STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:6093–6110
Published online 9 September 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3428

The exact distribution of Cochran's heterogeneity statistic in one-way random effects meta-analysis[‡, §]

Brad J. Biggerstaff[1, *, †] and Dan Jackson[2]

[1]*Division of Vector-Borne Infectious Diseases, National Center for Zoonotic, Vector-Borne, and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, U.S.A.*
[2]*MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.*

SUMMARY

The presence and impact of heterogeneity in the standard one-way random effects model in meta-analysis are often assessed using the Q statistic due to Cochran. We derive the exact distribution of this statistic under the assumptions of the random effects model, and also suggest two moment-based approximations and a saddlepoint approximation for Q. The exact and approximate distributions are then applied to obtain the corresponding distributions of the recently proposed heterogeneity measures I^2 and H_M^2, the power of the standard test for the presence of heterogeneity and confidence intervals for the between-study variance parameter when the DerSimonian–Laird or the Hartung–Makambi estimator is used. The methodology is illustrated by revisiting a recent simulation study concerning the heterogeneity measures and applying all the proposed methods to four published meta-analyses. Published in 2008 by John Wiley & Sons, Ltd.

KEY WORDS: normal quadratic form; DerSimonian–Laird estimator; saddlepoint approximation; homogeneity test

1. INTRODUCTION

Meta-analysis, the statistical process of pooling the results from separate studies concerned with the same treatment or issue, is frequently used in the context of medical statistics and provides the quantitative backbone of the evidence based medicine programme. Despite this, there are difficulties associated with the application of this type of technique. In particular, there is the difficulty in attempting to combine obviously disparate results in order to produce a single estimate

*Correspondence to: Brad J. Biggerstaff, Division of Vector-Borne Infectious Diseases, National Center for Zoonotic, Vector-Borne, and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521, U.S.A.
†E-mail: BBiggerstaff@cdc.gov

‡The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Received 28 August 2007
Accepted 24 July 2008

of treatment effect. In order to describe and quantify the differences in the studies' findings, the random effects model, described in detail in Section 2, has become a standard approach. This involves an unobserved random effect, which provides the necessary between-study variation in order to take into account the disparities in the studies' results. Although the conventional random effects model assumes that this random effect is normally distributed, other distributions have also been suggested for this purpose [1].

The most commonly used statistic in the context of formally assessing the magnitude of the random effect is Cochran's Q statistic [2]. The versatility of this statistic is considerable: it can be used to test the null hypothesis that there is no between-study variation [3], estimate the magnitude of this [4], and can be transformed, either to provide confidence intervals for between-study variance [5] or other measures that quantify the impact of between-study heterogeneity [6]. Although the distribution of Q is well known under the hypothesis that there is no between-study variation, and there is considerable interest in detecting and describing heterogeneity in meta-analysis [7], the properties of Q under the conventional random effects model more generally are currently poorly understood.

In this paper, the exact distribution of Cochran's Q statistic is derived. Although this distribution is evaluated numerically in practice, the necessary quantities can be obtained accurately enough for all practical purposes. The rest of the paper is set out as follows. In Section 2 the random effects model and Cochran's statistic are described, and the exact distribution of Q is obtained. In Section 3, some more easily computed approximations to the distribution of Q are developed. In Section 4, three important applications of the exact distribution of Q are examined: the recently proposed measures for the impact of heterogeneity; the power of the standard test for the presence of heterogeneity and confidence intervals for the magnitude of this. Hence no judgement is made concerning the appropriateness of the various uses of Q, all of which are shown to easily incorporate its exact distribution. The methodology is illustrated by obtaining the results from a recent simulation study [8] free of Monte Carlo error (Section 5), and then applying the proposed methods to some example data sets (Section 6). Section 7, the discussion, concludes the paper.

2. THE RANDOM EFFECTS MODEL AND COCHRAN'S Q STATISTIC

The random effects model [4, 5, 9] assumes that the outcome from each of k studies, Y_i for $i=1,2,\ldots,k$, may be modeled as

$$Y_i = \mu + u_i + \varepsilon_i \tag{1}$$

where $u_i \sim \mathrm{N}(0,\tau^2)$, $\varepsilon_i \sim \mathrm{N}(0,\sigma_i^2)$ and u_i and ε_i are mutually independent. It is conventional in the meta-analysis setting to assume that the within-study variances σ_i^2 are known [5, 9], though in application they are assumed to be well-estimated and are replaced by study-specific estimates. Formulae for estimating the values of σ_i^2 used in practice, for a wide range of measures of treatment effect used in meta-analysis, are described in detail by Sutton *et al.* [3]. Although this standard approach is adopted throughout, this approximation requires sufficiently large studies and it should be noted that some recent developments recognize that the within-study variances are given in the form of estimates and that these are typically functions of the underlying treatment effect [10, 11]. Furthermore, Hartung and Knapp [12] show that treating the within-study variances as fixed and known in this way can distort the distributions of test statistics and illustrate this finding

Published in 2008 by John Wiley & Sons, Ltd. *Statist. Med.* 2008; **27**:6093–6110
DOI: 10.1002/sim

in a simulation study. This point is returned to when applying the methodology to the example data sets in Section 6.

The random variable u_i denotes the random, study-specific deviation from the mean effect and the parameter τ^2 represents the between-study variance: $\tau^2>0$ reflects underlying study differences in a formal sense, and if $\tau^2=0$ then all studies have the same underlying effect μ, providing a fixed effects model.

Cochran's Q statistic is frequently used in conjunction with this model, and is conventionally written as the weighted sum of squares

$$Q=\sum_{i=1}^{k} w_i(Y_i-\widehat{\mu})^2$$

where $w_i=1/\sigma_i^2$ and $\widehat{\mu}=(\sum_i w_i Y_i)/(\sum_i w_i)$. In order to derive its exact distribution, it is however more convenient to write Q in terms of its matrix representation. Let $\mathbf{Y}$ be the vector containing the Y_i, and let $\mathbf{A}=\mathbf{W}-(1/w_+)\mathbf{w}\mathbf{w}^{\mathrm{t}}$, where $\mathbf{W}$ is the diagonal matrix containing the $w_i=1/\sigma_i^2$, $\mathbf{w}$ is the vector containing the w_i, $w_+=\sum_i w_i$ and t denotes matrix transpose. The matrix representation of Q is then given by

$$Q=\mathbf{Y}^{\mathrm{t}}\mathbf{A}\mathbf{Y}$$

Next, let $\mathbf{\Sigma}$ denote the variance of $\mathbf{Y}$, the diagonal matrix with entries $\sigma_i^2+\tau^2$, and let $\mathbf{Z}$ denote a standard multivariate normal k-vector. Noting that Q is location invariant, we may write

$$Q=\mathbf{Y}^{\mathrm{t}}\mathbf{A}\mathbf{Y}\stackrel{d}{=}\mathbf{Z}^{\mathrm{t}}\mathbf{\Sigma}^{1/2}\mathbf{A}\mathbf{\Sigma}^{1/2}\mathbf{Z}=\mathbf{Z}^{\mathrm{t}}\mathbf{S}\mathbf{Z}$$

defining $\mathbf{S}=\mathbf{\Sigma}^{1/2}\mathbf{A}\mathbf{\Sigma}^{1/2}$. Since $\mathbf{A}$ and hence $\mathbf{S}$ are symmetric, writing $\mathbf{S}$ in terms of its spectral decomposition yields

