2004

Meta-analysis of Two-Treatment Clinical Trials Including Both Continuous and Dichotomous Results

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

Meta-analysis of Two-Treatment Clinical Trials Including Both Continuous and Dichotomous Results

CHARLES F. BABBS, MD, PhD

Purdue University, West Lafayette, Indiana, USA 47907

(Medical Decision Making 24, 299-312, 2004)

ABSTRACT

To expedite the timely creation of medical practice guidelines, a meta-analytic method was developed to combine of both dichotomous survival data and continuous physiologic data from multiple studies comparing the same innovative clinical intervention to standard care. The method is adapted for synthesis of small, early studies of novel treatments. An aggregate ratio, R*, of the observed treatment effect to a clinically optimal treatment effect for studies in a series is computed, and compared to the 95% confidence limit for R* under the null hypothesis. Weights assigned to each study may reflect its precision, quality, or clinical relevance. Input data for continuous variables include sample means, standard errors, and sample sizes. Input data for dichotomous variables include group proportions and sizes. The analysis can be done using a simple, one-page spreadsheet. It allows one to judge biological significance, to test for statistical significance, to compare subgroups of studies for differences (heterogeneity of effect), to test for outliers, and to compute the power of the meta-analysis. These features are demonstrated for studies of interposed abdominal compression—cardiopulmonary resuscitation (IAC-CPR). This novel method of meta-analysis can provide rapid, quantitative, and accurate estimates of the amount of benefit or harm from an experimental clinical intervention, as reported in multiple small independent studies of differing experimental design.

Key words (not in title): abdominal, cardiopulmonary resuscitation, confidence intervals, continuous, data interpretation, evidence-based medicine, interposed abdominal compression (IAC)-CPR, orphan drugs, pediatric research

Abbreviations: CPR, cardiopulmonary resuscitation, IAC, interposed abdominal compression, ROSC, return of spontaneous circulation

INTRODUCTION

The fruits of research are sometimes like apples and oranges, i.e. they are not directly comparable. This paper presents a general approach to the systematic analysis of two-treatment experiments including a wide variety of study designs and end-points. Such heterogeneity is typical of early research on an emerging therapy or concept, before experimental methods become standardized. The approach was designed specifically as an aid to the evaluation of evidence during development of guidelines for CPR and advanced cardiac life support. In this field multiple large clinical trials of new interventions are rare or nonexistent—owing to the lack of funding, the difficulty in obtaining informed consent from persons in cardiac arrest, the low probability of survival from sudden cardiac death, the perception that the existence of standards obviates the need for research, and the chaotic nature of the clinical setting during cardiopulmonary resuscitation. Studies of innovative techniques in CPR include small numbers of patients. Some focus on survival; others focus on physiologic end-points such as blood pressure. Similarly, research synthesis relating to rare or orphan diseases, including many pediatric conditions, often involves a paucity of heterogenous data. In these cases there is an insufficient number of large controlled clinical trials to justify a formal Cochrane review[1]. Yet patients must be treated anyway, and hence there will always be a need for clinical guidelines based upon the best evidence available[*].

An important early question in analysis of a new treatment or intervention in such data poor areas of research is "does the new method work any better than the standard approach?" An important related question is "does the new method work better in some subgroups of patients than in others?" Quantitative statistical methods such as meta-analysis can help answer these questions[4]. The novel meta-analytic method presented herein is applicable to any studies that compare a new clinical intervention (drug, dose, or device) with a standard or control intervention in diverse animal or clinical models. Outcome measures may differ greatly. Some may be continuous data (e.g. physiologic measures like end-tidal CO_2 concentration or mean coronary perfusion pressure). Others may be dichotomous data (e.g. integer head counts of patients resuscitated, discharged, etc.). It is only necessary that all of the studies relate to a common focused question or test the same hypothesis and that the results of the different measures are generally consistent. (If the patients die with "good numbers" for continuous physiologic data, then, obviously, the survival data and the "numbers" are inconsistent, and caution is necessary in combining them.)

---

[*] The author has been personally involved in guideline creation in orphan research areas as Chair of the Research Working Group, Emergency Cardiovascular Care (ECC) Programs, American Heart Association for the years 2000 to 2002. This committee was charged with developing methods of evidence evaluation used in creation of international guidelines for cardiopulmonary resuscitation and emergency cardiovascular care[2] The present work was inspired by the challenges faced by the ECC committees, which had to make the most efficient possible use of available data. The ECC committees began with traditional vote counting procedures, in which each positive significant study “casts a vote” for, and each non-significant study “casts a vote” against, a proposed innovation. This approach is, of course, strongly biased toward the conclusion that the experimental treatment has no effect. Hence vote counting methods are no longer recommended for the task of research synthesis[3-5]. The present paper is an outgrowth of the author's search for a better alternative.

Existing statistical methods are not satisfactory for this task. A simple sign test of the numbers of studies having any observed positive effect of treatment, versus those with any observed negative effect of treatment can be done. However, this approach ignores information about the magnitudes of treatment effects. Conventional statistical tests using pooled data such as a t-test or analysis of variance assume homogeneity of variance, which is unlikely to be true in multiple small trials. Accordingly, such tests are usually deemed inappropriate for the purpose of combining results in a meta-analysis[5]. Conventional meta-analysis using effect size for continuous variables[5] will not work, without modification, for dichotomous data such as the proportion of survivors. Conventional meta-analysis using odds ratios[6, 7] will not work for physiologic data like blood pressure or blood flow. None of these methods deals formally with the issue of biological significance.

The method meta-analysis, described herein, is an extension of the response ratio method developed by Hedges and coworkers for synthesis of continuous data from studies in the field of ecology[8]. It permits assessment of the statistical and biological significance of the combined results from multiple small studies, which may include a mixture of continuous and dichotomous end-points. It is straightforward and easily implemented on ordinary personal computers using standard spreadsheet software.

METHODS

Terminology.

Definitions of symbols used to describe input data for the analysis are given in Table 1. The symbols $\overline{E}$ and $\overline{C}$ refer to sample means of continuous data, such as end-tidal CO_2, from experimental and control groups, respectively. The symbols p_E and p_C refer to proportions obtained from experimental and control groups, such as the proportions of surviving patients. True population variances for a random variable, X, are indicated by the symbol $\sigma^2(X)$. Estimates of population variances, obtained from measured data, are indicated by the addition of a "hat" symbol, $\hat{\sigma}^2(X)$. Standard errors of the mean are indicated by a combined symbol, for example $\hat{\sigma}(\overline{E})$. For continuous data $\hat{\sigma}(\overline{E})$ and $\hat{\sigma}(\overline{C})$ are taken as the published values for the standard errors of the mean reported by the investigators for treatment and control groups. If these values are not provided, they can be calculated using formulas in Table 2. For dichotomous data $\hat{\sigma}(p_E)$ and $\hat{\sigma}(p_C)$ may be computed from the expression for the variance of the binomial distribution, $\hat{\sigma}^2(p) = p(1-p)/(n-1)$. Division by n-1 rather than by n provides an unbiased estimate for the variance[9]. For simplicity of notation, "hats" are not used for sample means or proportions, which are understood to differ from the corresponding population values.

