difference across studies in the proportion of favorable outcomes between experimental and control groups is calculated. This difference is shown to have a t-distribution. Its significance can be ascertained with a simple t-test. The analysis can be implemented in a one-page computer spreadsheet. **Results:** Simplified meta-analysis provides high sensitivity and can be extended to include weighting of studies according to size or quality, comparison of subgroups of studies, tests for outliers, and calculation of the power of the meta-analysis. Sample analyses are presented for two experimental forms of cardiopulmonary resuscitation: interposed abdominal compression CPR and active compression-decompression CPR. **Conclusions:** Traditional narrative reviews, taking note of the proportion of individual studies with statistically significant results, can lead to erroneous conclusions and unnecessary delays in the clinical use of research findings. Simplified meta-analysis can provide rapid, quantitative, and accurate estimates of the amount of benefit or harm from an experimental intervention and can further empower physicians to practice evidence-based medicine.

Keywords: Abdomen, Active compression-decompression CPR, Clinical trials, Guidelines, Interposed abdominal compression-CPR, Methodology, Statistical analysis

1. Introduction

A key aspect of evidence-based medicine is the careful analysis of research literature concerning the effectiveness of a proposed novel treatment or procedure. Traditional methods of research synthesis involve tabulating the available studies that are related to a particular treatment or innovation and noting the number of significant positive studies versus the number of non-significant ones. This procedure has been termed “vote counting”, in which each study casts a vote for or against the innovation under consideration¹⁻³. The proportion of positive studies required to justify a change in clinical practice may vary among reviewers and regulatory bodies. In the usual collective review a positive recommendation is made when a majority, or more, of trials are statistically significant. If there are equal numbers of statistically significant and non-significant clinical studies, many reviewers will conclude the vote to be a draw and recommend that more research is needed³.

The vote counting approach to research review is now known to lead systematically to high percentages of erroneous conclusions¹⁻³. In particular, vote counting produces Type II statistical errors or false negative evaluations—that is, erroneously accepting the null hypothesis when the results of treatment are quite genuine^{2,4}. Conventional testing for statistical significance minimizes the probability of Type I errors or false positive evaluations—that is, erroneously concluding there is a real treatment effect when the results are due to sampling variability. However, consideration of conventional statistical significance does not protect against Type II errors.

Meta-analysis, defined as the quantitative synthesis of data from multiple clinical studies, minimizes both Type I and Type II errors. This approach is gaining popularity as an alternative to vote counting⁵⁻¹¹. The present paper introduces a simple and user-friendly form of meta-analysis for use by groups of physicians and others seeking to set evidence-based practice guidelines. It is an extension and simplification of the “observed minus expected” (O-E) approach of Yusuf and Peto^{5,12}, combined with the concept of cumulative meta-analysis introduced by Lau⁶. This simplified approach to meta-analysis has a number of advantages for systematic overviews of clinical trials.

1. The method is derived easily from fundamental principles and can be understood, verified, and trusted by physicians with knowledge of elementary statistics as taught in universities and medical schools. It is designed to demystify meta-analysis and to place control of the technical aspects of the process directly in the hands of clinical decision-makers. These decision makers can then combine clinical judgment with statistical inference to arrive at the best possible decisions about the timely adoption of new methods reported in the research literature.
2. It is easily performed. All that is needed is a basic spreadsheet program such as Microsoft Excel. Given a spreadsheet template, one can perform the technical computations for a meta-analysis about as easily as making a table of results from a literature survey. Special purpose software is not required.

3. It allows for a variety of optional calculations. These include the number needed to treat (or harm), the weighting of studies according to their size, quality, or category, a procedure for detecting outliers, calculation of the power of the meta-analysis, and easy comparison of subgroups of studies.

Here the method is described and illustrated with reviews of two newer techniques in cardiopulmonary resuscitation.

2. Methods

Suppose that one reviews a series of m randomized clinical trials and that in each trial an experimental group is compared with a control group and survival is the principle outcome measure. The proposed method focuses upon the differences in the proportions of survivors (or other favorable outcomes) in each study, which is denoted Δp . To say “the experimental treatment increased survival by 15 percent,” is to say $\Delta p = 0.15$. Any dichotomous outcome measure, indicating either benefit or harm, may be used in the place of survival. These might include the occurrence of tumor response, death, stroke, a significant cardiac event, the development of lung cancer, graft patency after one year, etc. If some studies measure harm (e.g. mortality) rather than benefit (e.g. survival), it is a simple matter to recast results in terms of the proportions of patients with favorable outcomes, so that the direction of the results is consistent for all studies.

2.1 Testing the mean difference in outcome

Let p_1 be the proportion of favorable outcomes in a control group of n_1 patients. Let p_2 be the proportion of favorable outcomes in an experimental group of n_2 patients. Let the corresponding values $p_{1i}, p_{2i}, n_{1i}, n_{2i}$ represent results of the i -th study in a series of m studies to be synthesized, and let $\Delta p_i = p_{2i} - p_{1i}$ be the difference in the proportions of favorable outcomes observed in the i -th study. Definitions of these and related variables used in the analysis are summarized in Table 1.

Table 1. Notation

Variables	Definitions
m	Number of studies in a series of studies to be analyzed
n	Number of patients in a study group
p	Measured proportion of survivors in a control group or an experimental group
Δp	Difference in proportions of survivors (experimental – control)
s(X)	Standard deviation of random variable, X
$s^2(X)$	Variance of random variable, X
w	Optional weighting factor for a particular study
W	Sum of weighting factors for all studies in a series to be analyzed
Subscripts	
1	Control group
2	Experimental group
i, j	Any single study in series of studies

The p_{1i} 's, p_{2i} 's, and Δp_i 's are random variables that depend upon the true probabilities of favorable outcome in the various experimental and control groups and also upon sampling variability. Suppose that each trial were repeated a large number of times and the null hypothesis were true. That is, there is no real treatment effect. Then for each study, i , the difference in the proportions of favorable outcomes, $\Delta p_i = p_{1i} - p_{2i}$, between experimental and control groups would have zero mean and variance best estimated as

$$s^2(\Delta p_i) = \frac{p_{1i}(1-p_{1i})}{n_{1i}-1} + \frac{p_{2i}(1-p_{2i})}{n_{2i}-1}. \quad (1)$$

Expression (1) is based upon the principle that the variance of a sum or difference of independent random variables equals the sum of the variances¹³. The control and experimental results are independent, because they come from different patients. The variance estimates of the form $\frac{p(1-p)}{n-1}$ for the measured proportions of survivors in the control and experimental groups are unbiased, as shown in Appendix 1. All data needed to evaluate expression (1) are readily obtained from published results.