$$Q\stackrel{d}{=}\sum_{i=1}^{k}\lambda_i(\mathbf{v}_i^{\mathrm{t}}\mathbf{Z})^2$$

where $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k\geqslant 0$ are the ordered eigenvalues of $\mathbf{S}$ and $\mathbf{v}_i$ is the ith eigenvector of $\mathbf{S}$ corresponding to λ_i. As $\mathbf{Z}\sim\mathrm{N}(0,\mathbf{I}_k)$ and the $\mathbf{v}_i$ are orthonormal, the $\mathbf{v}_i{}^{\mathrm{t}}\mathbf{Z}$ are independently distributed as $\mathbf{v}_i{}^{\mathrm{t}}\mathbf{Z}\sim\mathrm{N}(0,1)$, so that the distribution of Q may be expressed as

$$Q\stackrel{d}{=}\sum_{i=1}^{k}\lambda_i\chi_i^2(1) \tag{2}$$

where $\chi_i^2(1)$ are mutually independent chi-squared random variables with 1 degree of freedom.

Since the rows of $\mathbf{A}$ sum to zero, this matrix has an eigenvalue of 0, and hence so does $\mathbf{S}$. Thus $\lambda_k=0$, and the sum in equation (2) need only extend to $k-1$. Hence, Q has the distribution of a linear combination of independent, central $\chi^2(1)$ random variables, depending on τ^2 through the coefficients, the eigenvalues λ_i of $\mathbf{S}$. When $\tau^2=0$, $\lambda_i=1$ for each $i=1,2,\ldots,k-1$, so that Q correctly follows a $\chi^2(k-1)$ distribution. Furthermore, if $\sigma_1^2=\sigma_2^2=\cdots=\sigma_k^2\equiv\sigma^2$, then each positive $\lambda_i=1+\tau^2/\sigma^2$, so that $Q\sim(1+\tau^2/\sigma^2)\chi^2(k-1)$, as noted by Jackson [13].

Hedges and colleagues [14–16] also examine the distributional results in meta-analysis and note that the exact distribution of Q under the random effects model has, as we have just shown, the same distribution as a weighted average of chi-squared random variables. In none of these papers,

however, is this distribution stated explicitly, i.e. by noting that the weights are the eigenvalues of **S**. Hedges and Pigott [15, 16] use the same two-moment approximation for the distribution of Q given in Biggerstaff and Tweedie [5] (see below), arguing that this approximation should suffice for power calculations, a topic we consider in Section 4. Finally, Viechtbauer [17, p. 45] suggests that Q might reasonably be thought to have a non-central chi-squared distribution, though he then discounts this by noting that the exact variance would not match the proposed non-central chi-squared distribution with matching mean.

2.1. Obtaining the distribution of Q numerically

Denote the cumulative distribution function (CDF) of Q in equation (2) by $F_Q(x;\tau^2)$, where we stress the dependence on the heterogeneity variance τ^2. Although the evaluation of $F_Q(x;\tau^2)$ must generally be performed numerically, in the current computational environment this presents little difficulty. The eigenvalues of **S** can easily be computed, although note that these are functions of τ^2 and hence must be evaluated for each value of τ^2 under consideration. Once the eigenvalues λ_i have been obtained, the CDF of a positive linear combination of chi-squared random variables may be computed using a computational algorithm given in Farebrother [18]. An implementation of this algorithm in Pascal is available from statlib (http://lib.stat.cmu.edu/apstat/204), which we (B. J. B.) translated into C for use with S-Plus (Insightful Corp., Seattle, WA) or R (http://www.r-project.org). This translated algorithm, in conjunction with the S-Plus `eigen` command, was used to obtain $F_Q(x;\tau^2)$ in the computations that follow; the resulting S-Plus package 'lincombchisq', which implements Farebrother's algorithm, is available (http://csan.insightful.com). Finally, for computational convenience, we utilized the algorithm of Marsaglia [19] to compute the complementary CDF of the standard normal distribution when using Farebrother's algorithm; for all other uses of this CDF we utilized the native S-Plus and R routines.

3. APPROXIMATIONS TO THE DISTRIBUTION OF Q

Having derived the exact distribution of Q, it is of interest to compare this to some more simply computed approximations, in order to assess their suitability. Three such approximations to the distribution of Q are described in this section.

3.1. Approximation one: a two-moment gamma approximation

This approximation was initially developed by Biggerstaff and Tweedie [5] and was also adopted by Jackson [13]. Using what is essentially Satterthwaite's approximation [20], the approximating distribution is obtained by matching the first two moments of the gamma and Q distributions. Explicit formulas for the mean and variance of Q are

$$E[Q] = k - 1 + \left(S_1 - \frac{S_2}{S_1}\right)\tau^2$$

$$\mathrm{Var}[Q] = 2(k-1) + 4\left(S_1 - \frac{S_2}{S_1}\right)\tau^2 + 2\left(S_2 + \frac{S_2^2}{S_1^2} - 2\frac{S_3}{S_1}\right)\tau^4$$

where $S_r = \sum_i w_i^r$ for integer values of r [5]. These moments may also be obtained from the eigenvalues of $\mathbf{S}$, as $E[Q] = \sum_i \lambda_i$ and $\mathrm{Var}[Q] = 2\sum_i \lambda_i^2$.

The two-moment gamma approximation to the CDF of Q, with shape and rate parameters r and θ, respectively, is obtained by solving the equations $E[Q] = r/\theta$ and $\mathrm{Var}[Q] = r/\theta^2$. Emphasizing the dependence on τ^2, this gives

$$r(\tau^2) = \frac{(E[Q])^2}{\mathrm{Var}[Q]} \quad \text{and} \quad \theta(\tau^2) = \frac{E[Q]}{\mathrm{Var}[Q]}$$

The approximate CDF of Q is then given by computing the gamma CDF $F_G(x; \tau^2)$ with parameters $r(\tau^2)$ and $\theta(\tau^2)$.

3.2. Approximation two: a three-moment Pearson type III approximation

The Pearson type III distribution provides an extension of the previous two-moment gamma approximation. The third central moment (TCM) of Q can be derived, in a similar manner as Biggerstaff and Tweedie [5] obtain its variance, as

$$\mathrm{TCM}[Q] = E[(Q - E[Q])^3] = 8(k-1) + 24\left(S_1 - \frac{S_2}{S_1}\right)\tau^2 + 24\left(S_2 - 2\frac{S_3}{S_1} + \frac{S_2^2}{S_1^2}\right)\tau^4 + 8\left(S_3 - 3\frac{S_4}{S_1} + 3\frac{S_2 S_3}{S_1^2} - \frac{S_2^3}{S_1^3}\right)\tau^6$$

This moment can also be obtained from the eigenvalues of $\mathbf{S}$ as $\mathrm{TCM}[Q] = 8\sum_i \lambda_i^3$. Matching all three moments in a similar manner as above, and again emphasizing the dependence on τ^2, gives

$$r(\tau^2) = \frac{4\,\mathrm{Var}[Q]^3}{\mathrm{TCM}[Q]^2}, \quad \theta(\tau^2) = \frac{2\,\mathrm{Var}[Q]}{\mathrm{TCM}[Q]} \quad \text{and} \quad \gamma(\tau^2) = E[Q] - \frac{2\,\mathrm{Var}[Q]^2}{\mathrm{TCM}[Q]}$$

The approximate CDF of Q can easily be computed from the Pearson type III CDF $F_P(x; \tau^2)$ with location parameter $\gamma(\tau^2)$, shape parameter $r(\tau^2)$ and rate parameter $\theta(\tau^2)$; when $\gamma(\tau^2) = 0$, this distribution reduces to the previous gamma approximation. Note that although this approximate distribution is intended as an improvement on the previous one for the bulk of the distribution, as it matches a further moment, it has support $[\gamma(\tau^2), \infty)$, and hence it is not particularly suitable for obtaining the approximate CDF of Q for extremely small values, in particular those that are less than $\gamma(\tau^2)$.