Table 1. Nomenclature

Symbol	Definition for continuous data	Definition for dichotomous data
$\overline{E}$	Sample mean for treatment group	
$\overline{C}$	Sample mean for control group	
p_E		Proportion of individuals with favorable outcome in the treatment group
p_C		Proportion of individuals with favorable outcome in the control group
I	Ideal result for a clinical study, the best possible outcome	Ideal result for a clinical study, the best possible outcome, typically 1.0 or 100 % survival
n_E	Number of subjects in experimental group	Number of subjects in experimental group
n_C	Number of subjects in control group	Number of subjects in control group
$\hat{\sigma}(\overline{E})$ or $\hat{\sigma}(p_E)$	Published standard error of the mean for the experimental group (experimental group standard deviation divided by $\sqrt{n_E}$)	$\sqrt{p_E(1-p_E)/(n_E-1)}$, estimate of the standard deviation of observed proportions in multiple replications of the same study
$\hat{\sigma}(\overline{C})$ or $\hat{\sigma}(p_C)$	Published standard error of the mean for the control group (control group standard deviation divided by $\sqrt{n_C}$)	$\sqrt{p_C(1-p_C)/(n_C-1)}$, estimate of the standard deviation of observed proportions in multiple replications of the same study
R	$R = \frac{\overline{E}-\overline{C}}{I-\overline{C}} = 1-\frac{I-\overline{E}}{I-\overline{C}} \equiv 1-R'$	$R = \frac{p_E-p_C}{I-p_C} = 1-\frac{I-p_E}{I-p_C} \equiv 1-R'$
λ	$\lambda \equiv \ln(R') = \ln\left(\frac{I-\overline{E}}{I-\overline{C}}\right)$	$\lambda \equiv \ln(R') = \ln\left(\frac{I-p_E}{I-p_C}\right)$

Summarizing Research Findings as Clinical Result Ratios.

Clinical result ratios for two-group experiments are computed as shown in Table 2. They are obtained by dividing the measured difference in outcome between experimental and control groups by the clinically optimal difference in outcome for any particular end point. The clinically optimal or ideal outcome, denoted I, would usually represent return to normal physiologic status (for example, normal cardiac output or normal arterial blood oxygen content) or 100 percent survival in the case of proportions. Hence, the clinical result ratio is either

$$R = \left(\overline{E} - \overline{C}\right)/\left(I - \overline{C}\right) \qquad (1a)$$

in a study with continuous physiologic end points or

$$R = \left(p_E - p_C\right)/\left(I - p_C\right) \qquad (1b)$$

in a study of discrete endpoints such as survival. The result ratio, R, for each individual study can be interpreted as a fraction of the best possible treatment effect that could have been found. By convention outcomes are described as positive, such that a larger value indicates benefit. In turn, a ratio $R > 0$ indicates greater benefit in the experimental group than in the control group.* $R = 1$ indicates ideal benefit. $R = 0$ indicates no benefit. $R < 0$ indicates a worse result than control. Zero values for $\overline{E}$, $\overline{C}$, p_E, or p_C are allowed. Note that the ideal outcome, I, in expression (1a) is almost surely greater than the control group outcome, $\overline{C}$; otherwise the study would not have been done. In expression (1b) the value of I is typically 1.0 or 100 percent survival, which usually indicates the maximum possible clinical benefit. However, it is not strictly true that the ideal survival is always 100 percent. For example, in long-term studies the value of ideal outcome, I, could be taken as the predicted survival of healthy persons over the same time period. (Although a common convention in statistics is to use upper case letters to denote random variables and lower case letters to denote constants, in the present context I is a constant.) Using the ideal result, I, as a point of reference allows one to combine clinical response ratios for measures that have quite different scales and dynamic ranges. Since the relative effect of treatment across studies is more constant than the absolute effect[11-13], use of such response ratios is an advantage in meta-analysis.

* In epidemiology[10], outcomes are often reported as negative, e.g. mortality. Such results are easily translated into survival. It is an interesting mathematical diversion to show that if I=100 percent for survival and I'=0 percent for mortality, then R values computed directly from mortality data are the same as those computed from survival data. Although experts may disagree somewhat on the definition of "ideal", small changes in the choice of the parameter, I, do not influence the results of significance testing, as detailed in the Discussion section.

Table 2. Computation of result ratios for continuous and dichotomous variables for an individual study

	Definition for continuous data	**Definition for dichotomous data**
Clinical Response Ratio	$R = \frac{\bar{E} - \bar{C}}{I - \bar{C}} = 1 - \frac{I - \bar{E}}{I - \bar{C}} \equiv 1 - R'$	$R = \frac{p_E - p_C}{I - p_C} = 1 - \frac{I - p_E}{I - p_C} \equiv 1 - R'$
Log complement of response ratio	$\lambda = \ln(R') = \ln\left[\frac{I - \bar{E}}{I - \bar{C}}\right]$	$\lambda = \ln(R') = \ln\left[\frac{I - p_E}{I - p_C}\right]$
Variance of the log complement of response ratio*	$\hat{\sigma}^2(\lambda) = 2\frac{(n_E - 1)\,\hat{\sigma}^2(\bar{E}) + (n_C - 1)\,\hat{\sigma}^2(\bar{C})}{(N - 2)(I - \bar{P})^2}$	$\hat{\sigma}^2(\lambda) = \frac{\bar{p}(1 - \bar{p})}{(n_C - 1)(I - \bar{p})^2} + \frac{\bar{p}(1 - \bar{p})}{(n_E - 1)(I - \bar{p})^2}$

*For continuous data the estimates of the experimental and control group variances, namely

$$\hat{\sigma}^2(\bar{E}) = \frac{1}{(n_E - 1)}\sum_{i=1}^{n_E}(E_i - \bar{E})^2 \quad \text{and} \quad \hat{\sigma}^2(\bar{C}) = \frac{1}{(n_C - 1)}\sum_{i=1}^{n_c}(C_i - \bar{C})^2$$

are usually provided in the reports of studies to be synthesized. If not they must be calculated from the original data.

One practical issue in the computation of response ratios deserves mention here. For the purpose of meta-analysis it is useful to use a single composite figure of merit to describe the treatment effect in each particular study in a series. However, in many studies several different outcome measures are reported (primary and secondary end-points), which may have somewhat different ratios of experimental to control results. There are two approaches to this situation[14]. The first approach is to select a "best" end-point from each study that is most relevant to the question in hand. This may well be the primary end point identified by the original investigators. To avoid throwing away information in a data poor research area, however, it may be helpful to derive a composite measure of treatment effect for each such study. Hence a second approach is to compute an average composite R-value to represent the overall result of each study. (The general solution to this problem is to use a weighted average of outcome measures within each independent study, based on a predetermined framework defining the importance of the different measures reported. The first approach, just described, is the special case in which the "best"

outcome measure is given weight 1.00 and the all other measures are given weight zero. The second approach requires treating the various within-study outcomes as correlated measures, for which the standard deviation of the average is the average standard deviation.)

The result of this first phase of data abstraction is to create a single composite outcome measure, R, for each independent study. These individual indices of effect can then be combined into an aggregate index for the entire series of studies.