Assume initially that the studies to be synthesized are given equal weights. (Use of unequal weights, based for example upon study quality or size, is discussed in the next

section, 2.2.) The test statistic to be used in meta-analysis is the mean difference in proportions for the m studies, namely

$$\overline{\Delta p} = \frac{\Delta p_1 + \Delta p_2 + \cdots + \Delta p_m}{m}. \quad (2)$$

Under the null hypothesis the expected value of $\overline{\Delta p}$ is zero. Again, since the variance of a sum or difference of such independent random variables is the sum of the variances¹³, the estimated variance of $\overline{\Delta p}$ is

$$s^2(\overline{\Delta p}) = \frac{s^2(\Delta p_1) + s^2(\Delta p_2) + \cdots + s^2(\Delta p_m)}{m^2}. \quad (3)$$

Thus $s^2(\overline{\Delta p})$ is easily calculated from the p_1, p_2, n_1, n_2 data for the various studies. Taking the square root to obtain $s(\overline{\Delta p})$, we can compute the test statistic,

$$t = \frac{\overline{\Delta p}}{s(\overline{\Delta p})}. \quad (4)$$

One can show, along the lines of Welch¹⁴, that this test statistic is distributed very much like a ‘‘Student’’ t-distribution with a number of degrees of freedom roughly equal to twice the total number of survivors in all of the studies to be synthesized*, that is, a t-distribution with $df \approx 2 \sum_{i=1}^m p_{1i} n_{1i} + p_{2i} n_{2i}$. Note that a t-distribution with more than about 50 degrees of freedom is equivalent to the normal distribution. Thus for all most meta-analyses one can simply use the normal distribution.

To construct 95% confidence intervals for $\overline{\Delta p}$, let $t_{97.5}$ be the 97.5th percentile of the t-distribution with df degrees of freedom, as found in standard tables or functions. Typically this value is close to 2.0. For the normal distribution $t_{97.5} = 1.96$. The 95 percent confidence interval for $\overline{\Delta p}$ is $\overline{\Delta p} \pm s(\overline{\Delta p}) \cdot t_{97.5}$. If this confidence interval includes zero, one cannot reject the null hypothesis that $\overline{\Delta p} = 0$ at the $P = 0.05$ level. If the confidence interval does not include zero, there is a significant positive or negative effect of treatment across studies.

* A slightly better approximation can be obtained using the formula $df = 2S(1+3(S/N))$, where S is the total number of survivors in all studies and N is the total number of patients in all studies. If a majority of patients in all studies survive, then one should replace the number of survivors with the number of non-survivors.

2.2 Optional features of the analysis

Weighting of studies. In expression (2) $\overline{\Delta p}$ is an un-weighted average, to which all studies contribute equally, in keeping with the recommendation of Peto, Collins and coworkers.^{12, 15} If one wishes to deal with varying size or quality of the studies by introducing weights so that some studies contribute more to $\overline{\Delta p}$ than others, then for a weighted average with individual study weights w_i and sum of study weights W ,

$$\overline{\Delta p} = \frac{w_1 \Delta p_1 + w_2 \Delta p_2 + \dots + w_m \Delta p_m}{W}, \quad (5)$$

$$s^2(\overline{\Delta p}) = \frac{w_1^2 s^2(\Delta p_1) + w_2^2 s^2(\Delta p_2) + \dots + w_m^2 s^2(\Delta p_m)}{W^2}, \quad (6)$$

and the 95% confidence interval for $\overline{\Delta p}$ is $\pm s(\overline{\Delta p}) \cdot t_{97.5}$, as before.

Weights may be assigned to reflect the sizes (n 's) of various studies, such that a large study counts for more than a smaller one. One such approach^{1, 16-18} uses inverse variance estimates as a weighting factors. In this case weights are calculated as $w_i = 1/s^2(\Delta p_i)$. The inverse variance values are larger for larger and better-controlled studies. This type of weighting is especially recommended when combining several initial small trials with subsequent larger trials. Inverse variance weights produce minimum variance of the overall weighted average^{19, 20}. Subjective ratings of study quality can also be used as weights. However, the value of reviewers' impressionistic ratings of the quality or worthiness of studies is controversial^{21, 22}.

The mathematics of the present method accommodates both users who argue that unequal weighting of studies is arbitrary and capricious and also those who argue that equal weighting ignores obvious differences in study quality and size. For reviewers of the first persuasion all weights are simply set to 1.0. For reviewers of the second persuasion each individual study is weighted according to predetermined criteria. The use of inverse variance weighting is a popular compromise strategy and is rapidly becoming standard practice in medical meta-analyses. Typically, the main results of the meta-analysis will be insensitive to easily tested changes in the weighting method, providing evidence that the weights were not chosen to produce a particular result.

The inclusion of weighting factors in the mathematics and in spreadsheets for performing the analysis also has practical advantages. An examination of the influence of various individual studies upon the outcome of the meta-analysis can be done by the simple expedient of setting the weight of a particular study to zero. If the significance of the overall analysis changes greatly, then a single study is driving the results. By extension, analysis of sub-groups of studies can be performed by using non-zero weights for particular studies of interest and zero weights for all other studies.

