

Accepted Manuscript

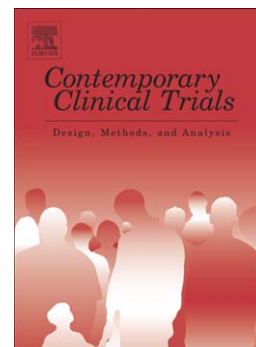
Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model

Suhail A.R. Doi, Jan J. Barendregt, Shahjahan Khan, Lukman Thalib, Gail M. Williams

PII: S1551-7144(15)30008-2
DOI: doi: [10.1016/j.cct.2015.05.010](https://doi.org/10.1016/j.cct.2015.05.010)
Reference: CONCLI 1196

To appear in: *Contemporary Clinical Trials*

Received date: 5 February 2015
Revised date: 12 May 2015
Accepted date: 15 May 2015



Please cite this article as: Doi Suhail A.R., Barendregt Jan J., Khan Shahjahan, Thalib Lukman, Williams Gail M., Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model, *Contemporary Clinical Trials* (2015), doi: [10.1016/j.cct.2015.05.010](https://doi.org/10.1016/j.cct.2015.05.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model

Suhail A. R. Doi¹, Jan J. Barendregt^{2,3}, Shahjahan Khan⁴, Lukman Thalib⁵, Gail M. Williams³

¹Research School of Population Health, Australian National University, Canberra, ²Epigear International, Sunrise Beach, Australia, ³School of Population Health, University of Queensland, Brisbane, Australia, ⁴School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Toowoomba, Australia, ⁵Department of Community Medicine, Kuwait University, Kuwait

Correspondence to

Suhail A. R. Doi
Research School of Population Health
Australian National University
Mills Road,
Canberra ACT 2601,
Australia
T +61 404 181134
E sardo@gmx.net

Abstract

This article examines the performance of the updated quality effects (QE) estimator for meta-analysis of heterogeneous studies. It is shown that this approach leads to a decreased mean squared error (MSE) of the estimator while maintaining the nominal level of coverage probability of the confidence interval. Extensive simulation studies confirm that this approach leads to maintenance of the correct coverage probability of the confidence interval, regardless of the level of heterogeneity. It also retains a lower observed variance compared to the random effects (RE) model. The QE model is robust to subjectivity in quality assessment down to completely random entry, in which case its MSE equals that of the RE estimator. When the proposed QE method is applied to a meta-analysis of magnesium for myocardial infarction data, the pooled mortality odds ratio (OR) becomes 0.81 (95% CI 0.61 – 1.08) which favours the larger studies but also reflects the reduced uncertainty around the pooled estimate. In comparison, under the RE model, the pooled mortality OR is 0.71 (95% CI 0.57 – 0.89) which is less conservative than that of the QE results. The new estimation method has been implemented into the free meta-analysis software MetaXL which allows comparison of alternative estimators and can be downloaded from www.epigear.com.

Keywords: Fixed effects, heterogeneity, meta-analysis, quasi-likelihood , quality effects, random effects

1. INTRODUCTION

In 2008, we had provided a solution to the problems with the random effects (RE) model for meta-analysis using a quality weighted model which we called the quality effects (QE) model [1,2]. In the previous paper in this series, we also discussed a variant of the QE model called the inverse variance heterogeneity (IVhet) model that does not require quality assessment because all studies by default are assigned the same quality [3]. The initial problem was that, as heterogeneity increases, the coverage of the RE confidence interval drops well below the nominal level [4], substantially underestimates the statistical error and produces overconfident conclusions [5,6]. In addition, we believe that the way the RE model modification of the inverse variance weights are conceptualized [7] lacks justification according to a strict view of randomization in statistical inference [8]. We therefore introduced these alternative models in an attempt to lower the estimator mean squared error and obtain a correct coverage of the confidence interval that keeps to the nominal level across different degrees of heterogeneity [1,3].

We now demonstrate that input of quality into the model can markedly improve the performance measures of the estimator as compared with the conventional random effects estimator or the IVhet estimator that replaces it [3]. Additionally, because quality is often viewed with suspicion as extremely subjective, the performance measures are obtained after subjecting the quality input to various degrees of random variation (at the point of input to the model) to see how this affects the estimator performance. The QE model examined in this paper updates the QE model of meta-analysis proposed in 2008 [2] in two important respects

[1]. First, overdispersion observed with the initial estimator has been corrected using an intra-class correlation based multiplicative scale parameter. Second, the quality scores were originally re-scaled between 0 & 1 for input into the model. Currently, they are still rescaled between 0 & 1 but then each of these rescaled scores is divided by the maximum value of the rescaled scores within the meta-analysis before it is input into the model. This still keeps the scores in the 0 – 1 range but now allows them to reflect the relative nature of these scores, i.e., relative to the best study in the meta-analysis. This will be discussed further in the next section.