The Aggregate Response Ratio and its Distribution Under H_0.

The objective of this section is to develop a test statistic from the combined response ratios from multiple studies that has a known sampling distribution under the null hypothesis, H_0. Then significance testing for a meta-analysis will be possible. First we shall find well behaved test statistics describing each individual study, and then we shall consider a weighted average of these statistics across all studies in the meta-analysis.

Statistics for individual studies

Suppose one has assembled a series of m independent studies of a particular treatment and has computed the clinical result ratio, R, for each study. To obtain a test statistic with a known distribution under H_0 , it is helpful to express result the ratio R for any particular study in terms of its complement, R' = 1 - R, that is,

$$R = \frac{\overline{E} - \overline{C}}{I - \overline{C}} = \frac{I - \overline{C} - \left(I - \overline{E}\right)}{I - \overline{C}} = 1 - \frac{I - \overline{E}}{I - \overline{C}} \equiv 1 - R' \qquad (1c)$$

or

$$R = 1 - \frac{I - p_E}{I - p_C} \equiv 1 - R' \qquad (1d)$$

and then to determine the distribution of R' values under the null hypothesis. Because the distribution of ratios of random variables is skewed, it is standard practice to work with the logarithms of ratios, which have a more symmetrical distribution that is better approximated by a normal distribution[15, 16]. (Of course, to work with logarithms we must have R' > 0.) Under the null hypothesis, when the expected values of the numerator and denominator of R' are equal, and the coefficients of variation of the numerator and denominator of R' are also equal, then the natural log of R' is very well approximated by a normal distribution having zero mean and having variance given in Appendix 1 (A1.1). As shown in Appendix 1, the variance of log R' can be determined directly, whereas the variance of R cannot, because the numerator contains the difference of two random variables. The distribution of log R' is also very close to normal under H_0 (Appendix 1(b)). Moreover, the Central Limit Theorem provides a further argument for approximate normality of the average logarithm of R' across studies, which will be used for significance testing in the meta-analysis.

Let the natural logarithm of $R' = 1 - R$ be denoted by the Greek letter, λ. Thus,

$$\lambda \equiv \ln(R') = \ln\left(\frac{I - \bar{E}}{I - \bar{C}}\right) \text{ for continuous data, or} \qquad (2a)$$

$$\lambda \equiv \ln(R') = \ln\left(\frac{I - p_E}{I - p_C}\right) \qquad (2b)$$

for dichotomous data in any particular study. Under the null hypothesis, H_0, the mean value of the distribution of $\lambda = \ln(R')$ is zero ($\ln(1) = 0$), and the estimated standard deviation is $\hat{\sigma}(\lambda)$, as defined in Table 2 and derived in Appendix 2. Because the distribution of λ is symmetrical and approximately Gaussian, one can readily compute the distribution of $R = 1 - R' = 1 - e^{\lambda}$ and its 95 percent confidence limits.

An aggregate test statistic for meta-analysis

Now to conduct a meta-analysis of studies related to a common experimental intervention, one may define a pooled test statistic for a series of m independent studies as

$$R^* = 1 - e^{\bar{\lambda}}, \qquad (3)$$

where $\bar{\lambda}$ is a weighted average value of $\lambda=\ln(R')$ for the series, namely,

$$\bar{\lambda} = \frac{1}{W}\sum_{j=1}^{m} w_j \lambda_j \qquad (4)$$

with individual study weights w_j, and W equal to the sum of the weights. The virtue of using R^* for meta-analysis is that confidence limits can be obtained for this statistic under H_0, based upon the known, approximately normal distributions of the random variables λ_j.

The weights can be assigned with due caution and good judgment[17, 18] to reflect the relevance or the quality of the studies. Generally, precise guidelines should be drawn up in advance. One option is to set all weights equal to 1, in which case $W = m$. In this case all studies that are deemed relevant to the question at hand are given weight 1 , and all studies that are deemed irrelevant are given weight zero. Another popular option is the use of inverse variance weighting, which gives greater weight to experiments whose estimates have greater statistical precision. Inverse variance weights produce minimum variance of the overall weighted average[19, 20].

A formal process of assigning weights makes the relative contribution of each study explicit. In addition, the explicit weights in expression (4) make it technically easy to repeat a meta-analysis with each study in turn omitted. One simply sets the weight of each study temporarily to zero and observes the effect upon the results. Such an exercise helps to determine if any one study

drives the final conclusions of the meta-analysis. The use of weights also expedites the comparison of subgroups of studies, as subsequently explained.

To test R* for statistical significance and to make cumulative meta-analysis plots we wish to find 95 percent confidence limits associated with R* when the null hypothesis is true. This task is easily done as follows. The estimated variance of $\bar{\lambda}$ is

$$S^2 \equiv \hat{\sigma}^2(\bar{\lambda}) = \frac{1}{W^2}\left[w_1^2\hat{\sigma}^2(\lambda_1) + w_2^2\hat{\sigma}^2(\lambda_2) + \cdots + w_m^2\hat{\sigma}^2(\lambda_m)\right]. \quad (5)$$

Expression (5) derives from the general principle that the variance of the sum of independent random variables is equal to the sum of the variances. Here the component λ-values are obviously independent, since they come from different studies, each of which must include different subjects. The variance estimates for the λ-values from each study are computed from measured data as shown in Table 2, the expressions for which are derived in Appendix 2.

If we assume that under the null hypothesis $\bar{\lambda}$ has a normal distribution with mean value zero and standard deviation S, as is reasonable**, then $\frac{\bar{\lambda}}{S}$ has a standard normal distribution. In turn, the 95 percent confidence interval under the null hypothesis for $\bar{\lambda}$ is $0 \pm 1.96S$. Hence, the lower and upper critical values for significance testing of R* with $\alpha = 0.05$ are

$$C_L = \left(1 - e^{+1.96\,S}\right) \text{ to } C_U = \left(1 - e^{-1.96\,S}\right). \quad (6)$$

By calculating the empirical R* value and comparing it to the critical values $\left(1 - e^{\pm 1.96\,S}\right)$, one obtains a rapid test of significance for biologically meaningful effects of treatment in the entire series of m studies. (Note that the confidence interval for the mean log R' value is symmetrical, but the back-transformed confidence interval for R* is not.)

Alternatively, one may calculate the classical p-value for the two-sided test H_0: $\lambda = 0$ vs. H_1 $\lambda \neq 0$ as $p = 2\int_{|\bar{\lambda}|/S}^{\infty} f(x)\,dx$, two-tailed, where f(x) is the probability density function for the normal distribution.

** A virtue of the log transformation is that the log result ratios are approximately normally distributed. Further, by virtue of the central limit theorem[21], an average of several log result ratios is even better approximated by a normal distribution. Because $S = \hat{\sigma}(\bar{\lambda})$ is actually an estimate derived from sample data, and hence a random variable, the distribution of $\bar{\lambda}/S$ will resemble a t-distribution. However, one can show, along the lines of Welch[22], that $\bar{\lambda}/S$ is distributed very much like a t-distribution with a number of degrees of freedom roughly equal to the number of patients in all studies synthesized. Thus for a typical meta-analysis assumption of a normal distribution is very reasonable.