Subgroup comparison. Suppose one finds among the studies in a meta-analysis two apparent subgroups of studies that differ in treatment effect, possibly on the basis differences in patient populations, treatment implementation, or clinical setting. To test whether there is there a significant difference in the treatment effect between the subgroups one can re-do the meta-analysis twice—first setting the weights for subgroup 2 to zero, leaving subgroup 1, and then setting the weights of subgroup 1 to zero, leaving subgroup 2. If the 95% confidence intervals do not overlap and just touch at one common point, then the results for one subgroup will be greater than the results for the other subgroup at least $(0.975)^2 = 95.0625$ percent of the time. In this sense one can say that if the 95 percent confidence intervals of $\overline{\Delta p}$ for mutually exclusive subgroups of studies do not overlap, then the subgroups are significantly different from each other. In many cases of apparent differences among studies, the 95 percent confidence intervals will overlap^{1,3}. Then one can avoid needless speculation over differences consistent with sampling variation alone. If the 95 percent confidence interval for a subgroup of $m=1$ study is clearly separated from that for the remaining $m-1$ studies, then the study may be unidentified as an outlier.

Number needed to treat (or harm). An omnibus number needed to treat can be calculated as $\frac{1}{\Delta p}$. This statistic, advocated by McQuay and Moore²³, represents the average number of patients that must be given an experimental treatment to obtain one additional survivor (or other good outcome). It is useful in evaluating the cost effectiveness of proposed innovations.

Power of the meta-analysis. Under the null hypothesis the mean value of the distribution of $t = \frac{\overline{\Delta p}}{s(\overline{\Delta p})}$ is zero, and the standard deviation of t is very close to 1.00, unless the aggregate number of survivors is very small. In this case one can also calculate the power of the meta-analysis for the alternative hypothesis that the true effect is at least, say, a 10 percent difference in the proportion of survivors. Then, for the alternative hypothesis $\overline{\Delta p}_1 = 0.10$, we have $t_1 = \frac{0.10}{s(\overline{\Delta p})}$. The statistical power of the meta-analysis is the probability of making a correct positive evaluation of the experimental treatment, assuming the alternative hypothesis is true. Using $P = 0.05$ for significance testing, the power is

$$F(t_1 - 1.96), \quad (7)$$

where F is the cumulative probability density function for the standard normal distribution.* The usefulness of computing the statistical power of a meta-analysis has been emphasized recently by Hedges and Pigott²⁴.

* In Microsoft Excel this function is named NORMSDIST().

2.3 Spreadsheets for performing meta-analysis

A standard spreadsheet program such as Microsoft Excel is sufficient to perform a simplified meta-analysis. In one compact format, each study is represented in a column of the spreadsheet. Successive steps in the calculations are performed in successive rows of the spreadsheet. Top rows include the study author, the study weight, the number of favorable outcomes and total number of patients in control and experimental groups. Subsequent rows include control and experimental group proportions, the Δp_i for each study, and its variance estimate (1). The next rows include the intermediate variables $w_i \Delta p_i$, $w_i^2 \Delta p_i$ and $w_i^2 s^2 (\Delta p_i)$ for weighted study calculations. Then $\overline{\Delta p}$ and its 95% confidence limits are calculated using expressions (4), (5) and (6) together with the t-distribution and inverse t-distribution functions. Finally the number needed to treat and power of the meta-analysis are calculated. Excel's SUMIF(weight > 0) and COUNTIF(weight > 0) functions can be utilized across columns to implement a new cumulative meta-analysis after the appearance of each successive study. Then each successive column represents a cumulative meta-analysis at one point in time. For inverse variance weighting of the studies, the weights are assigned values equal to $1/s^2 (\Delta p_i)$.

3. Results—sample analyses

Sample meta-analyses were performed on studies relating to two experimental techniques of cardiopulmonary resuscitation (CPR). The first is the addition of interposed abdominal compression (IAC) to otherwise standard CPR²⁵⁻²⁷. The second is CPR with active compression and decompression (ACD) of the chest²⁸⁻³⁰. Reviews of these techniques are found in references^{27, 31} and³². During IAC-CPR positive pressure is applied to the abdomen in counterpoint to the rhythm of chest compression, so that the abdomen is being compressed when chest pressure is relaxed. During ACD-CPR positive and negative pressures are applied alternately to the chest by means of a “plunger” that forms a seal with the anterior chest wall. Both methods improve hemodynamics in animal studies of electrically induced ventricular fibrillation^{33, 34}. Both improve CO₂ excretion as a measure of effective systemic perfusion in human resuscitations^{35, 28, 36}.

Relevant full length, peer reviewed publications were identified using evidence evaluation worksheets created by the research working group of the American Heart Association^{37, 38}. Individual trials were obtained from MEDLINE searches, the author's files, and reference lists of review articles on newer techniques in resuscitation as referenced in³⁹. The end points analyzed include short-term survival and long-term survival. Short-term survival is defined as return of spontaneous circulation (ROSC). Long-term survival is defined as hospital discharge with intact neurological function.

Results of simplified cumulative meta-analysis for IAC-CPR are shown in Table 2 and Figure 1. Results of simplified cumulative meta-analysis for ACD-CPR are shown in

Table 3 and Figure 2. Tables 2 and 3 were copied from a working spreadsheet, a template for which can be obtained from the author at no cost and modified by the user for similar meta-analyses. The top rows in Tables 2 and 3 show raw data abstracted from individual studies. The last five rows of Tables 2 and 3 show data for a cumulative meta-analysis. In these bottom rows successive columns from left to right represent successive stages of the cumulative meta-analysis. The mean $\overline{\Delta p}$ -value under Study 1 describes the first study only. The mean $\overline{\Delta p}$ -value under Study 2 describes the combined results of the first two studies. The mean $\overline{\Delta p}$ -value under Study 3 describes the combined results of the first three studies, etc. The cumulative nature of the data in the next four rows is similar.