The CDFs F_G and F_P provide easily computable approximations to F_Q. Neither requires the computation of the eigenvalues λ_i, and both may be evaluated using the CDF of the gamma distribution, available in standard statistical software.

3.3. Approximation three: a saddlepoint approximation

A further approximation that is expected to be more accurate in the tails of the distribution is the saddlepoint approximation, given in the present case by Kuonen [21] using the Barndorff–Nielsen

formulation. This requires the cumulant generating function of Q, denoted by $K(s)$, and its first two derivatives, given by

$$K(s)=-\frac{1}{2}\sum_{i=1}^{k-1}\log(1-2\lambda_i s),\quad K'(s)=\sum_{i=1}^{k-1}\frac{\lambda_i}{1-2\lambda_i s}\quad\text{and}\quad K''(s)=2\sum_{i=1}^{k-1}\left(\frac{\lambda_i}{1-2\lambda_i s}\right)^2$$

for $s<\frac{1}{2}\min_i \lambda_i^{-1}=\frac{1}{2}\lambda_1^{-1}$. To determine the saddlepoint approximating CDF $F_S(x;\tau^2)\approx P[Q\leqslant x]$, first solve the equation $K'(\hat{s})=x$ for $\hat{s}$, the solution being referred to as the saddlepoint. Next compute $a=\text{sign}(\hat{s})\sqrt{2[\hat{s}x-K(\hat{s})]}$ and $b=\hat{s}\sqrt{K''(\hat{s})}$. The saddlepoint approximation is then given by

$$F_S(x;\tau^2)=\Phi\left(a+\frac{1}{a}\log\left[\frac{b}{a}\right]\right)$$

where Φ is the standard normal CDF. Note that, as with the exact distribution for Q, the eigenvalues λ_i must be evaluated separately for each τ^2. Further, although Φ is readily computable using standard software, the saddlepoint equation $K'(\hat{s})=x$ must be solved for each x; standard root-solving routines are generally available, however, so this is not likely to be too great a burden on the implementation of the approximation. It may, indeed, be more accessible to researchers who do not have software implementing Farebrother's algorithm [18] for the positive linear combination of chi-squared random variables.

4. APPLICATIONS OF THE EXACT AND APPROXIMATE DISTRIBUTIONS OF Q

Now that the exact CDF of Q has been derived, along with suitable approximations, we turn our attention to some important applications of these. As noted in the introduction, the statistic Q can be used to measure the impact of heterogeneity, to test the null hypothesis $\tau^2=0$, and to provide confidence intervals for τ^2.

4.1. Measures of the impact of heterogeneity

Higgins and Thompson [6] proposed several measures of heterogeneity for meta-analysis, and one of the measures (I^2, below) has been adopted by the Cochrane Collaboration as the summary measure of heterogeneity in their Review Manager Software (RevMan 4.2 User Guide). Recently, Mittlböck and Heinzl [8] compared the properties of various measures using simulation. Since all of these statistics are functions of Q, their CDFs can be derived from that of Q. To ease presentation, let $v=k-1$. Define the heterogeneity measures

$$H_M^2=(Q-v)/v$$

$$I^2=100\times\frac{Q-v}{Q}=100\times\left(1-\frac{v}{Q}\right)$$

as given previously [6, 8]. In addition to these two measures, $H^2=Q/v=H_M^2+1$ has also been defined. This measure is not examined directly below, however, as it is merely a very simple function of H_M^2, and all its properties can be ascertained from this. Briefly, H^2 measures the relative excess of Q over its degrees of freedom, H_M^2 simply shifts H^2 so that homogeneity is at 0,

and I^2 is approximately 'the proportion of total variation in the estimates of treatment effect that is due to heterogeneity between studies' [6, p. 1552]. The measures H_M^2 and I^2 are typically defined, as here, without explicit truncation at 0, though this is then done when computed in application.

We thus have the following expressions for the CDFs of these heterogeneity measures in terms of F_Q, detailing the computation only for $F_{H_M^2}$:

$$F_{H_M^2}(x;\tau^2) = P[H_M^2 \leqslant x] = P[(Q-v)/v \leqslant x] = P[Q \leqslant v(x+1)] = F_Q(v(x+1);\tau^2)$$

$$F_{I^2}(x;\tau^2) = F_Q(v/(1-x/100);\tau^2)$$

where again we stress the dependence on τ^2. Exact properties of these heterogeneity measures, under the random effects model, may therefore be investigated directly. Replacing $F_Q(x;\tau^2)$ with $F_G(x;\tau^2)$, $F_P(x;\tau^2)$ or $F_S(x;\tau^2)$ provides the approximate CDFs of the heterogeneity measures using the corresponding approximation.

The exact means and variances of these measures can be computed, where necessary, using their CDFs with the identities for any continuous random variable X

$$E[X] = \int_0^\infty [1-F_X(x)]\,\mathrm{d}x - \int_{-\infty}^0 F_X(s)\,\mathrm{d}x$$

$$\mathrm{Var}[X] = \int_0^\infty 2x[1-F_X(x)+F_X(-x)]\,\mathrm{d}x - (E[X])^2$$

When a continuous random variable X is truncated at 0, as when H_M^2 and I^2 are used in practice, the mean and variance for the truncated version are available using the CDF of X from

$$E[X_+] = \int_0^\infty [1-F_X(x)]\,\mathrm{d}x \quad \text{and} \quad \mathrm{Var}[X_+] = \int_0^\infty 2x[1-F_X(x)]\,\mathrm{d}x - (E[X_+])^2$$

where $X_+ = X\mathrm{I}_{[0,\infty)}(X)$. In the applications below, these integrals are computed numerically using the implementation of Farebrother's algorithm for the exact distribution for Q for the appropriate CDF and the numerical integration routine provided in S-Plus (Insightful Corp., Seattle, WA); this may also easily be done in R (www.r-project.org).