Power of the Meta-analysis.

Under H_0 the mean value of the distribution of $\bar{\lambda}$ is zero (ln(1) = 0), and the standard deviation is S, computed from expression (5). Because the distribution of $\bar{\lambda}$ is normal, one can also calculate the power of the meta-analysis for an alternative hypothesis, H_1, that the true effect is at least, say, 10 percent of the ideal benefit, and assuming that S is the same under these circumstances, which is very nearly true. (Under H_1 there is a slight skewness to the distribution of λ , which can be ignored in power calculations.) Then, for R = 0.1 the mean value of λ is given by $R = 0.1 = 1 - e^{\bar{\lambda}'}$ or $\bar{\lambda}' = \ln(0.9) = -0.105$. (Note $\bar{\lambda}'$ is negative for a positive treatment effect.) In this case the beta error, or probability of a false negative evaluation of the experimental treatment, is

$$\beta = \int_{-\infty}^{-(\bar{\lambda}'/S + 1.96)} f(x)dx \,, \qquad (7)$$

where f(x) is the probability density function for the normal distribution, and $\bar{\lambda}' < 0$. The power of the meta-analysis is 1 - β. The usefulness of computing the statistical power of a meta-analysis has been emphasized recently by Hedges and Pigott[23].

Cumulative Meta-analysis.

Using the graphical approach similar to Lau and coworkers[7, 24] one can construct a cumulative meta-analysis plot showing successive values of $R^* = 1 - e^{\bar{\lambda}}$ in relation to the critical values for rejection of the null hypothesis after publication of each study in the series being analyzed. Such a plot shows how the significance of the overall treatment effect evolves with the publication of new data.

Omitting Individual Studies.

Critics will often object to one study or another on technical grounds, raising the question as to whether the entire analysis is flawed because an offending study has been included. To explore the influence of individual studies on the conclusions of the analysis one can make a table of the aggregate R^* statistics and their confidence limits, first when no study, and then when each study in turn is omitted from the analysis by setting its weight, w, equal to zero. If the results do not differ substantially (the usual outcome) then one can conclude that no single study drives the conclusions of the analysis.

Comparing Subgroups of Studies for Heterogeneous Effects.

Suppose one finds among the studies in a meta-analysis two apparent subgroups of studies that differ in treatment effect, possibly on the basis of differences in patient populations, treatment implementation, or hospital 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 for subgroup 1 to zero, leaving subgroup 2. The difference in mean log(R'), namely $d = \bar{\lambda}_1 - \bar{\lambda}_2$, between subgroup 1 and subgroup 2 can be tested for statistical significance. Since, the subgroups are independent, the variance of the difference is the sum of the variances, $S_1^2 + S_2^2 = \hat{\sigma}^2(\bar{\lambda}_1) + \hat{\sigma}^2(\bar{\lambda}_2)$, which are automatically available from the separate meta-analyses of the subgroups. The expected standard deviation of subgroup differences is $S_d = \sqrt{S_1^2 + S_2^2}$. In turn, the 95 % confidence interval for d is $0 \pm 1.96 S_d$ under the null hypothesis that d = 0. If the observed d lies within this confidence interval, there is no significant difference between the subgroups. Such a "d-test" can avoid needless speculation over differences explained merely by sampling variation.

Test for Outliers.

If the difference, d, between a subgroup of m=1 study and the remaining m-1 studies is clearly significant, for example, if the absolute value of d is greater than, say, 3 times S_d, then the study may be unidentified as an outlier.

Spreadsheets for Performing Meta-analysis.

A standard spreadsheet program such as Microsoft Excel is sufficient to perform a meta-analysis of clinical response ratios. No macros or programming are needed. Costly special purpose software is not required. There are two general phases of calculation. The first is obtaining the R- and λ-statistics for each study. The second is running the meta-analysis itself. These phases can be performed in different sections of the spreadsheet, as shown in Tables 3 and 4. These tables were copied directly form a working spreadsheet, a template for which can be obtained from the author at no cost[**] and modified for similar meta-analyses of various topics.

SAMPLE RESULTS FOR IAC-CPR

Sources.

During interposed abdominal compression (IAC)-CPR manual pressure is applied to the abdomen of the victim 180 degrees out of phase with the rhythm of chest compression, so that the abdomen is being compressed when chest pressure is relaxed, and vise versa. This technique has been studied in a variety of mathematical, mechanical, animal, and clinical models with generally positive results[25]. For simplicity only human studies are included in the present meta-analysis. These relatively small, initial studies gathered a mixture of continuous physiologic data and dichotomous survival data, which are well suited for analysis in terms of clinical response ratios. Full length, peer reviewed publications were identified using evidence evaluation

[**] E-mail: babbs@purdue.edu

worksheets created by the research working group of the American Heart Association[26]. Individual trials were obtained from MEDLINE searches, the author's files, and reference lists of review articles on newer techniques in resuscitation as referenced in[2]. All studies compared patients receiving IAC-CPR with those receiving standard CPR. The end points analyzed include blood pressure, end tidal CO_2, frequency of return of spontaneous circulation, and frequency of hospital discharge with intact neurological function. Altogether there are 7 separate results reported in the 5 studies. Criteria for quality and relevance of the studies are defined in terms of "Levels of Evidence", developed by the American Heart Association and described in detail previously[26]. The four Level 1 studies are weighted 1, and the one Level 2 study is weighted 1/2. The reduced weight for the Level 2 study is a conservative choice, since this initial non-randomized trial had strongly positive results.

Composite λ values for individual studies.

Results of data reduction for the 5 available clinical studies of IAC-CPR are shown in Table 3. Successive studies are arranged by columns from left to right. The column for the fourth study (Sack, 1992) is subdivided to accommodate multiple dichotomous end points. Data from each study are entered in two blocks—the upper block for continuous data and the lower block for dichotomous data. In Table 3 continuous data include mean arterial pressure and end-tidal CO_2, which is reflective of forward blood flow. Control and experimental means and their respective standard errors appear in successive rows. Next the experimental/control result ratio, R, and the more normally distributed logarithm of 1-R, denoted λ, are computed, together with the variance estimate for λ using the formula shown in Table 2 and Appendix 2 (A2.5a).