Table 2. Cumulative meta-analysis for ROSC and discharge survival data in studies of IAC-CPR. Here studies are given inverse variance weights, $1/s^2(\Delta p)$

	ROSC				Discharge	
<i>Raw data</i>						
Last Study ID	Mateer	Ward	Sack 1	Sack 2	Ward	Sack 1
Study Weight	349.93	40.47	116.25	154.96	256.00	255.75
Study #	1	2	3	4	2.1	3.1
# Cont Alive	45	3	14	21	0	3
n_1	146	17	55	76	17	55
# Exp Alive	40	6	29	33	1	8
n_2	145	16	48	67	16	48
<i>Proportions</i>						
p_1	0.3082	0.1765	0.2545	0.2763	0.0000	0.0545
p_2	0.2759	0.3750	0.6042	0.4925	0.0625	0.1667
Δp	-0.0324	0.1985	0.3496	0.2162	0.0625	0.1121
$s^2(\Delta p)$	0.0029	0.0247	0.0086	0.0065	0.0039	0.0039
<i>Cumulative data</i>						
$\overline{\Delta p}$	-0.0324	0.0831	0.1719	0.1830	0.0625	0.0873
SE $\overline{\Delta p}$	0.0535	0.0506	0.0444	0.0389	0.0625	0.0442
2-tail P(t)	0.5458	0.8680	0.0983	0.0062	0.4226	0.0599
NNT	-30.9	12.0	5.8	5.5	16.0	11.5
Power	0.46	0.51	0.61	0.73	---	0.62

Table 3. Cumulative meta-analysis for ROSC data in studies of ACD-CPR using inverse variance weights, $1/s^2(\Delta p)$

Last Study ID	C93N	T94J	L94J	S95J1	S95J2	L96J	S96J1	S96J2	S96R	P97C
<i>Raw Data</i>										
Study Weight	67.56	55.32	131.79	416.61	949.51	56.36	878.88	1761.87	221.72	557.88
Study #	1	2	3	4	5	6	7	8	9	10
Cont Alive	9	9	24	27	65	12	129	84	68	77
n_1	30	28	77	136	310	30	386	510	114	258
Exp Alive	20	15	24	20	56	10	140	91	54	114
n_2	32	25	53	117	297	26	405	501	106	254
<i>Proportions</i>										
p_1	0.3000	0.3214	0.3117	0.1985	0.2097	0.4000	0.3342	0.1647	0.5965	0.2984
p_2	0.6250	0.6000	0.4528	0.1709	0.1886	0.3846	0.3457	0.1816	0.5094	0.4488
Δp	0.3250	0.2786	0.1411	-0.0276	-0.0211	-0.0154	0.0115	0.0169	-0.0871	0.1504
$s^2(\Delta p)$	0.01480	0.01808	0.00759	0.00240	0.00105	0.01774	0.00114	0.00057	0.00451	0.00179
<i>Cumulative Data</i>										
$\overline{\Delta p}$	0.3250	0.3041	0.2198	0.0663	0.0151	0.0140	0.0132	0.0147	0.0097	0.0251
SE $\overline{\Delta p}$	0.1217	0.0902	0.0627	0.0386	0.0248	0.0244	0.0198	0.0152	0.0148	0.0140
2-tail P(t)	0.0098	0.0011	0.0006	0.0878	0.5447	0.5657	0.5061	0.3344	0.5123	0.0732
NNT	3.1	3.3	4.6	15.1	66.4	71.2	76.0	68.0	102.8	39.8
Power	0.13	0.20	0.36	0.74	0.98	0.98	1.00	1.00	1.00	1.00

In Table 2 the four data columns to the left represent a meta-analysis of short-term survival. The two columns to the right represent a separate meta-analysis of long-term, discharge survival. Studies of IAC-CPR showed a significant treatment effect for both short-term and long-term survival. $\overline{\Delta p}$ is 10.7 percent for ROSC, 8.7 percent for hospital discharge. The effect of IAC is highly significant for ROSC data ($P = 0.006$) and of borderline statistical significance ($P = 0.06$) for discharge survival data. These summary values can be immediately converted into the number-needed-to-treat, $NNT = 1/\overline{\Delta p}$, the number of patients that must be treated to obtain one additional survivor²³. For return of spontaneous circulation the NNT is 9. For discharge survival, the corresponding NNT is 12. The last row in Table 2 demonstrates that the process of cumulative meta-analysis gradually increases the statistical power for detecting a true positive treatment effect. Here power was calculated using expression (7) and assuming a true positive treatment effect of $\overline{\Delta p} = 10\%$. Increasing statistical power reduces the

probability of a false negative evaluation or Type II error, which is equal to one minus the power.

Figure 1(A) shows a cumulative meta-analysis plot for IAC-CPR to demonstrate historical trends with the publication of each successive study. The top data point and its 95 percent confidence interval represent the historically first trial, the next a combination of the first two trials, the third a combination of the first three, etc. These data points correspond to the successive meta-analyses of ROSC data in the columns of Table 2 from left to right. Statistical significance typically emerges after publication of just a few studies to reach a stable value. Thereafter, the addition of further studies merely narrows the 95 percent confidence interval.

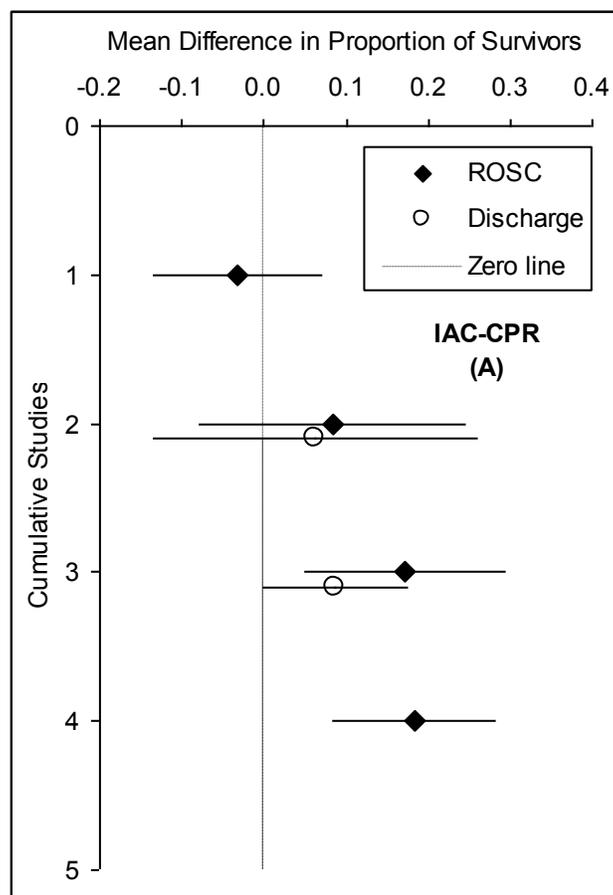


Figure 1(A)