2. DIFFERENCE BETWEEN THE RANDOM AND QUALITY EFFECTS WEIGHTED MEANS

Consider a collection of k independent studies, the j th of which has estimated effect size $\hat{\delta}_j$ which varies from its true effect size, δ_j through random error. Also consider that the true effects, δ_j , also vary from an underlying common effect, θ , through bias. This bias would include the possibility of some diversity of true effects (which remain similar) across studies (in which case θ would simply be the mean of the true (unbiased) effects). A greater diversity that leads to dissimilarity of effects would not be meta-analysed [9].

As previously described [3], the RE model weighted estimator, $\hat{\theta}_{RE}$, has weights given by:

$$\hat{w}_j = \frac{1}{v_j + \gamma^2}, \quad \{1\}$$

and weights that sum to 1 are given by

$$w_j = \frac{1}{\nu_j + \gamma^2} \bigg/ \sum_{j=1}^k \frac{1}{\nu_j + \gamma^2}, \quad \{2\}$$

with the sampling error variance of the j th study being ν_j and γ^2 being the methods of moment based between study variance estimate [7].

For the QE model, the weighted estimator, $\hat{\theta}_{QE}$, has weights that are adjusted from inverse variance weights based on the additional variance contribution from internal study biases (ϕ_j^2). Thus the QE weights are now given by:

$$\hat{w}_j'' = \frac{1}{\nu_j + \phi_j^2} \quad \{3\}$$

The problem is that ϕ_j^2 in expression {3} is unknown and thus a “synthetic” value for it needs to be created that keeps the proportional weights across studies similar to what would be expected with the true weights across studies in the meta-analysis. This works because we use normalized (sum to 1) weights in meta-analysis. One way of creating a synthetic value for ϕ_j^2 (and its estimator $\hat{\phi}_j^2$) is through the use of quality scores, derived from each study. If these quality scores are expressed on a scale between 0 and 1 by dividing by the maximum possible value under the scale, we can interpret the rescaled quality of each individual study as an intra-class correlation (q_j) and thus relate it back to ϕ_j^2 by considering this as the proportion of variance from between study bias (ϕ^2) not related to variance from internal study bias (ϕ_j^2) [10]. Thus $q_j = \phi^2 / (\phi^2 + \phi_j^2)$. Since q_j is also scale dependent, it can be expressed in relative terms that converts it into a study rank that starts from 1 in every meta-analysis and is now

scale independent. This is done by expressing q_j relative to the maximum ($q_{j\max}$) in the list of studies (which occurs in the study with the minimum internal bias variance, $\phi_{j\min}^2$). If we denote this study rank as Q_j , this equals the ratio $q_j/q_{j\max}$ and Q_j can then be expanded as follows:

$$Q_j = \frac{\phi^2}{\phi^2 + \phi_j^2} \bigg/ \frac{\phi^2}{\phi^2 + \phi_{j\min}^2} = \frac{(\phi^2 + \phi_{j\min}^2)}{(\phi^2 + \phi_{j\min}^2) + \hat{\phi}_j^2} \quad \{4\}$$

In expression {4}, the synthetic bias variance, $\hat{\phi}_j^2$, equals $\phi_j^2 - \phi_{j\min}^2$ and starts from zero and goes asymptotically to infinity. Similarly study rank, Q_j starts from 1 (since the best study has $\phi_j^2 = \phi_{j\min}^2$) and goes asymptotically to zero since $\lim_{\phi_j^2 \rightarrow \infty} Q_j = 0$. If we believe that the contribution of variance due to bias is not necessarily more costly than the contribution of variance due to chance (and is certainly not infinitely more costly otherwise we would be limiting this discussion to unbiased estimators) [11], then Q_j can be used to add the incremental contribution of variance due to bias to the variance due to chance and thus the synthetic sum of variances ($\hat{\phi}_j^2 + \nu_j$) is given by ν_j/Q_j , and ranges between ν_j (when Q_j is one) and increases asymptotically towards infinity (when Q_j is close to zero). Thus, by scaling the inverse variance weight proportional to Q_j we do not generate the true weight (which is unknown) but we do generate relative weights that should maintain the same proportional relationship to each other that would have been expected had the true (but unknown) weights been applied. This “synthetic” weight is given by:

$$\hat{w}_j'' \approx \frac{1}{v_j + \hat{\phi}_j^2} = \frac{Q_j}{v_j} \quad \{5\}$$

Keep in mind that \hat{w}_j'' in expressions {3} and {5} are not the same. To generate a weight that is influenced by both random chance and internal study bias thus requires weights to drop from $1/v_j$ to $\frac{Q_j}{v_j}$ due to the influence of variance contributed by internal study bias. An important aspect to keep in mind is that there is no attempt here at bias quantification and there is no connection between a quality score and the magnitude or direction of change in an effect size. This model utilizes the fact that variance due to bias can be modeled through a quality score.

