

**SIMULATION OF META-ANALYSIS FOR
ASSESSING THE IMPACT OF STUDY
VARIABILITY ON PARAMETER ESTIMATES FOR
SURVIVAL DATA**

by

Irina Karpova

Submitted to the Graduate Faculty of
the Graduate School of Public Health in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH
Graduate School of Public Health

This thesis was presented

by

Irina Karpova

It was defended on

April 10, 2006

and approved by

Thesis Advisor:
Stewart J. Anderson, PhD, MA
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Jong-Hyeon Jeong, PhD
Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Kevin Kip, PhD
Assistant Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

SIMULATION OF META-ANALYSIS FOR ASSESSING THE IMPACT OF STUDY VARIABILITY ON PARAMETER ESTIMATES FOR SURVIVAL DATA

Irina Karpova, M.S.

University of Pittsburgh, 2006

Meta-analysis is a statistical method of public health relevance that is used to combine the results of individual studies which evaluate the same treatment effect. A test that is commonly used to decide whether the results are homogeneous, and determines model choice for meta-analysis, is called Cochran's Q-test. A major drawback of the Q-test, when the outcomes are normally distributed, is its low power when the number of studies is small, and excessive power when the number of studies is large.

In this thesis, we propose a Cochran's Q-test for survival analysis data. Using simulations, we examine how the power of Cochran's test changes with different numbers of studies, different weight allocations per study, and the amount of censored observations. We show that the power increases with the increasing number of studies, but lowers with the increasing number of censored observations, and whenever one study comprises a large proportion of the total weight. We conclude that the test of heterogeneity should not be considered as the only determinant of the model choice for meta-analysis. Other methods such as graphical exploration, stratified analysis, or regression modeling should be used in conjunction with the formal statistical test.

TABLE OF CONTENTS

PREFACE	vii
1.0 INTRODUCTION	1
1.1 DEFINITION OF META-ANALYSIS	1
1.2 DIRECTION OF THESIS	4
2.0 LITERATURE REVIEW	6
3.0 METHODS AND RESULTS	8
3.1 SIMULATION METHODS	8
3.2 RESULTS	9
4.0 CONCLUSIONS	15
5.0 APPENDIX	16
BIBLIOGRAPHY	19

LIST OF TABLES

1	Power of Q -statistic (%) for variable number of studies n	12
2	Power of Q -statistic (%) for variable weights w_i	13
3	Power of Q -statistic (%) for variable censoring	14

LIST OF FIGURES

1	Histogram of standardized coefficients $\hat{\theta}_i, (i = 1, \dots, 10)$	10
2	Distribution of Q under Null	10
3	Power of Q -statistic (%) for variable number of studies	12
4	Power of Q -statistic (%) for variable weights	13
5	Power of Q -statistic (%) for variable censoring	14

PREFACE

I want to thank Dr. Stewart Anderson very much for his direction of this thesis and Drs. Jong-Hyeon Jeong and Kevin Kip for their careful reading and suggestions for improving this work.

I want to also thank my family for their love and support during the time I was a student and for their patience while I was writing this thesis.

1.0 INTRODUCTION

1.1 DEFINITION OF META-ANALYSIS

Meta-analysis can be defined as a statistical tool that deals with combining of the collection of findings from individual studies for the purpose of integrating them [1]. The results of the studies are pooled together in order to get a quantitative estimate of the effect of the intervention being studied. The interventions along with the control should be common to all of the studies included in the meta-analysis. The pooled quantitative estimate describes the observed relationship between the intervention and outcome. This could be summarized as an odds ratio or a relative risk for binary outcomes, a mean difference for continuous outcomes, or a hazard ratio for survival data.

Meta-analysis of randomized trials is based on the assumption that each trial provides an unbiased estimate of the effect of an experimental treatment, with the variability of the results between the studies being attributed to random variation. The overall effect calculated from a group of sensibly combined and representative randomized trials will provide a relatively unbiased estimate of the treatment effect, with an increase in the precision of this estimate. Such analyses are very important in medical research whenever information of a treatment performance is available from several clinical studies with similar treatment protocols.

The main requirement for performing meta-analysis is that the question under investigation should have some clinical importance and be biologically plausible [2]. In addition, some other reasons for doing meta-analysis include: the summarization of a large and complex body of literature on a topic, resolving conflicting research reports in scientific journals, avoiding the time and expense of conducting a big clinical trial, and improving the precision of an estimated treatment effect. Also, a meta-analysis can be performed in order to increase the statistical power through combining

several small studies into one big trial with a larger sample size.