Table 3. Raw and composite outcome data from independent human studies of IAC-CPR*

Study ID	Berryman	Mateer	Ward	Sack 1**		Sack 2
Date	1984	1985	1989	1992		1992
Continuous Data						
End point	MAP		ET-CO2			
Con mean	26		9.6			
Con SEM	1		1			
Exp mean	39		17.1			
Exp SEM	1.6		1.5			
Con N	6		33			
Exp N	6		33			
R	0.1884		0.3676			
R'	0.812		0.632			
Grand Mean	32.5		13.35			
$\lambda = \ln(R')$	-0.209		-0.458			
$V(\lambda)$	0.000911		0.0117			
Dichotomous Data						
End point		ROSC	ROSC	ROSC	Discharge	ROSC
Con survivors		45	3	14	3	21
Con N		146	17	55	55	76
Exp survivors		40	6	29	8	33
Exp N		145	16	48	48	67
Con p		0.3082	0.1765	0.2545	0.0545	0.2763
Exp p		0.2758	0.375	0.6041	0.1666	0.4925
R'		1.04677	0.7589	0.531	0.88141	0.70122
R		-0.04677	0.2411	0.469	0.11859	0.29878
Grand p		0.292096	0.2727	0.41748	0.1068	0.37762
λ		0.045713	-0.2758	-0.633	-0.1262	-0.35493
$V(\lambda)$		0.005711	0.0484	0.02852	0.00476	0.01728
Composite Data						
Mean λ	-0.2087	0.0457	-0.3671	**xxx**	-0.3796	-0.3549
Mean $V(\lambda)$	0.00091	0.00571	0.0301	**xxx**	0.01664	0.01728

*Abbreviations: Con = control, Exp = experimental, ET-CO2 = end tidal carbon dioxide concentration, hits = number of patients with favorable outcome, ID = identifier, MAP = mean arterial pressure, N = number of patients in a group, ROSC = return of spontaneous circulation, SEM = standard error of the mean

** Two columns accommodate two dichotomous variables, ROSC and discharge survival.

The lower block in Table 3 is for dichotomous survival data, including return of spontaneous circulation (ROSC) and discharge survival. The nature of dichotomous data requires different summary statistics. Numbers of good outcomes ("survivors") in both control and experimental groups and the respective group n's are tabulated, together with associated proportions. The variance estimate for the log result ratio with dichotomous data is computed from observed proportions as shown in Table 2 and Appendix 2 (A2.5b).

A key feature of the meta-analysis is generation of a single figure of merit, describing the result of each independent study. In Table 3 the mean log result ratio for all end points within a study is used to create such an estimate of treatment effect. Similarly, the mean variance estimate for all end points within a particular study is used to create an estimate of the typical variance of the result ratio for that study. This variance estimate is suitable for highly correlated variables within a particular study.

Cumulative Meta-analysis.

Results of cumulative meta-analysis for IAC-CPR are shown in Table 4 and in Figure 1. Table 4 is a continuation of the same spreadsheet shown in Table 3. The composite λ-values and their variances at the bottom of Table 3 for each of the 5 independent studies, involving entirely different patients, are transferred to the upper rows of Table 4 for each study. Here, working from left to right, a weighted mean λ-value and its estimated standard deviation are found using expressions (4) and (5). Successive columns from left to right represent successive stages of the cumulative meta-analysis. The mean λ-value under Study 1 describes the first study only. The mean λ-value under Study 2 describes the combined results of the first two studies. The mean λ-value under Study 3 describes the combined results of the first three studies, etc. Subsequent rows of R* data and associated critical values are obtained from the mean λ data by exponential transformation using expressions (3) and (6). Note that the power of the over-all analysis to detect an R* of 0.1 or greater increases substantially after publication of the third study (Table 4, bottom). Such calculations of beta error and statistical power can be useful to evaluate the possibility of Type 2 error in the event that the aggregate R* value is not significant.

Table 4. Cumulative meta-analysis of human studies of IAC-CPR

	Study 1	Study 2	Study 3	Study 4	Study 5
Study Weight	0.5	1	1	1	1
Study Number	1	2	3	4	5
Composite λ	-0.2088	0.0457	-0.3671	-0.3796	-0.3549
Composite V(λ)	0.0009	0.0057	0.0301	0.0166	0.0173
Weight*λ	-0.1044	0.0457	-0.3671	-0.3796	-0.3549
Weight^2*V(λ)	0.0002	0.0057	0.0301	0.0166	0.0173
$\bar{\lambda}$	-0.2088	-0.0391	-0.1703	-0.2301	-0.2578
$\hat{\sigma}(\bar{\lambda})$	0.0302	0.0514	0.0759	0.0656	0.0588
Ratio R*	0.1884	0.0384	0.1566	0.2055	0.2273
Critical pt.	-0.061	-0.10594	-0.1604	-0.13713	-0.1221
Critical pt.	0.0575	0.09579	0.13825	0.12059	0.1088
p-value	5E-12	0.44651	0.02488	0.00045	0.00001
Beta	4E-07	0.88469	0.38849	0.06062	0.00761
Power	1	0.11531	0.61151	0.93938	0.99239

Figure 1 is a plot of the aggregate result ratio, R*, and the associated critical values with publication of successive studies. The left-most data point represents the historically first trial, the second a combination of the first two trials, the third a combination of the first three, etc. The separate lighter weight lines, without data points, are upper and lower critical values; they include the 95 percent confidence intervals for R* under the null hypothesis. The data points and critical values plotted in Figure 1 correspond to the summary data in the columns of Table 4 from left to right. A significant aggregate effect of treatment in humans is achieved and maintained after the publication of the third study. The effect of IAC-CPR is both biologically and statistically significant. Biological significance can be judged from the absolute value of R*, here 20 percent of the ideal value of 1.0. Statistical significance can be judged from the difference between R* and the nearest critical value.

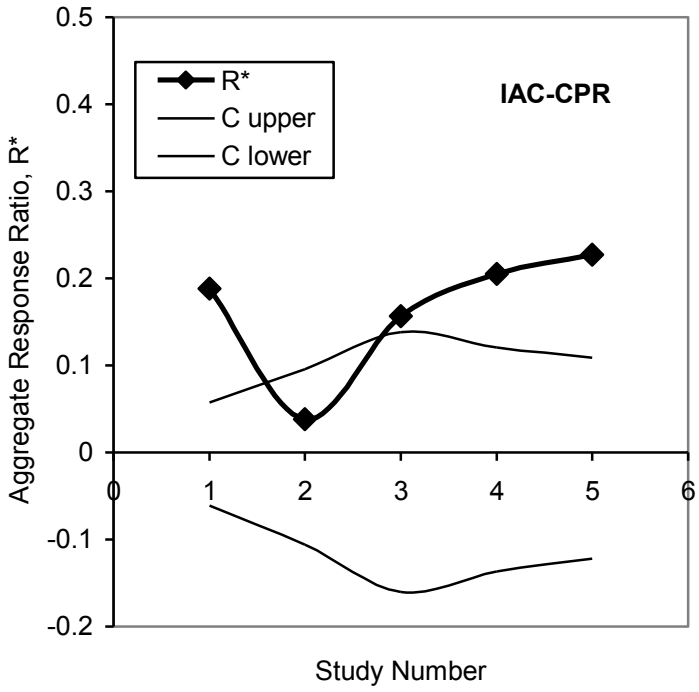

Figure 1. Cumulative meta-analysis of human studies of interposed abdominal compression CPR. In this horizontal format solid data points represent successive aggregate R* values computed from the data. The separate upper and lower curves represent critical values for statistical significance. Under the null hypothesis data points can be expected to occur between the upper and lower curves 95 percent of the time.

Analysis with Studies Omitted.

Table 5 shows results of meta-analysis of IAC-CPR when each of the 5 studies in turn is given zero weight. No single study drives the conclusion that IAC-CPR produces statistically significant benefit compared to standard CPR.