The decrease in the weight defined by expression {5} (from inverse variance weight) in each study will therefore be by an amount given by:

$$\frac{1}{v_j} - \frac{Q_j}{v_j} = \frac{1-Q_j}{v_j} \quad \{6\}$$

To reduce the inherent bias in any weighting scheme, we can pool the weight decrement across studies using expression {6}, and this is given by $\sum_{j=1}^k [(1-Q_j)/v_j]$, which then can be split into k parts whose size is *proportional to* Q_j and added to each study's weight so that the sum of weights remains unchanged from the sum of the inverse variance weights. This adjustment was first proposed in 2009 [2] and modified slightly subsequently [12]. This is an additional measure to reduce estimator bias because through this step we can achieve bias reduction by decreasing the correlation between weights and the study effect. The advantage gained from doing this

additional procedure is depicted in the simulation results below. The latter adjustment is made by adding a quantity, $\hat{\tau}_j$ (see appendix for computation), and the final QE model weight (\hat{w}_j'') for each study is given by:

$$\hat{w}_j'' = \frac{Q_j}{v_j} + \hat{\tau}_j \quad \{7\}$$

and the weights that sum to 1 are given by

$$w_j'' = \left(\frac{Q_j}{v_j} + \hat{\tau}_j \right) / \sum_{j=1}^k \left(\frac{Q_j}{v_j} + \hat{\tau}_j \right)$$

While the weights in expression {7} will not equal the expected value of the unknown weight which is $1 / (v_j + \phi_j^2)$, the relative distribution of weights that sum to 1 should now be correct to a large extent. Since we are only interested in relative weights (that sum to 1) this works quite well to decrease estimator variance beyond that achievable through inverse variance weights alone.

3. VARIANCE OF THE ESTIMATOR UNDER DIFFERENT MODELS

The difference between the RE model and QE model is that the former has all ϕ_j^2 replaced by γ^2 and thus $1 / (v_j + \gamma^2) = \hat{w}_j$ and there is a decreasing capacity to minimize error due to sampling variability by the weights as heterogeneity increases and weights equalize. In the case of the QE model, $\frac{Q_j}{v_j} + \hat{\tau}_j = \hat{w}_j''$ and thus when Q_j varies across studies, this estimator will discount studies with both greater random chance as well as internal study bias. The QE

estimator will thus be expected, with increasing heterogeneity, to have a lower true variance (true variance is estimated through simulation) and therefore MSE than the RE estimator.

The theoretical model variance (theoretical variance is derived through the model and given in expressions {9} below) is computed by considering only variance from random error and modeling overdispersion through a quasi-likelihood approach [13,14] as previously described [3]. This implies that the meta-analysis is performed under a fixed effect assumption ($\gamma^2 = 0$) where γ^2 is a moment-based estimate of the between-studies variance proposed by DerSimonian and Laird [7] and the variance of the estimator inflated to account for the heterogeneity, thus preventing a reduction in confidence interval coverage. Thus, we remedy the overdispersion expected by introducing a dispersion parameter into the model so that the conditional variance of the true study effects increases more rapidly than that based on the probability distribution assumed by a fixed effects approach. This is then a quasi-likelihood approach since the QE model specifies the conditional variance of the study specific true effects directly and this has the advantage of being based purely on the variance-to-mean relationship (rather than on distributional assumptions) with variance appropriately inflated using a scale parameter, ψ_j . The latter can be defined by interpreting the multiplicative factor as an intra-class correlation (ICC) as described by Kulinskaya & Olkin [15] where the $ICC_j = \gamma^2 / (\gamma^2 + \nu_j)$ and the scale parameter is defined as:

$$\psi_j = \frac{1}{1 - ICC_j} \quad \{8\}$$

The variance of any weighted estimator is given by $\sum_{j=1}^k \varpi_j^2 \text{var}(\hat{\delta}_j)$ where ϖ_j is any series of weights that sum to 1. The latter is then inflated to $\sum_{j=1}^k \varpi_j^2 \text{var}(\hat{\delta}_j) \psi_j$ and using this expression, the variance of the estimator under the QE model weights then is given by:

$$\text{var}(\hat{\theta}_{QE}) = \sum_{j=1}^k \left[(w_j'')^2 (v_j + \gamma^2) \right] \quad \{9\}$$

4. EXAMINING ESTIMATOR PERFORMANCE USING SIMULATION

We now proceed to examine the performance of the RE and QE estimators (Table 1) under varying degrees of heterogeneity. The odds ratio is used as the effect size (though the models can deal with any of the common effect measures) and the simulation is modeled around the magnesium meta-analysis [16] data which was previously reviewed by Al Khalaf et al [17]. Based on this meta-analysis a simulation study was set-up fixing the true effect size as the odds ratio (OR) and selecting an OR from between 0.4 and 4 and allowing the study sample size (N_j) as well as the proportion of events and non-events in the j th study to vary in a similar pattern as in the original studies. Randomly generated variance due to bias or random chance were added (as previously described [18]) to the true OR, the magnitude of the bias thus varying over runs to generate different levels of heterogeneity. The OR and four-fold cells required for each study were generated as previously described [18]. The only difference was that the simulation

generates q_j and we go a further step to compute Q_j from it which is then used as input into the QE model. Since the real-life Q_j is more uncertain when generated from a quality scale, we increase uncertainty around Q_j by creating a beta distribution around q_j that simulates a quality scale with a maximum score out of 10 points as follows:

$$Q_j \sim \text{Beta}\left(\{q_j \times 10\}, \{(10 \times \max q_j) + 0.1 - (10 \times q_j)\}\right) \quad \{10\}$$