Factors that can contribute to variability in the treatment effect between studies are design and conduct, clinical procedures, individual characteristics, exposures and outcomes studied [3]. Those may or may not be responsible for observed differences in the results across studies. Such factors are commonly referred to as clinical and methodological heterogeneity. The type of heterogeneity being referred to in this study is called statistical heterogeneity. Statistical heterogeneity exists when true effect estimates that need to be evaluated differ between studies, and the presence of it is indicated whenever the test of statistical homogeneity is significant.

A perceived drawback of meta-analysis is the lack of uniformity of results across studies. Pooling together the results of individual studies that address a common question into one meta-analysis can lead to the inclusion of the material with an element of diversity. Thus, investigation of heterogeneity of effect is an important part of any meta-analysis. In situations where the results of the studies that are combined differ greatly among each other, it may not be appropriate to calculate a summary effect statistic. Two types of models exist that can be used to produce summary effect measures: a fixed-effects model and a random-effects model [4]. A fixed-effects approach assumes homogeneity across the studies. When the above assumption is violated, a random-effect approach provides a way of incorporating between-study variability into the overall effect.

Cochran's Q -test is a statistical test that is commonly used to detect a statistical heterogeneity between individual results combined in one meta-analysis. If the test shows homogeneous results across studies then the differences between studies are assumed to be due to sampling variation, and a fixed effects model should be considered. If, on the other hand, the test shows significant heterogeneity across the studies, then a random effects model maybe appropriate. Let $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ represent the treatment effects in n studies. In order to get an overall estimate of the true effect, θ , a meta-analysis of the results of several studies should be performed. In the fixed effects approach, it is assumed that the true treatment effects are homogeneous across the studies, that is, $\theta_1 = \theta_2 = \theta_3 = \dots = \theta_n$. Accordingly, the overall treatment effect can be estimated as follows:

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i \hat{\theta}_i}{\sum_{i=1}^n w_i} \quad (1.1)$$

where w_i is the weight for study i often calculated as the reciprocal of the variance of $\hat{\theta}_i$, ($i = 1, \dots, n$ and $w_i = 1/v_i$). Hence, the variance component of the summary effect is only composed of terms for the within study variance of the individual trials.

The test of heterogeneity tests the hypothesis that the treatment effects are the same in all studies i.e. $H_0: \theta_1 = \theta_2 = \theta_3 = \dots = \theta_n$, with the alternative hypothesis stating that at least one treatment effect is different from others. The test statistics Q can be calculated as follows:

$$Q = \sum_{i=1}^n w_i(\hat{\theta}_i - \hat{\theta})^2 \quad (1.2)$$

where θ_i is an individual study effect, θ is an overall treatment effect from a meta-analysis of n separate trials, and w_i is the weight for study i . Under the null hypothesis, Q has an approximate χ^2 distribution with $n - 1$ *df* and if the p -value of the test is less than $\alpha = 0.05$, H_0 is rejected.

In situations where heterogeneity is present, it is inappropriate to combine separate estimates into a single number using fixed effects method. Random effects meta-analysis models heterogeneity as a variation of individual study effects around a population average effect. The adjusted weight for study i is given by $w_i^* = 1/(\nu_i + \sigma_B^2)$. Variance component includes a between study component σ_B^2 as well as a within study component, ν_i . If $\sigma_B^2 = 0$ there is no variability between treatments and this reduces to fixed effects model.

For both the fixed-effects and random-effects approaches, an approximate 95% confidence interval for the overall estimate of effect, $\hat{\theta}$, is given by $\hat{\theta} \pm 1.96\sqrt{(1/\sum_{i=1}^k w_i)}$, where w_i is the reciprocal of the variance ν_i of the estimated effect in the i^{th} study. Here, w_i is a measure of the random variation within the study for the fixed-effects approach, but includes the variation between the n estimated effects as well for the random-effects approach. Hence, the random effects model will usually produce a confidence interval wider, and never smaller, than that using the fixed-effects model.

Sometimes the Q -test fails to reject the null hypothesis of homogeneity even if a difference across the studies exists. Lack of power is a major drawback of the statistical test. Therefore, power is an important issue in the planning and conduct of meta-analysis. Several characteristics of meta-analysis such as the extent of heterogeneity present, the number of studies included, the weight

allocated to each study, can influence the power of the test. In addition, for survival analysis data, the degree of censoring can also influence the power of Cochran's test. Care must be taken in the interpretation of the test of heterogeneity when trials have small sample size or are few in number. While a statistically significant result may indicate a problem with heterogeneity, a non-significant result must not be taken as evidence of no heterogeneity. This is also why a $p = 0.10$, rather than the conventional level of $p = 0.05$, is sometimes used to determine statistical significance [5]. A further problem with the test is that when there are many studies in a meta-analysis, the test has high power to detect a small amount of heterogeneity that may be clinically unimportant.