4.2. Power of Cochran's test for the presence of heterogeneity

Under the hypothesis that $\tau^2 = 0$, Q has a chi-squared distribution with $k-1$ degrees of freedom [3, p. 39]. Hence in order to test $H_0 : \tau^2 = 0$, versus the alternative $H_1 : \tau^2 > 0$, Q is computed and compared with an appropriate critical value from $\chi^2(k-1)$. The exact power of this test as a function of τ^2 is $\beta(\tau^2;\alpha) = P[Q > \chi^2_{1-\alpha}(k-1)] = 1 - F_Q(\chi^2_{1-\alpha}(k-1);\tau^2)$, where $\chi^2_{1-\alpha}(k-1)$ is the $100(1-\alpha)$th percentile point of the $\chi^2(k-1)$ distribution. Approximate power functions $\beta_{\text{approx}}(\tau^2;\alpha)$, based on the approximating CDFs of the preceding section, may be computed using the appropriate approximate CDF in a similar manner as in the previous subsection. Jackson [13] used $F_G(\chi^2_{1-\alpha}(k-1);\tau^2)$ for this purpose and noted, for the case where $\sigma_i^2 = \sigma^2$ for all i, that this is exact and, if further that k is odd, the power can be expressed as a finite Poisson summation.

4.3. Confidence intervals for τ^2

Procedures for obtaining confidence intervals for τ^2 have previously been suggested [5, 9, 17, 22]. Using the exact CDF for Q, a corresponding exact confidence interval may be obtained. The untruncated version of the DerSimonian and Laird (DL) [4] estimator of τ^2, denoted by $\hat{\tau}^2_{\mathrm{DL}}$, is a linear function of Q, $\hat{\tau}^2_{\mathrm{DL}}=[Q-(k-1)]/c$, where $c=S_1-S_2/S_1$. Hence, following precisely the same procedure as Biggerstaff and Tweedie [5], a $100(1-\alpha_1-\alpha_2)$ per cent confidence interval is given by (τ^2_l, τ^2_u), where τ^2_l is the solution for τ^2 in $1-F_{\hat{\tau}^2}(\hat{\tau}^2_{\mathrm{DL}};\tau^2)=1-F_Q(c\hat{\tau}^2_{\mathrm{DL}}+k-1;\tau^2)=\alpha_1$ and τ_u is the solution for τ^2 in $F_{\hat{\tau}^2_{\mathrm{DL}}}(\hat{\tau}^2_{\mathrm{DL}};\tau^2)=F_Q(c\hat{\tau}^2_{\mathrm{DL}}+k-1;\tau^2)=\alpha_2$; these equations are solved numerically in practice and if $1-F_Q(c\hat{\tau}^2_{\mathrm{DL}}+k-1;0)>\alpha_1$ then the lower bound τ^2_l is truncated to zero. Approximating $F_Q(c\hat{\tau}^2_{\mathrm{DL}}+k-1;\tau^2)$ with $F_G(c\hat{\tau}^2_{\mathrm{DL}}+k-1;\tau^2)$ in these two equations gives the approximate confidence limits described previously [5, p. 757], and the other approximate distributions could also be used for this purpose.

Hartung and Makambi (HM) [23] developed positive variance estimators for one-way ANOVA and meta-analysis. In particular, their Corollary 1, expression (12), gives an easily computed estimator of the between-study variance in the one-way ANOVA model. When the within-study variances are assumed known, the parameters $b^*_i=\beta_i$ in expression (12) of HM; hence, we can then write their between-study variance estimator in our notation as

$$\hat{\tau}^2_{\mathrm{HM}}=\frac{Q^2}{c[2(k-1)+Q]}$$

Since $\hat{\tau}^2_{\mathrm{HM}}$ is a simple function of Q, we easily compute the exact CDF of $\hat{\tau}^2_{\mathrm{HM}}$ directly as

$$\begin{aligned}
F_{\hat{\tau}^2_{\mathrm{HM}}}(x;\tau^2) &= P\left[\frac{Q^2}{c[2(k-1)+Q]}\leqslant x\right] \\
&= P\left[\left(Q-\frac{cx}{2}\right)^2\leqslant 2(k-1)cx+\left(\frac{cx}{2}\right)^2\right] \\
&= F_Q\left(\frac{cx}{2}+\sqrt{2(k-1)cx+\left(\frac{cx}{2}\right)^2}\right)-F_Q\left(\frac{cx}{2}-\sqrt{2(k-1)cx+\left(\frac{cx}{2}\right)^2}\right)
\end{aligned}$$

As above, a $100(1-\alpha_1-\alpha_2)$ per cent confidence interval $(\tau^2_{l,\mathrm{HM}}, \tau^2_{u,\mathrm{HM}})$ for τ^2 may be obtained as the solutions to $1-F_{\hat{\tau}^2_{\mathrm{HM}}}(\hat{\tau}^2_{\mathrm{HM}};\tau^2)=\alpha_1$ and $F_{\hat{\tau}^2_{\mathrm{HM}}}(\hat{\tau}^2_{\mathrm{HM}};\tau^2)=\alpha_2$, respectively. The lower confidence limit (LCL) is truncated to zero in the same manner as for the corresponding interval based on the more conventional DL estimate.

5. EXACT PROPERTIES OF THE HETEROGENEITY MEASURES

Perhaps the most important use of Q is to quantify the impact of heterogeneity; as noted above, I^2 is now routinely provided by the systematic reviews reported by the Cochrane Collaboration. Mittlböck and Heinzl [8] (MH) recently performed a simulation study to examine the sampling properties of the various measures described in Section 4; now that exact distributions are available for these, one may undertake such investigations free of Monte Carlo error.

Following Hardy and Thompson [7], MH simulated the power of Cochran's heterogeneity test and the means of I^2 and H_M^2 for different numbers of studies and within-study variances. To illustrate our methods, we compute the exact power and means for their Scenario 2 (Table I of MH), in which they considered sub-scenarios by setting the numbers of studies and the within-study variances as $k=5, 10, 20$ and $\sigma^2=0.05, 0.1, 0.2$, respectively. In their study, a measure of 'typical' within-study variance [6] that was useful to assess the impact of within-study variance on the power of the homogeneity test was, in their notation,

$$\hat{\sigma}^2_{W,2}=\frac{(k-1)\sum_i w_i}{(\sum_i w_i)^2-\sum_i w_i^2} \tag{3}$$

and this was used in plotting power and the means of I^2 and H_M^2 to aid interpretation. The particular scenario we reproduce held constant measures of total information (as defined in MH).

In Figure 1, we update the graphs of the power of Cochran's test for heterogeneity (using a significance level of 0.05) and the means given in MH's Figure 1, Scenario 2, panels (g)–(l), using exact calculations as described in Sections 3, 4.1 and 4.2. To match MH's analysis, in Figure 1, we use the un-truncated versions of I^2 and H_M^2 to compute the means in the first two columns,

Table I. Heterogeneity measures for the example data sets.

Data	k	$\hat{\tau}^2_{\mathrm{DL}}$ (p-value)	$\hat{\sigma}^2_{W,2}$	Statistic	Trunc	Q	H_M^2	I^2
Aspirin	6	0.027	0.028	Observed		9.87	0.97	49.33
		(0.079)		Mean		9.87	0.97	13.74
				Mean	X	—	1.06	37.05
				Std. Dev.		6.49	1.30	44.34
				Std. Dev.	X	—	1.21	27.96
Diuretic	9	0.230	0.095	Observed		27.26	2.41	70.66
		(0.001)		Mean		27.26	2.41	57.69
				Mean	X	—	2.42	59.84
				Std. Dev.		16.57	2.07	28.05
				Std. Dev.	X	—	2.06	23.13
Glycerol	9	0.079	0.252	Observed		10.50	0.31	23.80
		(0.232)		Mean		10.50	0.31	−2.11
				Mean	X	—	0.42	21.70
				Std. Dev.		5.32	0.66	30.90
				Std. Dev.	X	—	0.56	22.10
Sclerotherapy	19	0.302	0.234	Observed		41.27	1.29	56.39
		(0.001)		Mean		41.27	1.29	50.38
				Mean	X	—	1.30	50.85
				Std. Dev.		14.52	0.81	19.03
				Std. Dev.	X	—	0.80	17.74

Included are the number of studies (k); the DerSimonian–Laird estimate of the heterogeneity variance ($\hat{\tau}^2_{\mathrm{DL}}$) and Cochran's test for heterogeneity (p-value); the 'typical' within-study variance ($\hat{\sigma}^2_{W,2}$); and the observed values, exact means and exact standard deviations of the heterogeneity statistics Q, H_M^2 and I^2 evaluated at the observed value of $\hat{\tau}^2_{\mathrm{DL}}$. The 'Trunc' column indicates (X) that the values are for the truncated versions.