Every run generated k studies (randomly between 5 and 19) with a sample size (N_j) from three different distributions giving 45 combinations possible that were uniformly distributed across all the simulation iterations. The three distributions of N_j were a Delaporte distribution (with parameters 0.1, 8000, 160) [18], a uniform distribution between 50 and 58000 in increments of 50 and finally a uniform distribution between 25 and 200 in increments of 25. The data from 10,000 iterations of these k studies for each model at each heterogeneity level were generated using the Monte Carlo simulation program Ersatz (Epigear International, Sunrise Beach, Australia; www.epigear.com) and meta-analyses results computed through MetaXL (Epigear International, Sunrise Beach, Australia; www.epigear.com). The performance measures were computed from the simulated data exactly as detailed by Burton et al [19]. The various performance measures were also plotted as a function of increasing heterogeneity, the latter being indicated by the median γ^2 in a particular simulation run. The Delaporte distribution had a median study size of 175 and a distribution that resembled the original magnesium meta-analysis with the occasional mega-trial. A total of 10 separate simulations (OR 0.4 to 4 in steps of 0.4) involving a million separate meta-analyses were therefore performed, but only selected results from two simulations are reported in this paper because they all concurred in terms of

estimator performance. We also ran the simulation protocol using the standardized mean difference effect size, but given that results were similar to the log odds ratio effect size, only the latter are reported.

The first observation from the simulation was that indeed the use of $\hat{\tau}_j$ decreased the bias in the estimator considerably (Figure 1). The second observation from this simulation was to confirm that the QE estimator had a clearly lower MSE than the RE estimator (Figure 1). Additionally, since both empirically weighted models (RE and QE) discount studies with larger sampling variability when heterogeneity is low, they did have a similar MSE when studies were homogenous (starting point in Figure 1). Since the MSE is lower for the QE estimator under increasing heterogeneity, the QE model estimator is more efficient than the RE estimator.

A comparison across the two models of the confidence interval width (not shown) and coverage probability (Figure 1) confirms that the QE estimator produced a CI with a slightly broader width to retain the coverage probability above the nominal level while the RE estimator fails to retain the coverage probability. Although the results of 2 simulations are shown in Figure 1, the results of the remaining 8 simulations were similar when the simulations were run with different effect sizes (range of OR from 0.8 to 3.6).

When quality was randomly generated at the point of input into the QE model without regard to the actually simulated quality in each iteration, the QE model had an identical MSE to the RE model (Figure 2). This suggests that the RE model is simply a QE model with random entry of quality information. Despite the performance estimates equalizing, the coverage probability of the QE confidence interval drops only to 90% unlike the greater drop seen with

the RE model. Finally, when the actual simulated variances were used as weights (the true QE model; Figure 2), the estimator was unbiased (as expected) and coverage probability was exactly at the nominal level of 95%. The MSE and variance using the true weights were nevertheless similar to the QE model MSE and variance.

Finally, we looked at the difference between the simulation generated “true” weight and the simulation based QE weight (with subjectivity added through expression {10}) for the first study in each iteration in run 1 (no heterogeneity) and run 10 (maximum heterogeneity) on the normalized scale (ie that sum to 1). The histogram in Figure S1 clearly depicts that the QE method is able to capture the “true” normalized weight quite well even when quality is measured subjectively.

5. REAL DATA EXAMPLES

To compute the meta-analysis results using real data requires the following steps:

- a) Quality assessment of individual studies using a quality scale and computing a univariate quality score. Each component is equally weighted given that we do not have sufficient information yet from meta-epidemiological studies to do otherwise. In the future differential weighting of quality components may be an option if data from such studies accrues.
- b) Conversion of the univariate score to Q_j by dividing each score by the maximum score in the list of studies

- c) Plugging summary statistics and Q_j into our software MetaXL (downloadable freely from www.epigear.com) which uses the methods defined in this paper to compute meta-analytic estimates and plots.

We first looked at the controversial magnesium meta-analysis whose data consists of 19 English language randomized trials (published prior to June 2006) that reported on early mortality after myocardial infarction [17]. Early mortality was defined as occurring in hospital during the acute admission phase or within 35 days of onset of myocardial infarction. This meta-analysis is well known because of the discrepancy between the mega-trial and the random effects meta-analysis result. When the meta-analysis estimates were computed using the two methods described here as well as when the quality score of the studies was considered to be the same (IVhet model), results were more conservative with the QE estimator (OR 0.81; 95% CI 0.61 – 1.08) and less so with the RE estimator. What the QE estimate depicts (Figure 3B) is support for the results of both the larger as well as the better quality studies (pooled estimate) while at the same time support for the smaller or poorer quality studies by increasing uncertainty around the pooled estimate as evidenced by the expanded (but presumably correct) confidence interval. The QE estimator, because it penalizes for both precision and quality, discounts the most precise study (ISIS-4) based on additional information from the study and thus provides more evidence for a possible benefit for magnesium (Figure 3B). When we applied the random effects model here (Figure 3C), it simply equalized the weights (more or less) and produced a very extreme effect without justification. Figure 3A depicts

results with the IVhet model which assumes studies all have the same quality. This result depicts a larger estimator variance and bias because of the presumption of equal quality.