The power of any statistical test refers to the probability that the null hypothesis indeed be rejected when some alternative hypothesis is true. In case of meta-analysis, it is necessary to avoid concluding falsely that the effect measures of different studies are homogeneous if in fact they are not. A simple way to reduce the chances of making such an error is to increase the number and size of the studies incorporated into one big investigation. In practice, it is not always possible to achieve that kind of situation, and, therefore, the power of the test for heterogeneity is often poor.

Provided that the weights are known, the expectation of the test statistics Q can be obtained, that is

$$E(Q) = (n - 1) + \sigma_B^2 \left[\sum_{i=1}^n w_i - \frac{\sum_{i=1}^n w_i^2}{\sum_{i=1}^n w_i} \right] \quad (1.3)$$

It can be shown from equation (1.3), when Q is roughly equal to the number of studies in the meta-analysis, there is a little evidence of heterogeneity. When Q is much larger than the number of studies, there is a significant evidence of heterogeneity. Increasing values of $E(Q)$ will generally imply increasing power.

1.2 DIRECTION OF THESIS

Given the above concerns, the purpose of this thesis is investigate the statistical power of the Q -test in the survival analysis setting. We will explore the relationship of the Q -statistic to the

sample size of individual trials. This had previously been done for outcomes that are normally distributed [4]. In addition we will investigate how the power of the Q -test changes with the weight distributions among the trials, and the degree of censoring within each study. In particular, we want to investigate how the power of the test for heterogeneity changes with changing number of studies included in a meta-analysis, weight allocations, and the amount of censoring in each trial. An answer to this question can provide a useful framework for evaluating meta-analytic methods in survival analysis.

For our research, we plan to use simulations in order to conduct 2000 meta-analyses under variable conditions described above. The survival data for each study will be randomly generated from an exponential distribution with a known distributional parameter. The hazard ratio for each hypothetical trial will be obtained and used as an effect measure. The overall effect measures $\hat{\theta}$, together with the Q statistics and the power of Cochran's Q -test will be obtained for each of the meta-analyses performed.

2.0 LITERATURE REVIEW

Two types of models are used to produce summary effect measures: fixed-effects models and random-effects models. A fixed-effects estimation method can be used to calculate an overall effect measure if there is no evidence of statistical heterogeneity. If statistical heterogeneity is believed to be present, a random effects estimation method may be appropriate. Vinh-Hung and Verschraegen (2004) conducted a meta-analysis of 15 published randomized clinical trials that compared radiotherapy versus no radiotherapy after breast-conserving surgery [6]. The outcomes studied were breast tumor recurrence and patient survival. Heterogeneity was assessed using the Cochran's Q test. For the recurrence outcome, the Q -test yielded a statistically significant heterogeneity across studies with substantial variation in relative risks. The pooled relative risk was estimated with a random effects model. The authors concluded that omission of radiotherapy is associated with a large increase in risk of breast cancer recurrence. On the other hand, the test of heterogeneity was not statistically significant when determining the relationship of radiotherapy versus no radiotherapy on patient mortality. However, as in the case of tumor recurrence, the overall relative risk was calculated with a random effects model. If the test of heterogeneity is not significant it is still possible that there may be important between-study variation since the test may have low statistical power.

The interpretation of the results of any meta-analysis depends on how substantial heterogeneity is, because the extent of heterogeneity determines how strongly it influences the conclusions. Higgins and Thompson et al. (2002) developed an approach that quantifies the effect of heterogeneity [7]. They provided measure of the degree of inconsistency between the results of the studies involved. They introduced a statistic, I^2 , that gives the percentage of total variation across studies that is due to heterogeneity other than chance. Their statistic, I^2 , can be calculated as $I^2 = 100\% \times (Q - df) / Q$

where Q is heterogeneity statistic and df denotes the degrees of freedom. I^2 can take values from 0% to 100%. 0% value implies no heterogeneity, and increasing values of I^2 indicates increasing heterogeneity.

Often meta-analyses include small numbers of studies and the power of Cochran's test in such cases is low. Jefferson et al. (2002) considered meta-analysis of eight randomized controlled trials of amantadine and rimantadine for preventing and treating influenza A in adults [8]. The treatment effects showed substantial variation with some of the confidence intervals not overlapping. The result of the test of heterogeneity was interpreted as being non-significant with a p -value of 0.09 which was not statistically significant with a cut-off of 5%. Using a 10% significance would, of course, change the conclusion with regard to the significance, but the risk of drawing a false positive conclusion would increase.