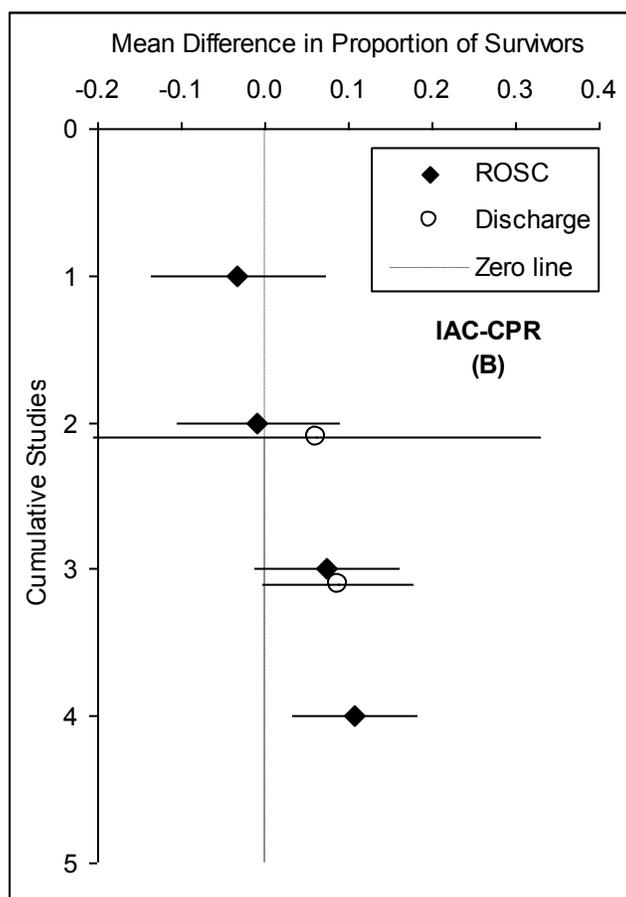


Figure 1(B)

FIGURE 1. Simplified cumulative meta-analysis for IAC-CPR. Successive points from top to bottom represent a separate meta-analysis after the appearance of each study in the series. (A) Equal weighting: each study is assigned weight 1.0. (B) Inverse variance weighting: each study is assigned weight, $1/s^2(\Delta p)$. The major conclusions of the meta-analysis are independent of the weighting method.

The cumulative meta-analysis shown in Figure 1(A) was conducted with equal study weights in keeping with the recommendation of Peto¹². The corresponding cumulative meta-analysis in Figure 1(B) was conducted with inverse variance weights in keeping with the recommendation of Hedges^{1,17}. The raw data for Figures 1(A) and 1(B) were identical. For dichotomous variables such as survival, individual study variances are highly dependent upon group size, as indicated in Equation (1). As expected, inverse variance weights diminish the impact upon the overall results of Study 2, which had small patient numbers. However, the major conclusions of the meta-analysis of IAC-CPR are insensitive to the particular weighting method used.

Critics often object to one study or another on technical grounds. Because the number of studies in the IAC series is small, the question arises "would the analysis be significant if the offending study were left out?" Table 4 shows results of meta-analysis of ROSC data for IAC-CPR, in terms of 95 percent the confidence limits for $\overline{\Delta p}$ when each of the 4 studies in turn is given zero weight in the analysis. A zero weight has the effect of excluding one particular study and can be assigned very easily in a spreadsheet format by changing the weight value to zero. Here the results, for inverse variance weighting, show that at least one of the studies by Sack, each of which included different patients, is needed to conclude that IAC-CPR produces more frequent ROSC than standard CPR.

Table 4. Sensitivity testing: effects of omitting single studies upon 95% confidence intervals for $\overline{\Delta p}$; ROSC data for IAC-CPR versus standard CPR

Study omitted	Lower 95% limit	Upper 95% limit
None	0.031	0.184
Mateer	0.152	0.375
Ward	0.022	0.180
Sack 1	-0.029	0.140
Sack 2	-0.014	0.162

At least one of the studies by Sack, which included different patients at different institutions, is needed to conclude that IAC-CPR produces more frequent ROSC than standard CPR.

Retrospectively, there does appear to be a difference in the benefit of IAC-CPR between the three in-hospital studies and the one pre-hospital study. A separate meta-analysis of the three inverse variance weighted in-hospital studies (numbers 2, 3, and 4) for ROSC data gives $\overline{\Delta p} = 26$ percent with a 95% confidence interval from 15 to 38 percent. This range does not overlap the 95% confidence interval of the one pre-hospital study, which ranged from -14 to +7 percent. (In the pre-hospital study patients assigned to IAC-CPR necessarily received some standard CPR from bystanders and also during transport to the hospital, perhaps diluting the effect of IAC.) Thus for in-hospital resuscitations, one can conclude from the three studies that IAC-CPR roughly doubles rates of initial resuscitation from about 25 to about 50 percent. For the two in-hospital studies of long-term survival with IAC-CPR also appeared to double from about 8 to 16 percent. However, the total number of patients studied for long-term survival is small and the result is of borderline statistical significance ($P = 0.06$).

A simplified meta-analysis of studies of ACD-CPR using equal weights for each study showed a statistically significant treatment effect for ROSC, with $\overline{\Delta p}$ equal to 7.7 percent (data not shown). However, when inverse variance weighting was used to emphasize the

relative importance of larger, better-controlled studies, $\overline{\Delta p}$ for ROSC was reduced to 2.5 percent. The 95% confidence interval for $\overline{\Delta p}$ was -0.002 to 0.053 , just including zero. For hospital discharge $\overline{\Delta p}$ was less than 0.01 and clearly not significant. Figure 2 shows a cumulative meta-analysis plot for ACD-CPR with inverse variance weights. Here the pattern is unusual in that early studies show a much greater effect than do later ones, with a trend toward zero until publication of the very last study, which was significantly positive individually.