Another example from the recent meta-analysis literature also reveals a similar comparison. A meta-analysis by Ha and colleagues [20] demonstrates a more conservative effect under the QE model (Figure 4A) while under the RE model (Figure 4B) there is a larger effect with a narrower confidence interval. Application of the random effects model, as the authors did, underestimates the statistical error without justification.

6. DISCUSSION

The QE model estimate differs from the RE model estimate in two perspectives: Pooled QE estimates favor both larger and better trials (as opposed to penalizing larger trials with the RE model) and have a more conservative confidence interval that retains the nominal coverage probability). The implication for the meta-analysis of the magnesium studies in myocardial infarction (Figure 3) is that the evidence for the intervention suggests less benefit when methodology is also assessed.

When quality information becomes available, several other options also open up which are different from our approach. Bias quantification has been proposed as a theoretical way to improve the estimator performance [21] but this remains impractical in meta-analysis because there is no definite relationship between a quality deficiency and the quantitative magnitude or

direction of bias in the study effect [11]. Our approach of using the additional information to model the component of between studies variance likely to be contributed by systematic error in individual studies is less problematic in this respect and clearly leads to gains in estimator efficiency as demonstrated here.

A final point is that the QE model remains robust to variability in quality assessment down to the point of completely random assessment of quality. When the latter happens, the MSE and variance of the QE estimator will tend to equal that of the RE estimator and its CI coverage will drop too. However, the latter does not occur on the magnitude seen with the conventional random effects model estimator. Figure 2 depicts the contrast between the performance under the QE model with random quality input (QE-r), the RE performance and the performance of true weights in the QE model (generated as the reciprocal of the sum of the simulated variances due to chance and bias for each study). It is seen that the MSE of the QE-r estimator equals that of the RE estimator while, as expected, the MSE with the true weights is the least and represents the best possible performance that an estimator can possibly have. The true weights also depict their expected property of zero bias and clearly maintained coverage at the nominal level (Figure 2).

We conclude that the QE model of meta-analysis is a clear improvement over the RE model to handle heterogeneity when quality information is available. Alternatively, when quality information is lacking, the IVhet estimator can be used (see previous paper in this series) [3]. To facilitate use and further evaluation of this new and improved method, our software, MetaXL (available for free download at www.epigear.com), has been updated to

version 2.0 to run the QE model as well as the IVhet model and all other conventional models for comparison.

FUNDING

There was no external funding for this study.

CONFLICT OF INTEREST

JJB owns Epigear International Pty Ltd which sells the Ersatz simulation software used in this study.

ACCEPTED MANUSCRIPT

Appendix 1: Computation of $\hat{\tau}_j$

In order to redistribute the inverse variance weights using Q_j , it first needs adjustment (Q'_j) to prevent the possibility of negative weights as follows:

$$Q'_j = \begin{cases} \frac{\left(\sum_{j=1}^k Q_j \right) \tau_j}{\left(\sum_{j=1}^k \tau_j \right) (k-1)} + Q_j & \text{if } (\exists Q_j) Q_j < 1 \\ Q_j & \text{otherwise} \end{cases}$$

where

$$\tau_j = \frac{1 - Q_j}{v_j} / (k-1)$$

and k is the number of studies in the meta-analysis. From this $\hat{\tau}_j$ is then computed given by

$$\hat{\tau}_j = \left(\left(\sum_{j=1}^k \tau_j \right) k \frac{Q'_j}{\sum_{j=1}^k Q'_j} \right) - \tau_j$$

The advantage of using this correction to reduce estimator bias is indicated in the comparison of QE versus QE(-t) in panel C of Figure 1.

References

1. Doi SA, Thalib L. A Quality-Effects Model for Meta-Analysis. *Epidemiology* 2008; **19**(1):94-100.
2. Doi SA, Thalib L. An Alternative Quality Adjustor for the Quality Effects Model for Meta-analysis. *Epidemiology* 2009; **20**(2):314.
3. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogenous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials* 2015;In process.
4. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20**(6):825-40.
5. Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Stat Med* 2011; **30**(28):3304-12.
6. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999; **150**(5):469-75.
7. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**(3):177-88.
8. Overton RC. A Comparison of Fixed-Effects and Mixed (Random-Effects) Models for Meta-Analysis Tests of Moderator Variable Effects. *Psychological Methods* 1998; **3**(3):354-379.
9. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009; **172**(1):137-159.
10. Doi SA. Evidence Synthesis for Medical Decision Making and the Appropriate Use of Quality Scores. *Clin Med Res* 2014; **12**(1-2):40-6.
11. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001; **2**(4):463-71.
12. Doi SA, Barendregt JJ, Mozurkewich EL. Meta-analysis Of Heterogenous Clinical Trials: An Empirical