When a meta-analysis includes a large number of studies, the Q -test may have excessive power. For example, Barbui et al. (2003) conducted a meta-analysis of 135 clinical trials of tricyclic antidepressants and selective serotonin reuptake inhibitors for treatment of depression [9]. The test of heterogeneity turned to be highly significant with a p -value of 0.005. As in the first case, the p -value does not reasonably describe the extent of heterogeneity among those trials.

Hardy and Thompson (1998) used normally distributed data to simulate how the power of the test for heterogeneity depends on the number of studies included in the meta analysis, the weight allocated to each study, and the extent of between-study variance present [4]. It was concluded that the test of heterogeneity for assessing the validity of the fixed effect model is of limited use especially when the total information (that is total weight or inverse variance) is low, or when the amount of information in each trial is highly variable. In these situations the test of heterogeneity has low power and should not be the only determinant of model choice in the meta-analysis. It may be more appropriate to consider the random effects analysis as a check of the robustness of conclusions to failure in the assumption of homogeneity in a fixed effect model.

3.0 METHODS AND RESULTS

3.1 SIMULATION METHODS

For this thesis, we will start by generating random numbers from an exponential distribution with a known distributional parameter. The numbers will represent survival endpoints for each of the two hypothetical treatments in study i . An estimated treatment effect $\hat{\theta}_i$ will be calculated for every study i . An overall estimate of the treatment effect will be determined using equation (2.1). Once the effect measures for each of n trials and the overall treatment effect are calculated, the test statistic Q will be obtained by applying equation (2.2). Simulations will be carried out for each of a variety of values of the between-study variance, namely, $\sigma_B^2 = 0, 0.1, 0.2$. From each meta-analysis, a value Q will be obtained together with the power of the test for heterogeneity. A total of 2000 meta-analyses will be performed for each of the variable situations we want to consider.

We are going to change the number of studies included into one meta-analysis, $n = 5, 10, 20$ with 200 observations per study, and investigate the behavior of the power of heterogeneity test for different between-study variance σ_B^2 . Also, we are going to vary the weights allocated to each of the studies within one meta-analysis. For the initial simulations, the weights allocated to each trial will be kept equal for simplicity, that is $w_i = w$ for all i . The behavior of the test statistic Q and the power of the test under varying allocations of weights will then be considered by fixing the number of studies, $n = 10$, and allowing a single weight to be different while all other weights remain equal to each other. Three cases for w_i will be chosen with the restriction that the total sum of the weights remains the same. For the equal weighting case, w_i will not change from one study to another. For the next case w_1 will take 25 percent of the weights, and the other studies will get equal weighting. In addition, a situation where one study takes 50 percent of the total

weight will be considered. We also plan to investigate how the degree of censoring within each study will influence the expectation of Q for a given between-study variance. In this case we will assume equal weight distribution among 10 individual studies. We want to consider 0, 10, and 50 percent of randomly censored values per study to see the impact of the degree of censoring on the expectation of Q .

The research will be carried out using the freely available R statistics software package. The program provides easy reading, manipulation, and plotting of simulated data sets.

3.2 RESULTS

The simulations for investigating the power of the test for heterogeneity were performed using survival data that was generated from an exponential distribution with a known parameter. Each meta-analysis consisted of n studies. For the study n , $i = 1 \dots = n$, proportional hazards ratio, $\hat{\theta}_i$, was calculated. In addition, an overall proportional hazards ratio $\hat{\theta}$, Q statistic, and corresponding $E(Q)$ value were obtained for every simulation N . For each of the studies, the simulations were based on known within and between study variances, as well as known weights for each study. A goodness of fit test was performed on the distributions of standardized values of overall proportional hazard ratios $\hat{\theta}$ and the values of Q statistic.

Based on the Figure (3.1), and as confirmed by goodness of fit test, $\hat{\theta}_i$'s have an approximate normal distribution. The overall treatment effects $\hat{\theta}$ for each meta-analysis N have been calculated using Equation (2.1). For each series of studies, a distribution of 2000 values of test statistic Q , calculated by applying Equation (2.2), was obtained and analysed using a goodness of fit test. Under the null hypothesis of homogeneity, Q statistic has an approximate χ_{n-1}^2 distribution. Figure (3.2) illustrates the distribution of 2000 Q values for the meta-analyses of 10 studies provided $\sigma_B^2 = 0$, $w_1 = \dots = w_{10}$, and no censored observations exist. The smooth curve in the Figure (3.2) is the probability density function of a χ^2 distribution with $df = 9$.