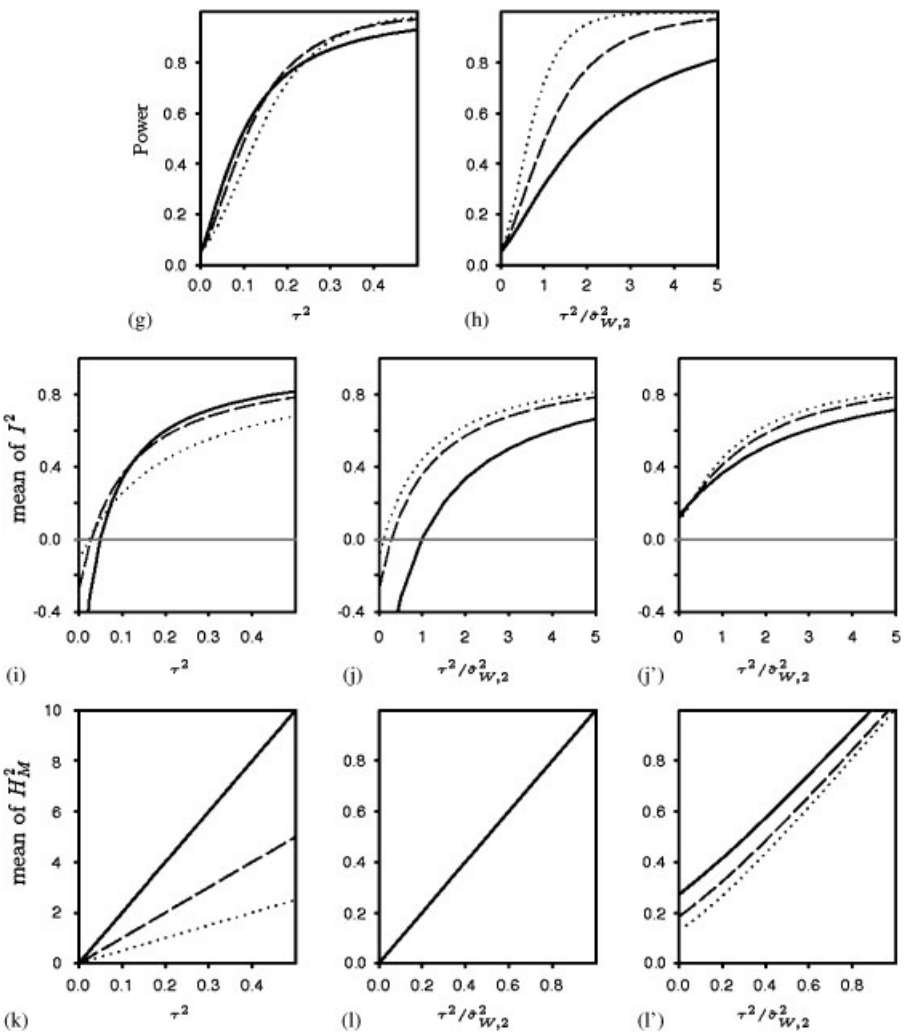

Figure 1. Update using the exact distributional results of Mittlböck and Heinzl [8] simulation study of power, mean of I^2 and mean of H^2_M, plotted against τ^2 (first column) and $\tau^2/\hat{\sigma}^2_{W,2}$ (second column), where $\hat{\sigma}^2_{W,2}$ is given by equation (3). To match Mittlböck and Heinzl, the solid line, dashed and dotted lines correspond to the sub-scenarios outlined in their text for Scenario 2, and the panel labels (g)–(l) correspond to Mittlböck and Heinzl's Figure 1; panels labeled (j′) and (l′) in column three present the means of the truncated values for I^2 and H^2_M. Note that the scale for panels (l) and (l′) has been adjusted from the original MH to make it easier to see the impact of truncation.

where I^2 is expressed as a decimal as in MH, and we label our graphs in Figure 1 as (g)–(l), i.e. omit (a)–(f), again for consistency with the corresponding figure of MH. In the final column of the figure, we include the truncated versions of I^2 and H^2_M to provide a comparison to these measures as used in practice. Of course truncating the measures to zero necessarily increases the

expectations, and note that we have adjusted the scales in panels (l) and (l′) from the original MH scales to show this clearly.

Interpretation of MH's graph of Scenario 2 is complicated somewhat by both Monte Carlo error and the fact that they evaluated the curves at most 11 points (for the power plot in MH's panel (g), these range from 0 to 0.5 by steps of 0.05). In contrast, our computations based on the exact calculations yield visually smooth curves, computed at 100 points along the horizontal axes. All of the results are in agreement with those of MH, however, and in particular when $\tau^2=0$, I^2 has a scaled, shifted inverse-chi-squared distribution, so that $E[I^2]=100\times[-2/(k-3)]$ for $k>3$ [8].

6. EXAMPLES

We further exhibit our methods with data from four published meta-analyses that have been used previously to illustrate statistical methodological developments. The four data sets are

- Aspirin and Heart attack: In which studies evaluated the potential benefit of aspirin use following heart attack [24].
- Diuretic and Pre-eclampsia: In which studies evaluated the prevention of pre-eclampsia with use of a diuretic [25].
- Glycerol for Acute stroke: In which studies evaluated the use of glycerol for preventing death in patients suffering from an acute stroke [13].
- Sclerotherapy and Cirrhosis and Oesophagogastric varices: In which studies evaluated the effectiveness of endoscopic sclerotherapy for preventing death in patients with cirrhosis and oesophagogastric varices [26–28].

For brevity, we do not reproduce the data here; they are available in the references provided and from the authors on request. All response variables have been converted, where necessary, to empirical log-odds ratios and their corresponding standard errors. It should be noted, however, that some studies in these examples are of modest size, and this should be borne in mind when interpreting the 'exact' results that follow; the random effects model (1), with σ_i^2 assumed known for all studies, is an approximation in reality. Note also that the distribution of Q depends on τ^2, and hence so do the distributions of the heterogeneity measures and the power of Cochran's test. As described below, *DerSimonian and Laird estimates of τ^2 are used to evaluate these quantities for each meta-analysis in Sections 6.1 and 6.2*, in order to obtain some illustrative results for these examples.