- Example. *Contemp Clin Trials* 2011; **32**:288-98.
13. 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**(3):255-65.
 14. Wedderburn RW. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* 1974; **61**:439-447.
 15. Kulinskaya E, Olkin I. An overdispersion model in meta-analysis. *Statistical Modelling* 2014; **14**(1) :49-76.
 16. Li J, Zhang Q, Zhang M, Egger M. Intravenous magnesium for acute myocardial infarction. *Cochrane Database Syst Rev* 2007;(2):CD002755.
 17. Al Khalaf MM, Thalib L, Doi SA. Combining heterogenous studies using the random-effects model is a mistake and leads to inconclusive meta-analyses. *J Clin Epidemiol* 2011; **64**(2):119-23.
 18. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Simulation Comparison of the Quality Effects and Random Effects Methods of Meta-analysis. *Epidemiology* 2015; DOI: 10.1097/EDE.0000000000000289 .
 19. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006; **25**(24):4279-92.
 20. Ha V, Sievenpiper JL, de Souza RJ *et al.* Effect of dietary pulse intake on established therapeutic lipid targets for cardiovascular risk reduction: a systematic review and meta-analysis of randomized controlled trials. *CMAJ* 2014; **186**(8):E252-62.
 21. Thompson S, Ekelund U, Jebb S *et al.* A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol* 2011; **40**(3):765-77.

Figure legends

Figure 1: The left (A – D) and right (E – H) panels are based on an effect size of OR = 0.4 and OR = 4 respectively. The panels depict MSE (A & E), variance (B & F), bias squared (C & G) and coverage probability (D & H). The MSE (A & E) is lowest for the QE model estimator and the effect of adding $\hat{\tau}_j$ (QE) or not (QE(-t)) is shown. The QE estimator (in contrast to the QE(-t) estimator has much less bias leading to a lower MSE.

Figure 2: The left (A – D) and right (E – H) panels are based on an effect size of OR = 0.4 and OR = 4 respectively. The panels depict MSE (A & E), variance (B & F), bias squared (C & G) and coverage probability (D & H) for the true weight used in the simulation (QE true; where $\frac{Q_j}{v_j} + \hat{\tau}_j$

was replaced with $\frac{1}{v_j + \phi_j^2}$) and a simulation where a random quality (out of 10) was input into the model (QE-r). The MSE (A & E) is lowest for the QE true model estimator and the effect of using a random quality is shown. The QE true model has no bias while the random quality leads to deterioration in coverage of the QE estimator, but nowhere as near to what is seen with the random effects estimator.

Figure 3: The magnesium meta-analysis results are depicted. The IVhet model (A) favours larger studies and has a larger uncertainty around the point estimate. With the addition of quality (B) there is less uncertainty and the point estimate does not simply favour larger studies and indeed suggests a trend towards some effect. When the random effects model is used (C), there is clearly an effect, but this is because it underestimates the statistical error and thus we have an overconfident result. Forest plots created using MetaXL version 2.0 (www.epigear.com).

Figure 4: Example from a 2014 meta-analysis by Ha and colleagues [20] on the effects of pulse intake on serum LDL levels. The IVhet model (A) demonstrates no significant effect but taking quality into consideration (B) suggests that there is indeed an effect as the uncertainty decreases. The random effect model (not shown) concurs with the QE result, but simply because it tends to underestimate the statistical error. Forest plots created using MetaXL version 2.0 (www.epigear.com).

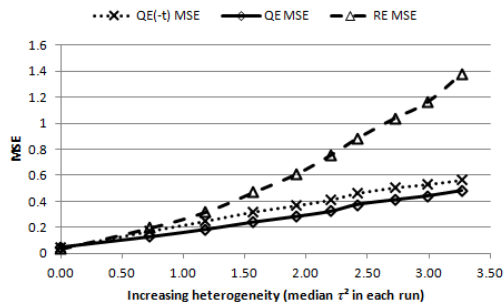
Table 1: Summary of the Two Methods*

	RE	QE
Weights that sum to 1	$w_j = \frac{1}{v_j + \gamma^2} / \sum_{j=1}^k \frac{1}{v_j + \gamma^2},$	$w_j'' = \left(\frac{Q_j}{v_j} + \hat{\tau}_j \right) / \sum_{j=1}^k \left(\frac{Q_j}{v_j} + \hat{\tau}_j \right)$
Pooled effects	$\hat{\theta}_{RE} = \sum_{j=1}^k w_j \hat{\delta}_j$	$\hat{\theta}_{QE} = \sum_{j=1}^k w_j'' \hat{\delta}_j$
Variance of pooled effect	$\text{var}(\hat{\theta}_{RE}) = \frac{1}{\sum_{j=1}^k \frac{1}{(v_j + \gamma^2)}}$	$\text{var}(\hat{\theta}_{QE}) = \sum_{j=1}^k \left[(w_j'')^2 (v_j + \gamma^2) \right]$
Comments	More fully specified model	Quasi-likelihood model