Statistical power is defined as the probability of rejecting a null hypothesis when that hypothesis is false. The null hypothesis for the test for heterogeneity states that the effect measures are the

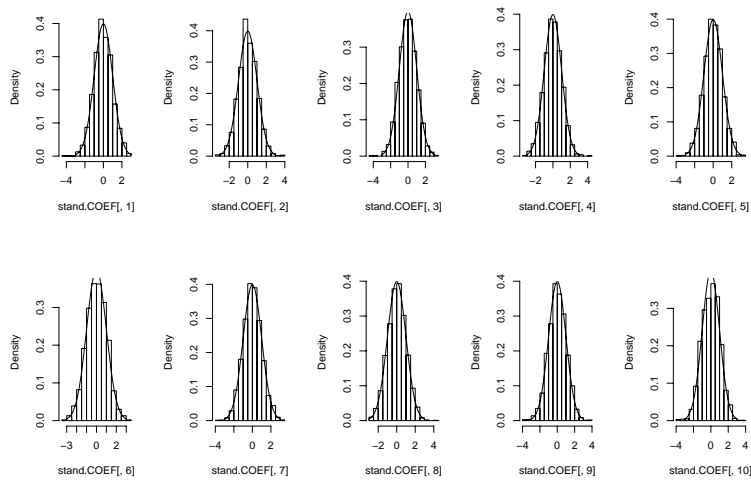


Figure 1: Histogram of standardized coefficients $\hat{\theta}_i$, ($i = 1, \dots, 10$)

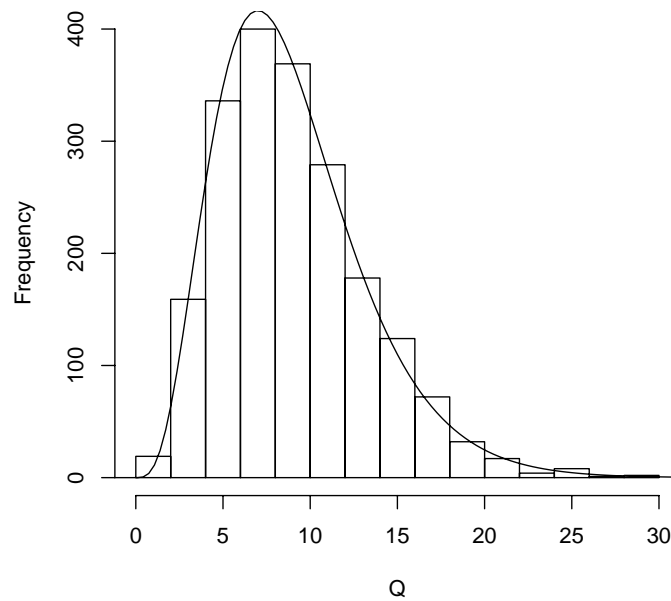


Figure 2: Distribution of Q under Null

same in all studies i.e. $H_0: \theta_1 = \theta_2 = \theta_3 = \dots = \theta_n$, with the alternative hypothesis stating that at least one treatment effect is different from others. For each series of simulations, the proportion of rejected values of Q was obtained by calculating the number of statistics greater than the critical value of χ_{n-1}^2 ($\alpha = 0.05$, $df = (n - 1)$) and dividing by N . The proportion of rejected Q values multiplied by 100% represents the power of the test for heterogeneity for a given conditions.

For our research, we used randomly generated survival data and determined the distribution of the calculated individual effect measures $\hat{\theta}_i$. A similar study was performed by Hardy and Thompson (1998), who simulated how the power of the Cochran's test changes with different number of studies and allocation of the weights [4]. However, for their investigation they used a normal distribution with the known parameters to randomly generate the effect measures $\hat{\theta}_i$'s. It was found that the power increases when the number of studies per meta-analysis gets larger and when the weights for each study are equal.

Based on equation (2.3), the value of $E(Q)$ and hence, the power of Q -test, depends on the number of studies n (assume 200 observations per study), weight allocations w_i , and the extent of between-study variability σ_B^2 in a meta-analysis. Figure (3.3) demonstrates how the power of Cochran's Q -test changes with increasing n and σ_B^2 , given the weight w_i allocated to each trial remains the same and no censored observations exist. The individual data points are listed in Table (3.1)

For fixed non-zero σ_B^2 , the expectation of Q increases with increasing n , and for a specific n , the expectation of Q increases with the increasing between-study variability. Furthermore, the power of Q -test increases faster for a given σ_B^2 for a larger n . When considering the situation with different weight allocations w_i , it was shown that for a given σ_B^2 , the values of $E(Q)$ are greatest when the weights are the same. The power of the Q -test decreases as the weights within a meta-analysis for a given σ_B^2 become more different (Table 3.2, Figure 3.4)

In addition, the impact of variable degree of censoring within a meta-analysis was investigated. As Figure (3.5) illustrates, the test for heterogeneity for a fixed non-zero σ_B^2 has more power in the situation when no censored observations exist. The power tends to decrease with the increasing amount of censored observations. As the degree of censoring increases from 10% to 50% per study, the power of Q -test decreases substantially for a given between-study variance (Table 3.3).