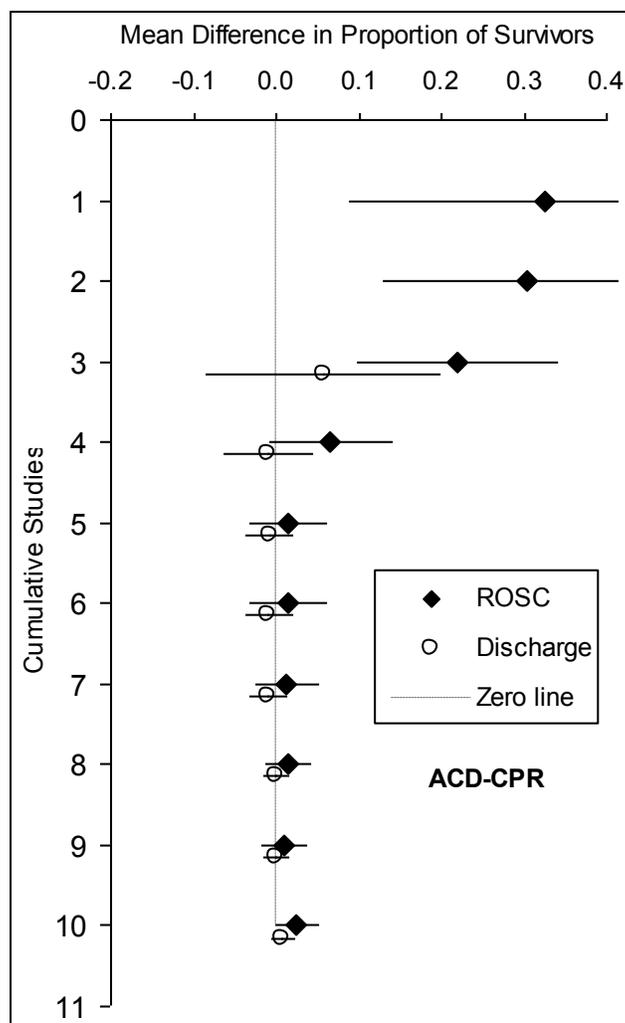


FIGURE 2. Simplified cumulative meta-analysis for ACD-CPR using inverse variance weighting.

Evidently, the studies of ACD-CPR were not all cast from the same mold. Inspection of the Δp values reveals one group of studies indicating a significant positive benefit (studies 1, 2, 3, and 10) and a second group having zero benefit upon survival (studies 4-9). The 95 percent confidence interval for $\overline{\Delta p}$ in the first group of studies is 0.10 to 0.24. The 95 percent confidence interval for $\overline{\Delta p}$ in the second group of studies is -0.03 to +0.03. Although this subgroup analysis was performed retrospectively with knowledge of the results, the impression for the ROSC data is unmistakably bimodal and not that expected for a homogeneous population. Those close to the research suggest rigorous training in the technique, which is physically more difficult to perform than standard CPR⁴⁰, is required for good results. Such training was clearly described in the last positive trial, and was probably done in the first few positive trials. This effect cannot be demonstrated for long-term survival following ACD-CPR, however, which is uniformly similar to that for standard CPR.

4. Discussion

Research is producing growing numbers of important innovations in healthcare generally, and in the field of resuscitation in particular. However, there are often years of delay between achievement of positive results in valid research studies and their subsequent implementation in routine clinical practice^{6, 41, 42}. One reason for delayed implementation of research findings is the excessive processing and analysis that evidence requires before application⁴¹. Busy individual clinicians have limited time to pour over seemingly conflicting data. Panels of experts meet for at most a few days to consider multiple practice guidelines in rapid succession with no time scheduled for seeking answers to questions that arise. Overly complex statistical analysis is simply avoided by such panels, and conservative conclusions are drawn, often using an explicit or implicit vote-counting approach that is biased toward Type II errors¹⁰.

Simpler, user-friendly techniques of research synthesis are needed to bridge the gap between research and practice. Simpler techniques of meta-analysis would empower more physicians to practice evidence-based medicine, especially in conjunction with the increased availability of the necessary raw data on the Internet. Such techniques implemented on laptop computers with standard software would allow guideline-writing groups to address a variety of "what if" questions in the process of deliberation, such as "what if this particular study were left out" or "what if certain high quality studies counted more".

This paper presents an approach to meta-analysis of a series of two-group clinical trials that is straightforward both in theory and in execution. The simplification of technical mathematics, which leads to a t-test of the overall significance of included studies, makes the approach personally verifiable by physician analysts. The actual calculations require no more than a one-page spreadsheet. One can create or borrow, and validate, a sample spreadsheet quickly. Subsequent meta-analyses can be done easily by replacing the raw

data in the sample spreadsheet with raw data from other studies and expanding or contracting the table appropriately. Special purpose software is not needed.

Assumptions of the method are few. One is that the variance of a sum of independent or uncorrelated random variables equals the sum of the variances¹³. This assumption is reasonable since the component Δp values are derived from independent studies. (It is important to confirm that the same patients do not reappear in more than one study, as can happen with a preliminary report and a final report of an ongoing trial.) The use of a t-distribution for expression (4) requires the further assumption that the average difference in proportions will be normally distributed. This assumption is reasonable, since the component binomial distributions approach the normal distribution in shape for group $n's > 20$ (true for all but the very smallest clinical trials)¹³. As the number m of studies that are combined increases, the central limit theorem¹³ further strengthens this assumption. Finally, the assumption that a normally distributed random variable, divided by its estimated standard deviation has a Student's t distribution, as originally described by Welch¹⁴, is well preceded in statistics and can be confirmed for proportional data by numerical examples.

Meta-analysis in terms of the proportional differences, Δp_i , is especially significant for those assumptions that are not present. Just as in the summed observed minus expected method of Yusuf, Peto, and coworkers⁵, there is no assumption of homogeneity. The test statistic $\overline{\Delta p}$ is treated simply a weighted sum of random variables. Hence expressions (2) through (6) do not require that the Δp_i 's be similar in magnitude.

The studies may vary in terms of patient populations, drug doses, procedural skills, hospital settings, exclusion criteria, etc. It is not necessary to assume that the trials synthesized are exactly comparable—only that they test the same basic intervention as it might be implemented in various settings in the real world. The Δp_i 's need not even measure proportions of exactly the same things. One study could measure proportion of 24-hour survivors and another study could measure the proportion of 48-hour survivors after the same treatment. A reviewer might ask the question: “is there evidence of treatment effect upon long term survival?” The meta-analysis combining these studies is perfectly valid as long as “long term” is clearly defined as 24 hours or greater. Similarly, if some studies measured mortality and others measured the occurrence of stroke or myocardial infarction, they could be combined, as above, in a meta-analysis of clearly defined “major adverse outcomes” following standard vs. experimental therapy.