Table 5. Effects of omitting single studies

	Study omitted					
	None	1	2	3	4	5
Ratio, R*	0.227	0.232	0.291	0.203	0.200	0.206
Upper critical value	0.109	0.121	0.132	0.106	0.121	0.121

Subgroup Analysis.

Retrospectively, however, there does appear to be a difference between the four in-hospital studies and the one pre-hospital study of IAC-CPR. When the weights of all other studies are set to zero, the R* value for the one out-of-hospital study (study number 2) is -0.05. When the weight of study 2 is set to zero, the R* value for the remaining in-hospital studies is +0.29. A d-test, as described in Methods, for the difference in log(R') shows that the difference is greater than 3 standard deviation from zero. The failure of the one pre-hospital study to demonstrate a difference is explained by the substantial periods of standard CPR necessarily received by patients in the IAC-CPR group both before arrival of emergency medical personnel and during transport to the hospital[2]. Thus there is strong evidence, in particular, for the use of IAC-CPR in a hospital setting.

DISCUSSION

To speed the translation of valid research findings into clinical practice, guideline writers must make the most efficient use of available data, using methods such as cumulative meta-analysis[7]. The present statistical approach is a form of cumulative meta-analysis applicable to both continuous and dichotomous data, which tend to crop up heterogeneously in emerging research areas and in studies involving less common disease entities. Because the patients in one study are never compared directly with those in another study, it is not necessary to assume that the trials synthesized are exactly comparable (a Petonian approach[27, 28])—only that they test the same basic intervention as it might be implemented in various settings in the real world. Using clinical response ratios, the particular studies to be included in a meta-analysis can be selected on scientific and medical grounds, not on the basis of technical statistical requirements such as homogeneity of variance or the need to treat continuous and dichotomous data differently.

There are a variety of other helpful aspects of working with clinical response ratios. Clinical response ratios describe biologically meaningful effects of treatment, allowing synthesis of various end-points on a scale from 0 to 1.0, representing no improvement to maximal desirable improvement. $R^* = 0$ indicates absolutely no evidence of clinical benefit versus control. $R^* = 1$ indicates maximal or ideal clinical benefit. $R^* > C_U$ indicates significantly better performance of

the experimental treatment than the control. $R^* < C_L$ indicates significantly poorer performance than control.

Use of R* statistics rather than simple ratios of experimental to control end-points[8] allows synthesis of measures with a wide variation in dynamic ranges, for example arterial blood pH, for which the biological range is about 7.1 to 7.6 versus arterial blood pO_2, for which the biological range (including 100 percent oxygen breathing) is about 40 to 400 torr. A strength of the present method of meta-analysis is that the test of the null hypothesis itself does not depend on the fixed effect assumption. There is no formal requirement that the components of $\bar{\lambda}$ in expression (4) have the same mean or that they have the same variance. It is simply an average of independent random variables. However, these values are normalized to a biologically meaningful range by introduction of the ideal response factor, I.

Importantly, the significance test for Type 1 error is not sensitive to small differences in expert opinion regarding the choice of the normal values, I. This fact can be demonstrated numerically in specific examples and also analytically by calculus[***].

There may be some minor abuse potential in the choice of values, I , because there is some latitude in deciding what the ideal therapeutic response would be. Typically changes of continuous variables from abnormal to the mid normal value would be reasonable choices for ideal therapeutic benefit. Changes in dichotomous variables to 100 percent survival or response rates are similarly obvious choices. It would be difficult for biased analysts to circumvent these obvious choices, which will be visible for readers to judge and dispute if they wish.

A minor technical advantage of the present method using the ideal response factor over the simple response ratio method[8] is that it can be used for experiments in which the mean control value is zero or near zero. In such cases the denominator of the simple response ratio method would be too small.

The general notion of combining various types of data in one analysis is based upon the idea that the first step in the generation of clinical practice guidelines is simply to determine if a proposed treatment, in general, produces favorable effects[27]. If it does, then the effects that are seen in the selected trials are likely to generalize to the even broader range of circumstances found in widespread practice. When data are abundant possible heterogeneity of treatment effect among

---

[***] The analytical result (given here without detailed proof) is that as the value of ideal outcome, I, is changed by a modest amount, ΔI, the relative change in the upper limit for significance testing, Δc_U is approximately proportional to the change in clinical response ratio, ΔR. That is, $\frac{dc_U}{c_U} \approx \frac{\Delta R}{R}$. Accordingly, as I is changed, both the test statistic, R, and the critical value for significance testing, c_U, increase or decrease by the same percentage. Hence the result of the significance test is not influenced by small changes in I. In outline, the proof includes differentiation of expression (1) with respect to I to obtain dR/R, use of the small value approximation of e^x in expression (6) and noting that $\frac{dc_U}{c_U} \approx \frac{dS}{S}$, then finding dS/S by differentiation of expression (5), and noting that the result is the same as dR/R.

sub-populations becomes a major question. In the example of IAC-CPR, even with only five clinical studies to analyze, there is evidence that the method may be more effective for in-hospital resuscitation than for out-of-hospital resuscitation. In meta-analysis there seems to be a tradeoff between combining heterogeneous studies, which increases generalizability, and combining only homogeneous populations, which reduces variation but also reduces generalizability. The use of an overall significance test, followed by subgroup analysis, allows one to do the former, followed by the latter, obtaining the benefits of both.

The quality of any meta-analysis sis is heavily influenced by the quality of the studies included. Selection of studies depends greatly upon the viewpoint of the meta-analyst and the framing of the question to be addressed. Selection also may depend on subjective ratings of study quality, which can vary greatly[17]. Language bias may exclude trials published in languages other than English[29]. Data can be double counted inadvertently, for example in a separate single-center report of some of the same patients that are included in a multi-center trial. Studies with non-significant results -- especially ones with small sample sizes -- may be less likely to be published (publication bias), and hence may not be accounted in a formal meta-analysis[29]. Such biases can be minimized by a vigorous effort to include all relevant studies.

The explicit weighting factors in expressions (4) and (5) for the various studies may be viewed negatively as easily abusable fudge factors that could be used to skew an analysis in one direction or another, depending on the bias of the reviewer[17, 18]. In general, inclusion of all studies is recommended, using either equal weighting for quality [5, 14, 30] or inverse variance weighting[19, 20]. The seeming arbitrariness of weights may be unsettling to those unfamiliar with meta-analysis. If concern arises one can perform a sensitivity analysis by re-running the meta-analysis with alternative weights and presenting the results in a table, similar to Table 5. Typically the main results of the meta-analysis will be insensitive to changes in the weighting scheme, giving the reader confidence that the weights were not chosen to produce a particular result.

It is important to realize, however, that any synthesis of research findings will include the unavoidable selection bias of the reviewer, who must choose which studies to include and which to omit from the review (i.e. assigning weights equal to zero). If the explicit weights were omitted from expressions (4) and (5), then all included studies would have weight 1 and all excluded studies would have weight zero. The explicit weights merely highlight the necessary judgments required, forcing the reviewer to disclose subjective decisions.

A weighting scheme of particular interest is the inverse variance weighting. As shown by Hedges[19], selecting weights that are inversely proportional to the variances of individual random variables minimizes the variance of their sum. Such weights give the narrowest 95 percent confidence interval for the sum. Inverse variance weighting is easily implemented in spreadsheets such as Table 4, because variances of study log R' values are already calculated as shown in Tables 2 and 3.