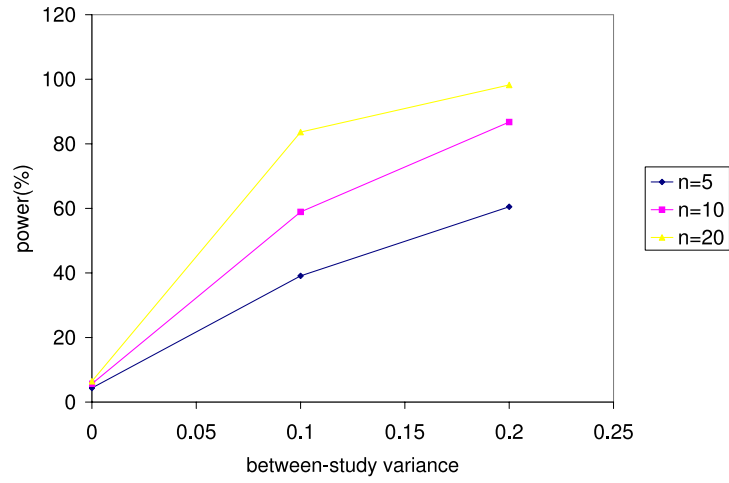


Figure 3: Power of Q -statistic (%) for variable number of studies

Table 1: Power of Q -statistic (%) for variable number of studies n

	number of studies (n)		
	$n = 5$	$n = 10$	$n = 20$
$\sigma_B^2 = 0.0$	4.35	5.65	6.45
$\sigma_B^2 = 0.1$	39.10	58.90	83.65
$\sigma_B^2 = 0.2$	60.50	86.75	98.25

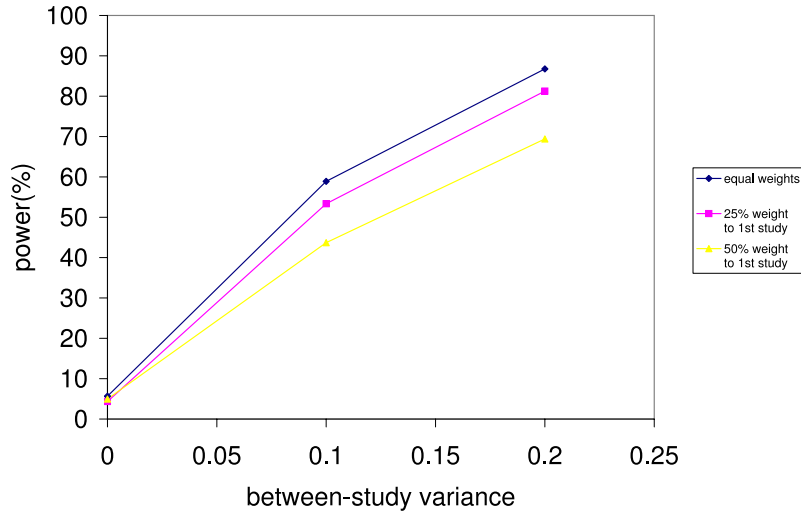


Figure 4: Power of Q -statistic (%) for variable weights

Table 2: Power of Q -statistic (%) for variable weights w_i

	weight (w_i)		
	equal weight	$w_1 = 25$	$w_1 = 50$
$\sigma_B^2 = 0.0$	5.65	4.35	4.95
$\sigma_B^2 = 0.1$	58.90	53.35	43.70
$\sigma_B^2 = 0.2$	86.75	81.25	69.40