6.1. Measures of the impact of heterogeneity

The first row corresponding to each meta-analysis in Table I records the numbers of studies, along with the DL estimate of τ^2, the 'typical' within-study variance $\hat{\sigma}^2_{W,2}$ and the observed values of Q and the various heterogeneity measures discussed. The heterogeneity measures and tests show that these four examples include meta-analyses across a range of degrees of heterogeneity as, for example, I^2 ranges from roughly 24 per cent for the Glycerol data set to 71 per cent for the Diuretic data set.

Also shown in Table I are the means and standard deviations of the heterogeneity measures, where X denotes that the truncated value has been used. These quantities are evaluated using their exact distributions and using the DL estimates of τ^2. The relative values for the truncated versions

of H_M^2 and I^2 reflect the truncation at 0, as the means increase and the standard deviations decrease relative to their non-truncated counterparts. This is seen most dramatically for the Glycerol data set, the least heterogeneous, where $E[I^2] = -2.11$ and $E[I^2 \mathrm{I}_{[0,\infty)}] = 21.70$. The standard deviations of the various measures of heterogeneity are generally very large relative to the magnitude of the estimates, reflecting the skewness of the distributions involved and the imprecise estimation for examples with small numbers of studies.

6.2. *The power of Cochran's test*

In order to explore the power of Cochran's test for the presence of heterogeneity, for each of the example data sets we plot in the left-hand column of Figure 2 the exact and approximate CDFs of Q (using the DL estimates for τ^2 in order to evaluate the quantities required). Having chosen a significance level and therefore a critical value for the test, the left-hand column of Figure 2 can be used to give an indication of the power of the test. In the corresponding right-hand column, we further illustrate the error in utilizing the various approximations by plotting the percentage relative error (PRE), obtained as $100[F_{\mathrm{approx}}(x; \hat{\tau}^2_{\mathrm{DL}}) - F_Q(x; \hat{\tau}^2_{\mathrm{DL}})]/F_Q(x; \hat{\tau}^2_{\mathrm{DL}})$, against $F_Q(x; \hat{\tau}^2_{\mathrm{DL}})$. This again uses the DL estimates of τ^2 when evaluating the CDFs of Q, over a wide range of values of x.

For the Diuretic data set, the CDF panels (first column) show that the moment approximations do not track the exact CDF particularly well, especially for small values. This is seen more clearly in the PRE panels (second column), where we see as much as 10 per cent relative error for the two-moment approximation, and 5 per cent relative error for the three-moment approximation. Note that the saddlepoint approximation appears generally better than both moment approximations, even though theoretically it is expected to perform best in the tail. For the Glycerol and Sclerotherapy data sets in particular, little difference is found among the approximations compared with the exact CDF, with PREs under 1 per cent for nearly the whole range of the distributions.

These observations are consistent with the expectation that the approximations should suffer as heterogeneity increases. Recall the heterogeneity measures (Table I) indicated that the Glycerol data set was least heterogeneous, whereas the Diuretic data set was the most. This is reflected in the CDF/PRE graphs, where one sees that the approximations are best (smallest PRE) for the Glycerol data set and worst (largest PRE) for the Diuretic data set. It is however interesting to note that the Aspirin and the Sclerotherapy data sets have similar I^2 but the approximations are better for the latter, which may be a reflection of the larger number of studies in the Sclerotherapy data.

As shown in Figure 3, analogous results are obtained when computing power functions (as a function of τ^2) for Cochran's heterogeneity test, using a significance level of 0.05. Here in particular we see the accuracy of the saddlepoint method, as the power curves only rely on this approximation in the tail of the distribution, where theoretical accuracy is $O(n^{-3/2})$ [21].

The relative performance of the approximations appears consistent across the examples. The saddlepoint approximation provides, as expected, the best performance. The three-moment approximation appears to match the right tail of the exact distributions better than the two-moment approximation, but provides similar errors as this at the left-tail of the distribution. The implications of these errors are illustrated further in the next section where confidence intervals (CIs) are computed.

6.3. *Confidence intervals*

Table II records 95 per cent CIs computed from the example data sets, where LCL and UCL denote the lower and upper confidence limits, respectively. We computed symmetric CIs in

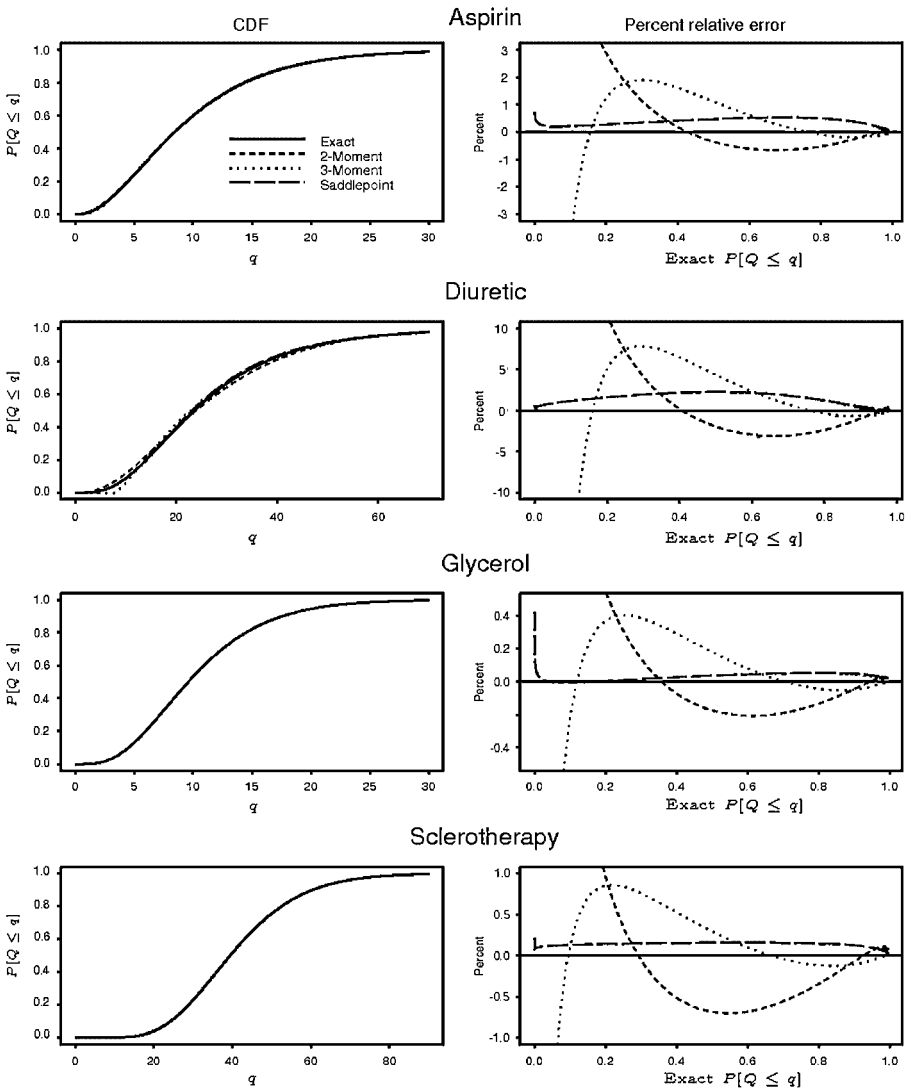

Figure 2. Cumulative distribution functions (CDF) and associated percentage relative errors (PRE) for the example data sets, evaluated with τ^2 equal to the DerSimonian–Laird estimator, $\hat{\tau}^2_{\mathrm{DL}}$. Note that the vertical scales differ among the PRE panels.