CONCLUSIONS

When used properly and wisely, meta-analysis of clinical response ratios is a flexible tool to expedite the timely and accurate syntheses of early research findings in clinical medicine. In particular, meta-analysis may be used for combining studies in research areas where large clinical trials would not be practical or would be unethical—for example in the field of cardiopulmonary resuscitation where the issue of informed consent becomes quite difficult. The present method provides quantitative tests of both Type 1 and Type 2 statistical errors, which would lead respectively to either false positive or false negative evaluations of emerging treatments, approaches, or concepts. Such information can guide individual and institutional practice and shorten the time between research discoveries and their clinical implementation, especially in neglected "orphan" areas.

Appendices

Appendix 1. Variance estimate for the logarithm of a ratio of random variables

(a) Application of the delta method

The probable error method or delta method[16, 32] may be used to approximate variances of functions of random variables. If X is a random variable with mean μ and variance σ^2, the variance $\sigma^2(f(X)) \approx \sigma^2(X)(f'(\mu))^2$ where $f'(X)$ is the first derivative of function f(X) with respect to X. To appreciate the approximation one can visualize the function f(X) as a graph with a tangent of slope $f'(\mu)$ at point $(\mu, f(\mu))$. By deduction from such a graph, it follows that the standard deviation of f(X) is approximately $f'(\mu)$ times the standard deviation of X, as long as $f'(X)$ does not change greatly over the range of X.

For the case of $f(X) = \ln(X)$, the delta method gives $\sigma^2(\ln(X)) \approx \frac{1}{\mu^2}\sigma^2(X)$.

Hence,

$$\hat{\sigma}^2\left[\ln\left(\frac{X}{Y}\right)\right] = \hat{\sigma}^2\left[\ln(X) - \ln(Y)\right] = \hat{\sigma}^2[\ln(X)] + \hat{\sigma}^2[\ln(Y)] \cong \frac{\hat{\sigma}^2(X)}{(\overline{X})^2} + \frac{\hat{\sigma}^2(Y)}{(\overline{Y})^2}. \qquad \text{(A1.1)}$$

To explore limits of approximation (A1.1) and the shape of the distribution of ln(X/Y), a more detailed treatment is as follows.

(b) Application of a series expansion

Let X and Y be two independent random variables, each always > 0. Let us represent X and Y as $X = \mu_x + \sigma_x U_1$ and $Y = \mu_y + \sigma_y U_2$, where constants μ_x and μ_y are population means, σ_x

and σ_y are standard deviations such that $\sigma/\mu < 1$, and random variables U_1 and U_2 are distributed as N(0,1), that is U_1 and U_2 are independent standard normal deviates. Now consider the ratio

$$\frac{X}{Y} = \frac{\mu_x(1 + c_x U_1)}{\mu_y(1 + c_y U_2)}, \tag{A1.2}$$

where c_x and c_y are coefficients of variation (σ/μ), which are typically < 0.3 in order to keep X and Y > 0. Now

$$\ln\left(\frac{X}{Y}\right) = \ln\left(\frac{\mu_x}{\mu_y}\right) + \ln(1 + c_x U_1) - \ln(1 + c_y U_2)\ . \tag{A1.3}$$

Using two terms of the series expansion $\ln(1+\varepsilon) = \varepsilon - \frac{1}{2}\varepsilon^2 + \frac{1}{3}\varepsilon^3 - \frac{1}{4}\varepsilon^4 + \cdots$ for $|\varepsilon| < 1$, which can represent all but the largest occasional values of X and Y, we have

$$\ln\left(\frac{X}{Y}\right) = \ln\left(\frac{\mu_x}{\mu_y}\right) + c_x U_1 - c_y U_2 - \frac{1}{2}\left[c_x^2 U_1^2 - c_y^2 U_2^2\right] + \cdots\ . \tag{A1.4}$$

From inspection of the series expansion, the random variable (A1.4) equals a constant, $\ln(\mu_x / \mu_y)$, plus a normal deviate with variance $c_x^2 + c_y^2$ (i.e. the combination of $c_x U_1$ and $c_y U_2$), plus a smaller correction term. The variance of the first three terms is given by (A1.1). The correction term is the difference of two random variables that are distributed as U^2, namely a chi-square distribution with one degree of freedom. If $c_x = c_y$, then the distribution of the correction term will be symmetrical about zero, and the correction will add a small amount of extra noise or variance to the distribution of the normal approximation.

Given that standard normal deviates 1 and 2 are independent, that U^2 is not correlated with U, and that the variance of a chi-square distribution with 1 degree of freedom is 2, we have

$$\sigma^2\left(\ln\left(\frac{X}{Y}\right)\right) \cong c_x^2 + c_y^2 + \left(c_x^2\right)^2 + \left(c_y^2\right)^2, \text{ or}$$

$$\sigma^2\left(\ln\left(\frac{X}{Y}\right)\right) \cong \left(c_x^2 + c_y^2\right)\left[1 + \frac{\left(c_x^2\right)^2 + \left(c_y^2\right)^2}{c_x^2 + c_y^2}\right]. \tag{A1.5}$$

If $c_x = c_y = c$, then the ratio of the variance of the actual distribution to that of the approximation of (A1.1) is $1 + c^2$. For example if c = 1/4, then the variance ratio is 1 + 1/16. The ratio of the standard deviations is approximately $1 + \frac{1}{2}c^2$, or about a 3 percent

difference. Indeed, expression (A1.5) could be used to further refine variance estimates in Table 2; although the correction would be small. Thus when $c_x = c_y$ the normal approximation to the distribution of ln(X/Y) is quite good.

In contrast, however, if $c_x >> c_y$, or if $c_x << c_y$, then the mean value of the chi-square correction term $-\frac{1}{2}\left[c_x^2 U_1^2 - c_y^2 U_2^2\right]$ will be nonzero, introducing skewness as well as noise to the actual distribution of ln(X/Y). In the limiting case, in which $c = c_x >> c_y$ or $c_x << c_y = c$ the amount of the shift in the mean will be $0.5\, c^2\, E(U^2) = 0.5\, c^2$, where the expected value of a chi-square distribution with 1 degree of freedom, $E(U^2)$, is 1.0. For example, if $c_x = c_y = 0.25$ for typical data, then the bias due to skewness would be 0.0625. In practical meta-analyses the observed value of log response ratios ln(X/Y) may be in the neighborhood of ± 0.10 . Thus, the effect of skewness can be a substantial fraction of the effect of experimental treatment when $c_x \neq c_y$. For this reason variance estimates using (A1.1) and the assumption that ln(X/Y) is normally distributed apply most accurately to cases where the coefficients of variation of X and Y are the same, in particular $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y$.