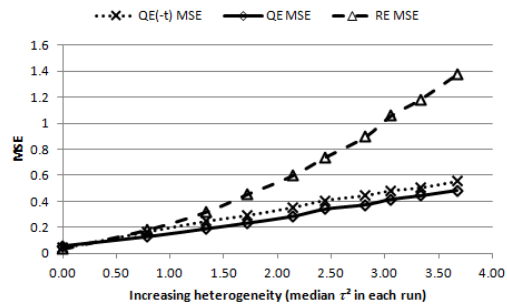
*For abbreviations or expansion of the notation please see the text

Figure 1

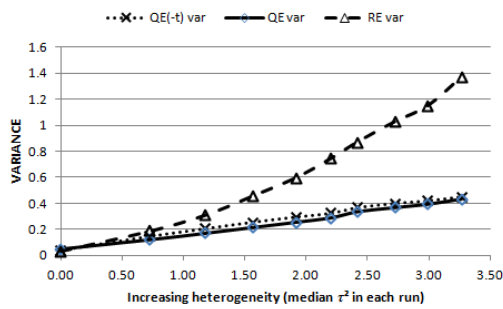
A)



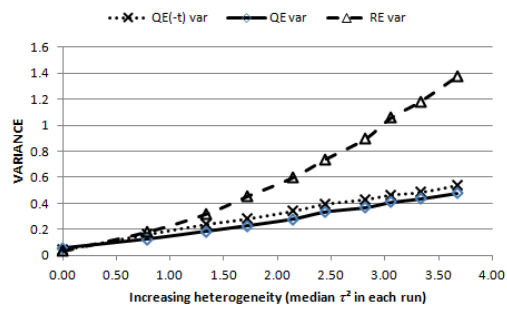
E)



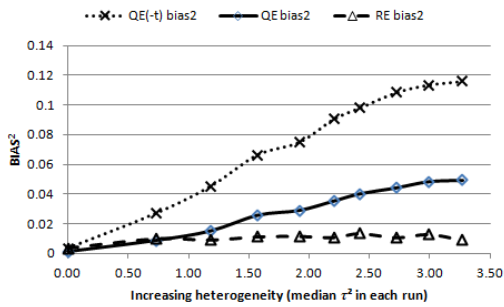
B)



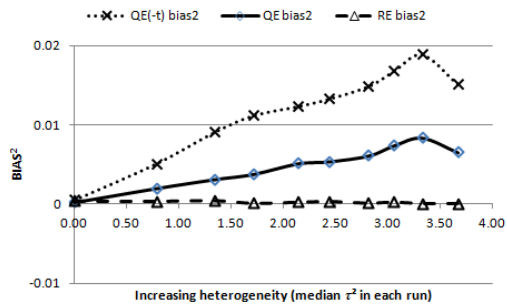
F)



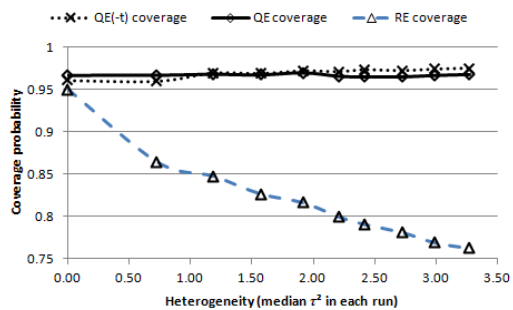
C)



G)



D)



H)

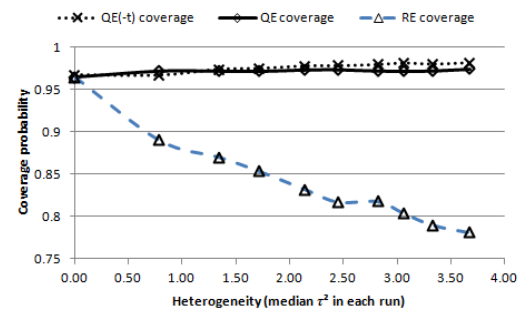
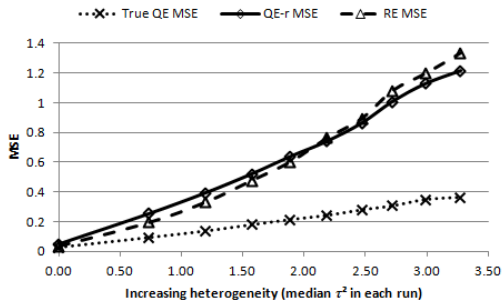
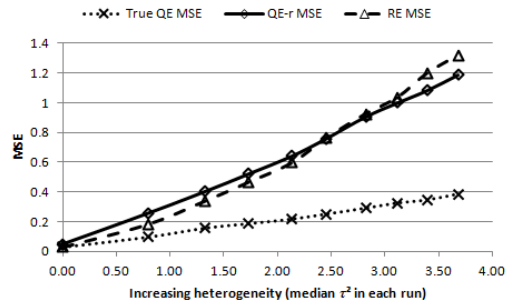