these examples, taking $\alpha_1 = \alpha_2 = 0.025$, but other choices for α_1, α_2 are possible, such as those yielding so-called minimum or shortest length intervals. For comparison, the associated point estimates for τ^2 are given, as are relative values for the CI limits (lower and upper endpoints separately), and CI length and relative CI length. For brevity, we have included the approximating distribution function-based CIs for the DL estimator only; they could be used with the HM estimator as well, for which qualitatively similar results are expected. The LCLs for the

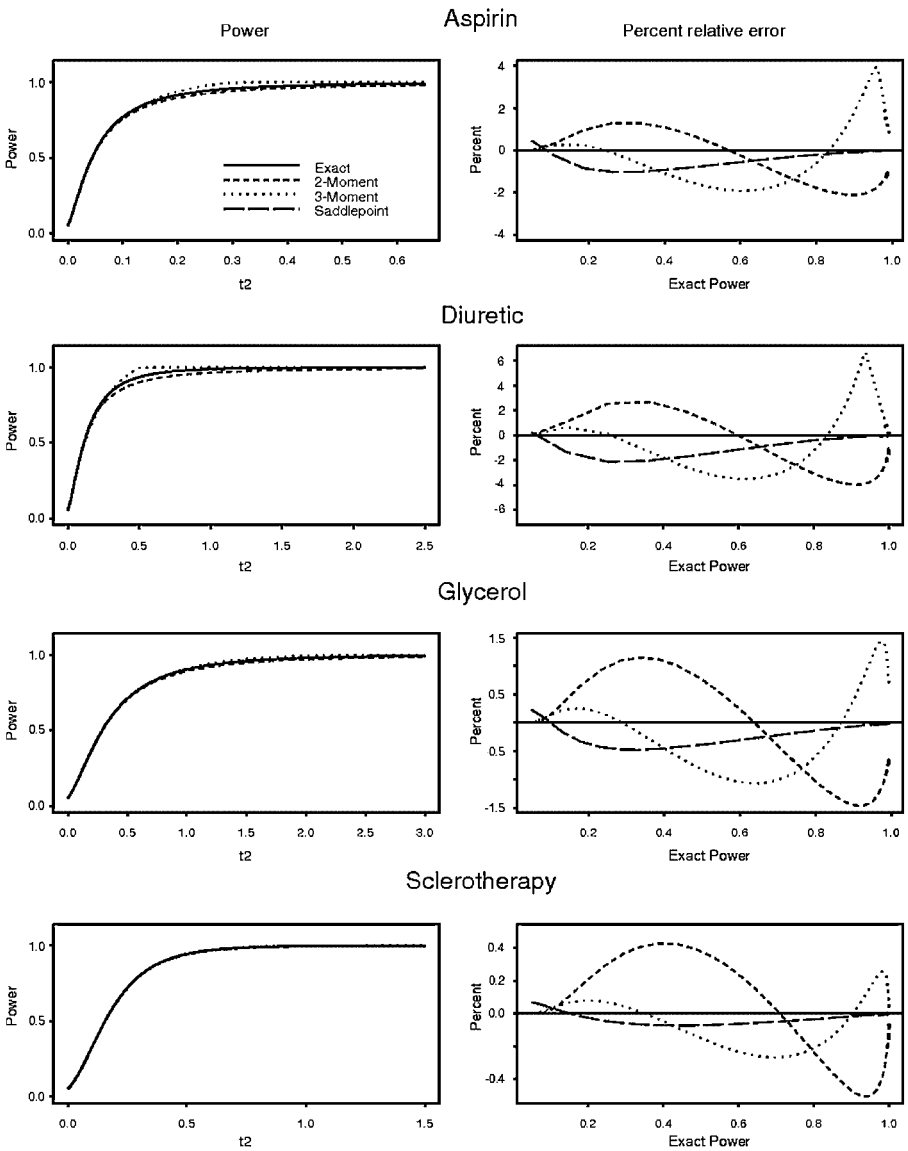

Figure 3. Power functions and associated percentage relative errors (PRE) for the example data sets, plotted against τ^2. Power was computed with size $\alpha=0.05$, as reflected in the power curves equalling 0.05 at $\tau^2=0$. Note that the vertical scales differ among the PRE panels.

Aspirin and Glycerol data sets are truncated to zero as $1-F_Q(c\hat{\tau}^2_{\mathrm{DL}}+k-1;0)>0.025$ for these two examples.

In all four examples, the CI using the three-moment approximation is narrower than the exact DerSimonian and Laird (Exact-DL) CI, due to each time the relatively small value for the UCL,

Table II. Ninety-five per cent confidence intervals for τ^2 for the example data sets. Point estimates $\hat{\tau}^2$ are the DerSimonian-Laird (DL) estimator $\hat{\tau}^2_{DL}$ for CI methods two-, three-moment, Saddlepoint and Exact-DL; and Hartung and Makambi's estimator $\hat{\tau}^2_{HM}$ for the CI method Exact-HM derived using this statistic. Relative (Rel.) values are with respect to the Exact-DL method.

Data	$\hat{\tau}^2$	CI method	LCL	UCL	Rel. LCL	Rel. UCL	CI length	Rel. CI length
Aspirin	0.027	Two-moment	0	0.451	—	1.328	0.451	1.328
	0.027	Three-moment	0	0.219	—	0.646	0.219	0.646
	0.027	Saddlepoint	0	0.340	—	1.003	0.340	1.003
	0.027	Exact-DL	0	0.339	—	1	0.339	1
	0.027	Exact-HM	0	0.338	—	0.995	0.338	0.995
Diuretic	0.230	Two-moment	0.048	2.355	1.016	1.646	2.307	1.667
	0.230	Three-moment	0.047	0.875	0.987	0.611	0.828	0.598
	0.230	Saddlepoint	0.049	1.437	1.029	1.004	1.388	1.003
	0.230	Exact-DL	0.047	1.431	1	1	1.384	1
	0.205	Exact-HM	0.058	1.570	1.228	1.097	1.512	1.093
Glycerol	0.079	Two-moment	0	1.340	—	1.192	1.340	1.192
	0.079	Three-moment	0	0.952	—	0.846	0.952	0.846
	0.079	Saddlepoint	0	1.127	—	1.002	1.127	1.002
	0.079	Exact-DL	0	1.124	—	1	1.124	1
	0.131	Exact-HM	0	0.733	—	0.652	0.733	0.652
Sclerotherapy	0.302	Two-moment	0.071	1.075	1.004	1.050	1.004	1.054
	0.302	Three-moment	0.071	0.997	1.000	0.974	0.926	0.972
	0.302	Saddlepoint	0.071	1.024	1.001	1.000	0.953	1.000
	0.302	Exact-DL	0.071	1.023	1	1	0.953	1
	0.286	Exact-HM	0.082	1.072	1.157	1.048	0.991	1.040

as each LCL for this interval is nearly equal to the Exact-DL LCL. In contrast, the CI using the two-moment approximation is relatively wider, owing to the relatively large UCL as, again, the LCL nearly agrees with the Exact-DL LCL. In each case, the saddlepoint approximating interval is nearly the same as the Exact-DL interval. Except for the Glycerol data set, which is the least heterogeneous, the exact CI of Hartung and Makambi (Exact-HM) agrees very well with the Exact-DL CI. In contrast, for the Glycerol data set, the Exact-HM CI is shortest. These results show that circumstances can easily be found where the various approximations have considerable implications for the resulting confidence intervals, suggesting that the extra effort required when using the exact distribution of Q or the saddlepoint approximation, which is nearly exact, can be worthwhile.