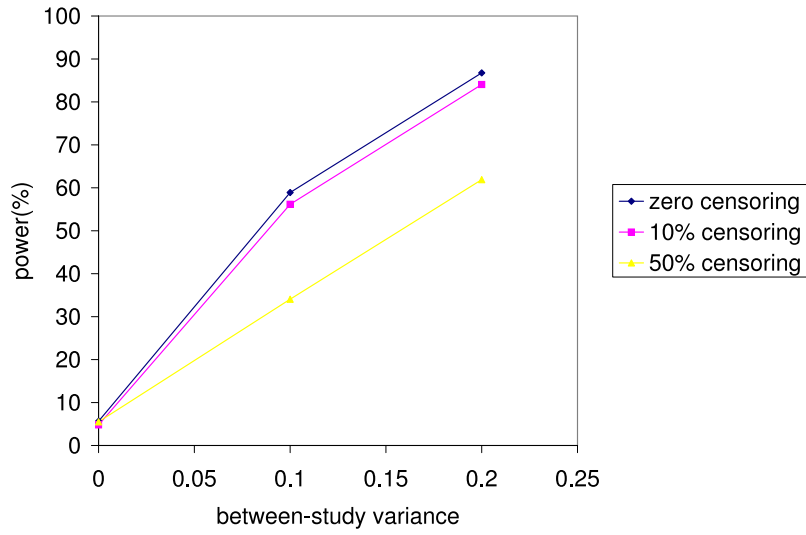


Figure 5: Power of Q -statistic (%) for variable censoring

Table 3: Power of Q -statistic (%) for variable censoring

	censoring		
	0% cens	10% cens	50% cens
$\sigma_B^2 = 0.0$	5.65	4.80	5.50
$\sigma_B^2 = 0.1$	58.90	56.15	34.05
$\sigma_B^2 = 0.2$	86.75	84.05	61.90

4.0 CONCLUSIONS

Our results confirm that the power of Cochran's test for heterogeneity in the survival analysis setting is low when the number of studies within a meta-analysis is small, or when the weight given to one study is substantially greater than the weights allocated to the other studies. Also, in survival data, the power of the Q -test decreases with an increasing amount of censored observations.

As the results of this investigation imply, the Cochran's test which is commonly used for determining whether a fixed effects or random effects model should be used for the calculation of an overall effect measure, cannot be considered as the only criterion of model choice. At the simplest level, heterogeneity of results between studies can be visually examined with the help of a forest plot of individual results. The plot can be used in conjunction with a formal statistical test for heterogeneity. Another possibility to check the assumption of homogeneity is to perform both fixed and random effects analyses and use the resulting information to investigate the sensitivity of the tests. Different results obtained from those two models usually imply between-study heterogeneity. In subgroup analyses, subsets of studies chosen according to one or more specific characteristics can be assigned to a separate meta-analysis. The differences between summary estimates in the subsets should also indicate heterogeneity. Another possible way to explore heterogeneity of results is by using meta-regression, an approach where an overall effect measure can be obtained by controlling for the effects of study covariates, given an adequate number of studies.

Our research can be extended to explore the behavior of the power of Cochran's test for many different combinations of study conditions. Using our R code, we can estimate its power for many different situations just by manipulating the number of studies, the number of observations per study, the amount of censoring, and the between-study variances.

5.0 APPENDIX

R CODE FOR SIMULATION STUDY

Listing of Program to generate simulations for the Q statistic

```
#####  
# This program simulates survival data for several  
# studies. Both failure and censoring distributions  
# are generated for n studies. A "Q" statistic  
# defined similar to that by  
# Hardy and Thompson, Stat in Med, 17, 841-856, 1998,  
# is calculated for each n studies. This process is  
# repeated N times and distributional properties of the  
# Q statistics under different null or alternative  
# hypotheses are obtained.  
# Program by Irina Karpova and Stewart Anderson  
#####  
  
library(survival)  
  
#####  
# First, initialize parameters that will change  
# from simulation to simulation  
#####  
  
N<-2000 # Number of simulations  
n<-10 # Number of studies per meta-analysis  
sample.size<-rep(200,n) # Assign sample size per study (Make it divisible by 2)  
lambda.1 <- rep(2,n) # Assign rates for two exp functions  
lambda.2 <- rep(2,n)  
cens.1 <- 1.0 # 1 - proportion censored in sample 1  
cens.2 <- 1.0 # 1 - proportion censored in sample 2  
true.sigma.b.sq <- 0.0  
bet.study.var.1 <- 0.0 # Add noise to the parameter in sample 1  
bet.study.var.2 <- 0.0 # Add noise to the parameter in sample 2
```