Figure 2

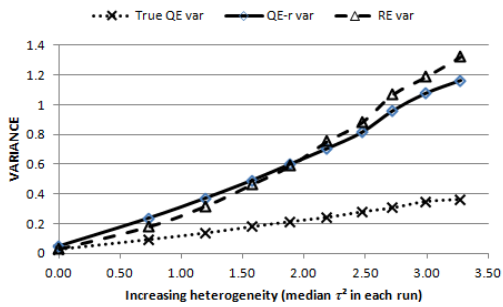
A)



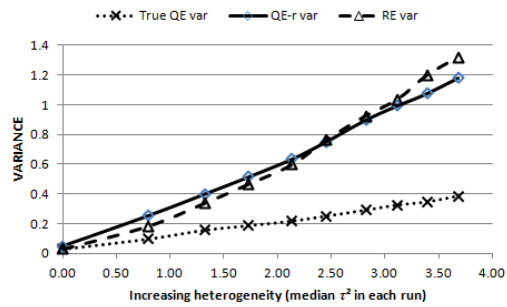
E)



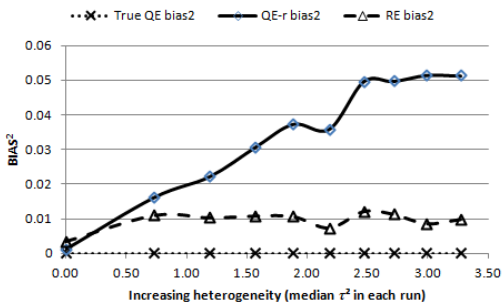
B)



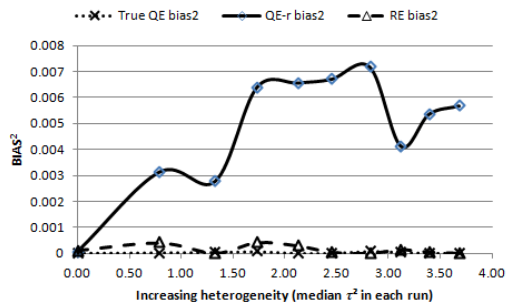
F)



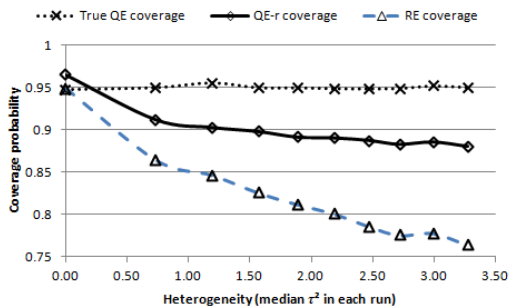
C)



G)



D)



H)

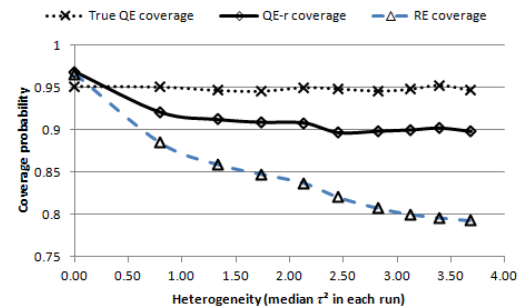


Figure 3

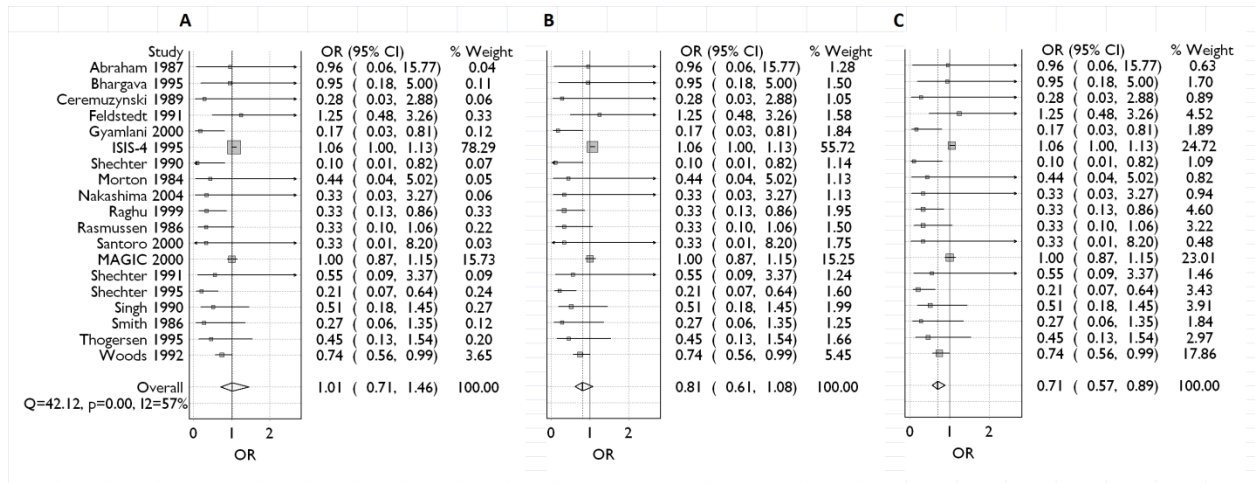


Figure 4