For those accustomed to the high false negative results of traditional literature reviews and research synthesis^{1,3}, meta-analysis may seem to be a rush to judgment, provoking one or more now classical objections^{3,21}. Each objection is a variation on the common theme that “you shouldn't combine apples and oranges”. One such concern is the challenge of dealing with multiple end points recorded in individual studies, such as long-term survival (apples) and short-term survival (oranges). Strictly speaking, ROSC and discharge survival data from the same studies of experimental CPR cannot be combined in the same simplified meta-analysis because they are not independent. A conservative

and now technically easy remedy, as shown in Figures 1 and 2, is to perform a separate analysis for each type of end-point, even though the number of studies measuring a less popular or hard to obtain end point, like discharge survival, may be smaller.

A second objection relates to variable methodological quality of included studies, although subjective quality ratings do not necessarily correlate with the size of treatment effects^{3, 21, 22}. Critics argue that less worthy studies (apples) should not be compared with more worthy studies (oranges). This issue can also be dealt with empirically. Should subjective quality ratings be available and agreed to, one can perform a separate meta-analysis (subgroup comparison) for studies of high and low quality by alternately setting the weights of low and high quality studies to zero. If the results differ, one can rely on the “strong” studies; if they do not, then all studies should be included³.

A third objection to meta-analysis relates to variable research methods among studies with different patient populations, exclusion criteria, data recording methods, hospital settings, or other kinds of “apples and oranges”. Again this objection can be answered empirically. One may simply perform a separate meta-analysis for each class of studies—that is, perform a subgroup comparison. If the 95% confidence intervals for the Δp values for two subgroups do not overlap, then one has statistical evidence for a case of apples and oranges. Otherwise one only has evidence for normal sampling variation. Thus, all three classical objections to meta-analysis can be dealt with by doing a separate meta-analysis of the apples versus the oranges. Heated philosophical debates can be attenuated by taking a closer look at the data.

Earlier workers in the field of meta-analysis anticipated several aspects of the approach presented in this paper. The method of Rosenthal and Rubin⁴³ began with reported p-values as inputs and later¹⁶ included differences in proportions. The present approach represents a fresh look at the problem from the standpoint of the physician. In this context meta-analysis is a tool that must be used in conjunction with clinical judgment. In the first phase of evaluation of a new intervention one is simply trying to determine if a proposed treatment, in general, produces an effect on survival or some other favorable outcome¹². If it does, then the effect seen in the selected trials is likely to generalize to the even broader range of circumstances found in widespread practice. As more data become available a second phase of meta-analysis may be done to determine if a particular intervention produces different outcomes in particular patient groups. The present method can be used to help answer both of these questions. Simplified meta-analysis provides a tool to distill relevant data, but, of course, it cannot substitute for sound clinical judgment in dealing with the sensitive issues of life and death and quality of life.

Conclusion

Meta-analysis allows the reviewer to deal with the situation of multiple small studies having promising effects but small n's and somewhat varying methodology, owing to limitations of cost, time, patient recruitment, or the inherent difficulty of resuscitation research. Formal meta-analysis is immune from the systematic and excessive Type II errors associated with the traditional "vote-counting" methods of research synthesis^{2,4}. Such analysis can guide individual and institutional practice and shorten the time between medical research discoveries and their clinical implementation. The method presented here makes meta-analysis accessible to any physician.

References

1. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Orlando: Academic Press, 1985.
2. Hedges LV, Olkin I. Vote-counting methods in research synthesis. *Psychological Bulletin* 1980; 88:359-369.
3. Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, California: Sage Publications, 1990:592.
4. Babbs CF. Consensus evidence evaluation in resuscitation research: analysis of Type I and Type II errors. *Resuscitation* 2001; 51:193-205.
5. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases* 1985; 27:335-371.
6. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; 327:248-54.
7. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Treatments for myocardial infarction*. *JAMA* 1992; 268:240-8.
8. Nichol G, Stiell IG, Laupacis A, Pham B, De Maio VJ, Wells GA. A cumulative meta-analysis of the effectiveness of defibrillator-capable emergency medical services for victims of out-of-hospital cardiac arrest. *Ann Emerg Med* 1999; 34:517-25.
9. Thacker SB. Meta-analysis. A quantitative approach to research integration. *JAMA* 1988; 259:1685-9.
10. Roper WL, Thacker SB. Doing good before there's harm. *Ann N Y Acad Sci* 1993; 703:33-9; discussion 39-40.
11. Kornbluth AA, Salomon P, Sacks HS, Mitty R, Janowitz HD. Meta-analysis of the effectiveness of current drug therapy of ulcerative colitis. *J Clin Gastroenterol* 1993; 16:215-8.
12. Peto R. Why do we need systematic overviews of clinical trials? *Statistics in Medicine* 1987; 6:233-240.