Appendix 2. Variance estimate for the complement of the log clinical result ratio, R', for an individual study under the null hypothesis

For a particular study,

$$\lambda \equiv \ln(R') = \ln\left(\frac{I - \overline{E}}{I - \overline{C}}\right) \text{ for continuous data, and} \qquad \text{(A2.1a)}$$

$$\lambda \equiv \ln(R') = \ln\left(\frac{I - p_E}{I - p_C}\right) \text{ for dichotomous data.} \qquad \text{(A2.1b)}$$

We wish to estimate the variance of the λ's under the null hypothesis that there is no true effect of treatment on the sampling distributions for experimental and control data. One can estimate the variance of the distribution of λ=ln(R') values, using the relationship

$$\hat{\sigma}^2\left[\ln\left(\frac{X}{Y}\right)\right] \cong \frac{\hat{\sigma}^2(X)}{\left(\overline{X}\right)^2} + \frac{\hat{\sigma}^2(Y)}{\left(\overline{Y}\right)^2} \qquad \text{(A2.2)}$$

derived in Appendix 1. Note that A2.2 assumes that X and Y are independent, which is true for independent experimental and control groups in most clinical trials, because they contain different patients. The required variance estimate for continuous variables is

$$\hat{\sigma}^2(\lambda) = \hat{\sigma}^2\left[\ln\left(\frac{I-\bar{E}}{I-\bar{C}}\right)\right] = \frac{\hat{\sigma}^2(\bar{E})}{(I-\bar{E})^2} + \frac{\hat{\sigma}^2(\bar{C})}{(I-\bar{C})^2} \ .$$

Under H_0: $\bar{E} = \bar{C} = \bar{P}$, the common population mean, and $\sigma^2(\bar{E}) = \sigma^2(\bar{C}) = \sigma^2(\bar{P})$, so

$$\hat{\sigma}^2(\lambda) = 2\frac{\hat{\sigma}^2(\bar{P})}{(I-\bar{P})^2} .$$

The best estimate of the population mean under H_0 is the pooled common sample mean

$$\bar{P} = \frac{n_E\,\bar{E} + n_C\,\bar{C}}{N} \qquad \text{(A2.3a)}$$

for continuous data, or the pooled common proportion of successful outcomes

$$\bar{p} = \frac{n_E p_E + n_C p_C}{N} \text{ for dichotomous data,} \qquad \text{(A2.3b)}$$

where $N = n_C + n_E$ is the total number of subjects in the study.

For continuous data a well-accepted estimate of the common population variance under H_0 is

$$\hat{\sigma}^2(\bar{P}) = \frac{(n_E - 1)\,\hat{\sigma}^2(\bar{E}) + (n_C - 1)\,\hat{\sigma}^2(\bar{C})}{N-2}, \qquad \text{(A2.4a)}$$

which is a weighted average of the sample variance estimates, the weights being the respective degrees of freedom.

For dichotomous data we assume that the true proportion of survivors under H_0 is given by (A2.3b). For binomial distributions the variances of control and experimental proportions depend upon the sample sizes n_C and n_E, which could be different[31] (p 234). In this case an unbiased estimate of the variance for the control group under H_0 is

$$\hat{\sigma}^2(\bar{p}_C) = \frac{\bar{p}(1-\bar{p})}{n_C - 1} \ . \qquad \text{(A2.4b)}$$

Similarly, an unbiased estimate of the variance for the experimental group under H_0 is

$$\hat{\sigma}^2(\bar{p}_E) = \frac{\bar{p}(1-\bar{p})}{n_E - 1} \ . \qquad \text{(A2.4c)}$$

Now using expressions (A2.3a) for $\overline{P}$ and (A2.4a) for $\hat{\sigma}^2(\overline{P})$, we can compute the desired variance estimate for λ derived from continuous data under H_0 as

$$\hat{\sigma}^2(\lambda) = 2\frac{\hat{\sigma}^2(\overline{P})}{(I-\overline{P})^2} = 2\frac{(n_E-1)\,\hat{\sigma}^2(\overline{E})+(n_C-1)\,\hat{\sigma}^2(\overline{C})}{(N-2)(I-\overline{P})^2} \quad . \qquad \text{(A2.5a)}$$

Similarly, we can compute the desired variance estimate for λ derived from dichotomous data under H_0 as

$$\hat{\sigma}^2\left[\ln\left(\frac{I-p_E}{I-p_C}\right)\right] = \frac{\hat{\sigma}^2(p_E)}{(I-p_E)^2} + \frac{\hat{\sigma}^2(p_C)}{(I-p_C)^2}, \text{ where under } H_0 \; p_E = p_C = \overline{p}, \text{ and}$$

substituting expressions (A2.4b) and (A2.4c), we have for dichotomous data

$$\hat{\sigma}^2(\lambda) = \frac{\overline{p}(1-\overline{p})/(n_C-1)}{(I-\overline{p})^2} + \frac{\overline{p}(1-\overline{p})/(n_E-1)}{(I-\overline{p})^2} . \qquad \text{(A2.5b)}$$

References

1. The Cochrane Manual, www.cochrane.org. 2003.
2. 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.
3. Hedges LV, Olkin I. Vote-counting methods in research synthesis. Psychological Bulletin 1980; 88:359-369.
4. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Orlando: Academic Press, 1985.
5. Hunter JE, Schmidt FL. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Newbury Park, California: Sage Publications, 1990:592.
6. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.
7. 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.
8. Hedges LV, Gurevitch J, Curtis PS. The meta-analysis of response ratios in experimental ecology. Ecology 1999; 80:1150-1156.
9. Babbs CF. Simplified meta-analysis of clinical trials in resuscitation. Resuscitation 2003; In Press.
10. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994; 47:881-9.
11. Walter SD. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. Stat Med 1997; 16:2883-900.

12. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med 2000; 19:1707-28.
13. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 1998; 17:1923-42.
14. 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.
15. Feller W. An Introduction to Probability Theory and Its Applications. Vol. 1. New York: John Wiley & Sons, Inc., 1957:461 pages.
16. Beaton GH, Martorell R, Aronson KJ, et al. Effectiveness of vitamin A supplementation in the control of young child morbidity and mortality in developing countries. ACC/SCN State-of-the-art series--nutrition policy discussion paper. Vol. 13. Suffolk, England: Lavenham Press Ltd., 1993:120.
17. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.
18. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992; 45:255-65.
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. Mosteller F, Rourke REK. Sturdy Statistics: Nonparametrics and Order Statistics. Reading Massachusetts: Addison-Wesley, 1973:395 pages.
22. Welch BL. The significance of the difference between two means when the population variances are unequal. Biometrika 1937; 29:350-362.
23. Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. Psychological Methods 2001; 6:203-217.
24. 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.
25. Kern KB, Paraskos JA. 31st Bethesda Conference--Emergency Cardiac Care (1999). Journal of the American College of Cardiology 2000; 35:825-880.
26. Cummins RO. Introduction to the international guidelines 2000 for CPR and ECC. Circulation 2000; 102 (suppl I):I-1--I-11.
27. Peto R. Why do we need systematic overviews of clinical trials? Statistics in Medicine 1987; 6:233-240.
28. 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.
29. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol 2000; 53:207-16.

30. 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.
31. Altman DG. Practical Statistics for Medical Research. Boca Raton: Chapman & Hall, 1991:611.
32. Crumpler TB, Yoe JH. Chemical Computations and Errors. New York: John Wiley & Sons, Inc., 1940:174-179.