We note here one computational inconvenience. The numerical root-finding routine `uniroot` in S-Plus (Insightful Corp., Seattle) cannot, by design, call functions that themselves call `uniroot`. Because of this, S-Plus's `uniroot` function may not be used to find CIs using the saddlepoint approximation, which requires such iterative calls, because of the need to solve the saddlepoint equation for each new value of τ^2 and also to solve the equations that provide the confidence limits. The corresponding function `uniroot` in R (www.r-project.org) does not have this difficulty, so R may be used to find CIs using the saddlepoint approximation; in our examples, we used R for this computation, while using S-Plus for all the others.

7. DISCUSSION

We have derived the exact distribution of Cochran's Q statistic and have shown how this can be used with regard to all its various applications in meta-analysis. The theory developed has been demonstrated in both practical and more theoretical contexts: in particular, the distribution provides insights into the power of the standard test for presence of heterogeneity and the sampling distributions of measures of the impact of this, and can also be used to provide confidence intervals for τ^2. The methodology is therefore of interest to all who use standard meta-analysis techniques, irrespective of their preference concerning the various uses of Q. The exact distributional result for Q and the related results extend naturally to standard meta-regression under the normal model, by a simple modification to the matrix formulation utilized here, and this may form the subject of future work.

Despite the broad applicability of the theory, it should be repeated that the standard random effects model has been assumed throughout. This model requires sufficiently large studies, so that the Central Limit Theorem can be invoked to justify approximate within-study normality; the random effects are also assumed to be normally distributed. Since the studies can only ever be finite in size and are often, in fact, relatively small, the 'exact' distribution only provides an approximation to the true distribution of Q in reality.

Despite this, situations in meta-analysis are frequently found where the studies are regarded as sufficiently large to make the standard assumptions, and the use of the exact distribution for Q in these cases removes one layer of approximation. In situations where studies are of more modest size, meta-analysts can choose study outcomes, and accompanying measures of treatment effect, where the usual within-study assumptions are most appropriate. Meta-analyses based on many small studies are also fairly commonplace, however, and this is a scenario where the random effects model is unlikely to provide a reasonable approximation. Further work should allow for the variation in the estimates of the within-study variances, perhaps by incorporating a simulation study, which could reflect the binary nature of the outcomes used in the four examples examined in Section 6. It is of considerable interest, therefore, to see how robust the distribution of Q is to the two assumptions of normality and this provides an avenue for further research.

The approximations we have evaluated were found to perform well for moderate degrees of between-study variance, but somewhat less satisfactorily for examples with more considerable heterogeneity. Although the approximations require further investigation before a definitive statement can be made, it seems that the use of the exact distribution may be especially desirable for examples with very large degrees of heterogeneity. Of course, some may doubt the appropriateness of conducting a meta-analysis under such circumstances and it may prove that the approximations are suitable in situations where most would contemplate pooling the results in the usual way. In particular, the saddlepoint approximation may be sufficiently accurate for most applications, which merely require accuracy in the tails of the distribution of Q. Even though the saddlepoint method is an approximation, our implementation of the exact distribution was computationally faster than the saddlepoint approximation, due to the extra effort required to solve the saddlepoint equation for each value of τ^2 and so each set of eigenvalues λ_i.

Finally, now that an exact distribution of Q is available, attention is inevitably drawn toward other statistics used in meta-analysis. For example, the estimate of treatment effect is conventionally assumed normally distributed with known variance but this is merely justified asymptotically, as the number of studies tends toward infinity. For real meta-analyses, such as the four examined here,

the number of studies is typically somewhat more modest and the accuracy of this approximation is brought into question. The exact distribution of this and other statistics routinely used in meta-analysis awaits a full investigation.

REFERENCES

1. Lee KJ, Thompson SG. Flexible parametric models for random effects distributions. *Statistics in Medicine* 2008; **27**:418–434. DOI: 10.1002/sim.2897.
2. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**:101–129.
3. Sutton AJ, Abrams KR, Jones DR, Sheldon DR, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: New York, 2002.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
5. Biggerstaff BJ, Tweedie RL. Incorporating variability of estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine* 1997; **16**:753–768.
6. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**: 1539–1558.
7. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 1998; **17**:841–856.
8. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Statistics in Medicine* 2006; **25**:4321–4333.
9. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
10. Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A. Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics* 2002; **3**:445–457.
11. Malzahn U, Böhning D, Holling H. Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika* 2000; **87**:619–632.
12. Hartung J, Knapp G. An alternative test procedure for meta-analysis. In *Meta-analysis*, Schulze H, Holling H, Böhning D (eds). Hogrefe & Huber: Göttingen, 2003.
13. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine* 2006; **25**:2688–2699.
14. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 1998; **3**(4): 486–504.
15. Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychological Methods* 2001; **6**(3):203–217.
16. Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychological Methods* 2004; **9**(4):426–445.
17. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine* 2007; **26**:37–52.
18. Farebrother RW. Algorithm AS 204: the distribution of a positive linear combination of χ^2 random variables. *Applied Statistics* 1984; **33**(3):332–339.
19. Marsaglia G. Evaluating the normal distribution. *Journal of Statistical Software* 2004; **11**(4):1–11.
20. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 1946; **2**(6):110–114.
21. Kuonen D. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika* 1999; **86**:929–935.
22. Knapp G, Biggerstaff BJ, Hartung J. Assessing the amount of heterogeneity in random-effects meta-analysis. *Biometrical Journal* 2006; **48**:271–285.
23. Hartung J, Makambi KH. Positive estimation of the between-group variance component in one-way ANOVA and meta-analysis. *South African Statistical Journal* 2002; **36**:55–76.
24. Draper D, Gaver Jr DP, Goel PK, Greenhouse JB, Hedges JV, Morris CN, Tucker JR, Waternaux CM. *Combining Information*: *Statistical Issues and Opportunities for Research*. National Academy Press: Washington, DC, 2002.
25. Collins R, Yusuf S, Peto R. Overview of randomized trials of diuretics in pregnancy. *British Medical Journal* 1985; **290**:17–23.

Published in 2008 by John Wiley & Sons, Ltd. *Statist. Med.* 2008; **27**:6093–6110
DOI: 10.1002/sim

26. Pagliaro L, D'Amico G, Sorensen TIA, Lebrec D, Burroughs AK, Morabito A, Tiné F, Politi F, Traina M. Prevention of first bleeding in cirrhosis: a meta-analysis of randomized trials of nonsurgical treatment. *Annals of Internal Medicine* 1992; **117**:59–70.
27. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
28. Jackson D. The implications of publication bias for meta-analysis' other parameter. *Statistics in Medicine* 2006; **25**:2911–2921.

Published in 2008 by John Wiley & Sons, Ltd. *Statist. Med.* 2008; **27**:6093–6110
DOI: 10.1002/sim