Listing of Program to generate simulations for the Q statistic (cont.)

```
#####
# Initialize output vectors
#####

theta<-rep(NA,N) # Initialize overall treatment effect per m-a
Q<-rep(NA,N) # Initialize Q statistic
E.Q<-rep(NA,N) # Initialize expectation of Q (power)
coef.result<-rep(NA, n) # Initialize treatment effects for n studies in a m-a
se.result<-rep(NA,n) # Initialize results
total.lambda.1<-rep(NA,n)
total.lambda.2<-rep(NA,n)
var.result<-rep(NA,n)
weight.result<-rep(NA,n)
weight.coef<-rep(NA,n)
study.COEf<-matrix(rep(-999,N*n),nrow=N)
stand.COEf<-matrix(rep(-999,N*n),nrow=N)
var.coef.result<-rep(NA,N)

for(j in 1:N){ # START BIG LOOP
  noise.lambda.1<- rnorm(n,0,sqrt(bet.study.var.1))
  noise.lambda.2<- rnorm(n,0,sqrt(bet.study.var.2))
  total.lambda.1<-lambda.1+noise.lambda.1 # All vectors have length n
  total.lambda.2<-lambda.2+noise.lambda.2 # All vectors have length n
  for(i in 1:n){ # Start small loop (for the n studies)
    trt<-c(rep(0,sample.size[i]/2),rep(1,sample.size[i]/2)) # generate treatments 0 and 1
    survtimex<-c(rexp(sample.size[i]/2,total.lambda.1[i])) # generate surv time x values
    survtimey<-c(rexp(sample.size[i]/2,total.lambda.2[i])) # generate surv time y values
    cens.time.1 <- rbinom(sample.size[i]/2,1,cens.1) # gen the prop of censored sample 1
    cens.time.2 <- rbinom(sample.size[i]/2,1,cens.2) # gen the prop of censored sample 2
    censtimes <- c(cens.time.1,cens.time.2) # Put censoring variables in both samples
    fail.time<-c(survtimex,survtimey)
    mfit<-survfit(Surv(fail.time,censtimes==1)~trt)
    result<-coxph(Surv(fail.time, censtimes==1)~trt) # calculate prop haz ratio
    coef.result[i]<-result$"coefficients"[1] # haz ratio for n studies
    se.result[i] <- sqrt(result$"var"[1]) # from result above take variance
    var.result[i]<-(se.result[i])^2
    weight.result[i]<-(1/var.result[i])
    weight.coef[i]<-(weight.result[i]*coef.result[i])
  } # end small loop

  study.COEf[j,<-coef.result
  stand.COEf[j,<-coef.result/se.result
  theta[j]<-sum(weight.coef)/sum(weight.result)
  theta.temp <- rep(theta[j],n)
  Q[j]<- sum(weight.result[i]*(study.COEf[j,] - theta.temp)^2)
  wgths<-sum(weight.result) - sum(weight.result^2)/sum(weight.result)
  E.Q[j]<-(n-1) + true.sigma.b.sq*wgths
} # END BIG LOOP
```


Listing of Program to generate simulations for the Q statistic (cont.)

```
#####  
# Get summary results across the N simulations  
#####  
  
prop.reject <- sum(ifelse(1-pchisq(Q,n-1)<=0.05,1,0))/N  
  
#####  
# graphs of theta's and Q's  
#####  
par(mfrow=c(2,5))  
hist(stand.COEf[,1],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,2],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,3],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,4],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,5],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,6],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,7],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,8],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,9],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
hist(stand.COEf[,10],20,prob=T)  
curve(dnorm(x,0,1),-3,3,add=T)  
  
par(mfrow=c(1,1))  
hist(Q,main="")  
curve(N*2*dchisq(x,n-1),add=T)  
title(paste("Between variance=",true.sigma.b.sq))
```

BIBLIOGRAPHY

1. DerSimonian R., and Liard N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7, 177-188.
2. Lau J., Ioannidis J.P., Schmid C.H. (1998). Summing up evidence: one answer is not always enough. *Lancet* 351, 123-127.
3. Thompson S.G.(1994) Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 309, 1351-1355.
4. Hardy R.J., Thompson S.G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 17, 841-856.
5. Petitti D.B. (2001). Approaches to heterogeneity in meta-analysis. *Statistics in Medicine* 20, 3625-3633.
6. Ving-Hung V., Verschraegen C. (2004). Breast-conserving surgery with or without radiotherapy: pooled analysis for risks of ipsilateral breast tumor recurrence and mortality. *JNCI* 96, 115-121.
7. Higgins J.P.T., Thompson S.G.(2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21, 1539-1558.
8. Jefferson T.O., Demicheli V., Deeks J.J., Rivetti D. (2002). Amantadine and rimantadine for preventing and treating influenza A in adults. *Cochrane Database Systematic Reviews* 4, CD002791.
9. Barbui C. Hotopf M., Freemantle N., Boynton J., Churchill R., Eccles M.P., Geddes J.R. et al. (2003). Treatment discontinuation with selective serotonin reuptake inhibitors (SSRIs) versus tricyclic antidepressants (TCAs). *Cochrane Database Systematic Reviews* 3, CD002791.