13. Mosteller F, Rourke REK. *Sturdy Statistics: Nonparametrics and Order Statistics*. Reading Massachusetts: Addison-Wesley, 1973:395 pages.
14. Welch BL. The significance of the difference between two means when the population variances are unequal. *Biometrika* 1937; 29:350-362.
15. Collins R, Gray R, Godwin J, Peto R. Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. *Statistics in Medicine* 1987; 6:245-250.
16. Rosenthal R. Comparing effect sizes of independent studies. *Psychological bulletin* 1982; 92:500-504.
17. Hedges LV, Gurevitch J, Curtis PS. The meta-analysis of response ratios in experimental ecology. *Ecology* 1999; 80:1150-1156.
18. Meier P. Variance of a weighted mean. *Biometrics* 1953; 9:59-73.
19. Hedges LV. Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 1982; 92:490-499.
20. Hedges LV. Combining independent estimators in research synthesis. *British Journal of Mathematical and Statistical Psychology* 1983; 36:123-131.
21. Wolfe FM. *Meta-Analysis: Quantitative Methods for Research Synthesis*. In: Lewis-Beck MS, ed. *Quantitative Applications in the Social Sciences*. Vol. 59. Newbury Park: Sage Publications, 1986:65 pages.
22. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990; 11:339-52.
23. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 1997; 126:712-20.
24. Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychological Methods* 2001; 6:203-217.
25. Ralston SH, Babbs CF, Niebauer MJ. Cardiopulmonary resuscitation with interposed abdominal compression in dogs. *Anesthesia and Analgesia* 1982; 61:645-651.
26. Sack JB, Kesselbrenner MB, Bregman D. Survival from in-hospital cardiac arrest with interposed abdominal counterpulsation during cardiopulmonary resuscitation. *JAMA* 1992; 267:379-385.
27. Sack JB, Kesselbrenner MB. Hemodynamics, survival benefits, and complications of interposed abdominal compression during cardiopulmonary resuscitation. *Acad Emerg Med* 1994; 1:490-497.
28. Cohen TJ, Tucker KJ, Lurie KG, et al. Active compression-decompression. A new method of cardiopulmonary resuscitation. *Cardiopulmonary Resuscitation Working Group. JAMA* 1992; 267:2916-23.
29. Lurie KG, Shultz JJ, Callahan ML, et al. Evaluation of active compression-decompression CPR in victims of out-of-hospital cardiac arrest. *JAMA* 1994; 271:1405-11.
30. Plaisance P, Adnet F, Vicaut E, et al. Benefit of active compression-decompression cardiopulmonary resuscitation as a prehospital advanced cardiac life support. A randomized multicenter study. *Circulation* 1997; 95:955-61.
31. Babbs CF, Tacker WA. Cardiopulmonary resuscitation with interposed abdominal compression. *Circulation* 1986; 74(suppl IV):37-41.

32. Wik L, Mauer D, Robertson C. The first European pre-hospital active compression-decompression (ACD) cardiopulmonary resuscitation workshop: a report and a review of ACD- CPR. *Resuscitation* 1995; 30:191-202.
33. Einagle V, Bertrand F, Wise RA, Roussos C, Madger S. Interposed abdominal compressions and carotid blood flow during cardiopulmonary resuscitation: support for a thoracoabdominal unit. *Chest* 1988; 93:1206-1212.
34. Lindner KH, Ahnefeld FW, Bowdler IM. Cardiopulmonary resuscitation with interposed abdominal compression after asphyxial or fibrillatory cardiac arrest in pigs. *Anesthesiology* 1990; 72:675-681.
35. Ward KR, Sullivan RJ, Zelenak RR, Summer WR. A comparison of interposed abdominal compression CPR and standard CPR by monitoring end-tidal PCO₂. *Ann Emerg Med* 1989; 18:831-7.
36. Orliaguet GA, Carli PA, Rozenberg A, Janniere D, Sauval P, Delpech P. End-tidal carbon dioxide during out-of-hospital cardiac arrest resuscitation: comparison of active compression-decompression and standard CPR. *Ann Emerg Med* 1995; 25:48-51.
37. Cummins RO, Chamberlain D, Hazinski MF, et al. Recommended guidelines for reviewing, reporting, and conducting research on in-hospital resuscitation: the in-hospital "Utstein style". *Circulation* 1997; 95:2213-2239.
38. Cummins RO. Introduction to the international guidelines 2000 for CPR and ECC. *Circulation* 2000; 102 (suppl I):I-1--I-11.
39. Cummins RO. American Heart Association in collaboration with the International Liaison Committee on Resuscitation. Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care: international consensus on science. *Circulation* 2000; 102 (suppl I):I-1--I-384.
40. Wik L, Schneider T, Baubin M, et al. Active compression-decompression cardiopulmonary resuscitation-- instructor and student manual for teaching and training. Part II: A student and instructor manual. *Resuscitation* 1996; 32:206-12.
41. Haynes RB, Hayward RS, Lomas J. Bridges between health care research evidence and clinical practice. *J Am Med Assoc* 1995; 273:342-50.
42. Haynes RB. Some problems in applying evidence in clinical practice. *Ann N Y Acad Sci* 1993; 703:210-24; discussion 224-5.
43. Rosenthal R, Rubin DB. Comparing significance levels of independent studies. *Psychological Bulletin* 1979; 86:1165-1168.

Appendix 1: Estimating the variance of a binomial distribution from measured proportions.

Let p be the measured proportion of survivors among n treated patients, for whom the true probability of survival is π . The mean (expected value) of the binomial distribution of p -values is $E(p) = \pi$ and the variance, $V(p)$, of the binomial distribution is $V(p) = \pi(1-\pi)/n$. It is well known that the maximum likelihood estimate of π is p ¹³. In turn, the maximum likelihood estimate of V is $\hat{V}(p) = p(1-p)/n$. This expression provides an initial estimate of the binomial distribution variance.

Let us check the expected value of $\hat{V}(p) = p(1-p)/n$ to see if it is biased compared to the true variance $V(p)$.

$$E(\hat{V}) = \frac{E(p)}{n} - \frac{E(p^2)}{n} \quad \text{A1.1}$$

By definition,

$$V(p) = E((p - \pi)^2) = E(p^2 - 2p\pi + \pi^2) = E(p^2) - 2\pi E(p) + \pi^2 = E(p^2) - \pi^2 \quad \text{A1.2}$$

Combining the above,

$$E(\hat{V}(p)) = \frac{\pi}{n} - \frac{V(p) + \pi^2}{n} = \frac{\pi(1-\pi)}{n} - \frac{V(p)}{n} \quad \text{A1.3}$$

But $\pi(1-\pi)/n = V(p)$, so

$$E(\hat{V}(p)) = V(p) \left(1 - \frac{1}{n}\right) = V(p) \left(\frac{n-1}{n}\right) \quad \text{A1.4}$$

The initial estimate is biased by a factor of $(n-1)/n$. However, if we multiply this estimate by a factor of $n/(n-1)$ the bias is removed. Hence,

$$\hat{V}'(p) = \frac{p(1-p)}{n} \cdot \frac{n}{n-1} = \frac{p(1-p)}{n-1}$$

is unbiased, as can be confirmed by repeating steps A1.1 through A1.4 for the new estimate.