J Syst Sci Complex (2022) 35: 2381–2397

Incorporating Variation and Quality of the Underlying Effects in Meta-Analysis*

FU Jinyu · LIN Jinguan

DOI: 10.1007/s11424-022-1429-5
Received: 2 November 2021 / Revised: 17 November 2021

Abstract This paper proposes a model to further explore the effects of the quality information and variation of the underlying effects on the summary effect measure in meta-analysis. A shape parameter is used in this model to quantify the asymmetry of the effect sizes of studies that are included. Estimation of the proposed model parameters is carried out by the Bayesian MCMC method. Performances of the resultant estimates are examined in the simulations and empirical case with data obtained from a total of 22 meta-analyses taken from three different designs. A conclusion would be drawn that it is advisable to take the proposed model, when quality information becomes available, in particular with a situation where the underlying effects approximately follow a normal distribution. If, however, the quality information is absent, the skew-normal distribution for random effect model should be adopted.

Keywords Bayesian estimation, heterogeneity, overall effect size, quality score, skew-normal distribution.

1 Introduction

Meta-analysis has become a powerful and widely used tool to integrate findings from different studies and inform decision-making in evidence-based medicine[1]. Statistical methods are used to combine the results of several studies that concentrate on similar research goals. It makes sense to compute a summary effect from studies if they have the same metric, for example, several randomized controlled trials are compulsorily on a continuous or a dichotomized scale. For continuous outcomes, the common effect size is the standardized mean difference in which the sample size, mean, and standard deviation are required of. To get the sample mean and

FU Jinyu
School of Mathematics, Southeast University, Nanjing 210096*, China.* Email: fujinyu163@163.com.
LIN Jinguan (Corresponding author)
School of Mathematics, Southeast University, Nanjing 210096*, China; School of Statistics and Data Science, Nanjing Audit University, Nanjing* 211815*, China.* Email: jglin@nau.edu.cn.
*This research was supported by the Natural Science Research Foundation of China (Key Projects) under Grant No. 11831008, the Natural Science Research Foundation of China under Grant No. 11971235.
◇*This paper was recommended for publication by Editor LI Qizhai.*

variance, for example, Hozo, et al. employed inequalities to establish some estimators[2], Wan, et al. introduced a quartile method to improve the sample standard deviation estimation, compared to Hozo, et al.'s method[3], and Luo, et al. incorporated the sample size in a smoothly changing weight in the estimators to reach the optimal estimation[4]. For dichotomized data, "OR", "RR", and "RD" are widely used as the effect sizes, of which, the issues are extensively discussed and reviewed.

The summary effect plays a significant role in many systematic reviews. Two types of models are most frequently employed at the early stage to deal with the summary effect. A combination of study effect sizes is conducted with the approach termed as "Fixed Effect Model", assuming that the differences observed between these studies are due to chance alone, or with the generally advocated approach known as "Random Effect Model", which hypothesizes that the observed variance includes chance and between-study heterogeneity that cannot readily be explained[5]. When the random effect (RE) is applied in some badly designed studies, it would result in biased estimates, even though there is an adjustor in the heterogeneity[6], an apparent reason for which is that the RE model fails to take the study quality into consideration. In terms of trail quality, Verhagen, et al., for example, provided some evidence that one study quality will affect the magnitude of the combined estimates as well[7]. At the same time, some contributions incorporated the quality score as a covariation of the study weight into a calculation of a summary effect[8–10]. The study weights based on sampling theory (e.g., inverse-variance weights) can be multiplied by quality scores[11], increasing the weight of high-quality studies and decreasing the weight of poorer ones. To further handle this problem, a quality-adjusted model called the quality effects (QE) model was first proposed by Doi and Thalibb[12], and was then adequately developed by some researchers[13, 14]. In the QE model, quality scores are viewed as a component of study weights of the weighted mean. Some rigorous justifications for the weighting by quality scores are given in medical research[14]. Moreover, Doi, et al. also discussed a variant of the QE model that is called the inverse variance heterogeneity (IVhet) model, not requiring quality assessment because all studies by default are assigned the same quality[15].

While the QE model and RE model can to some extent alleviate the problem of heterogeneity, they are hampered by two shortcomings. One is that despite this QE model avoids the artificial inflation in the variance, quality weights lead to bias[16, 17]. They only capture the (methodological) heterogeneity in quality, one important explainable differences, without paying attention to other inter-study difference, such as different definitions of treatment effects and unexplained differences between studies. A second shortcoming of the models is that it may be occasionally appropriate to make an assumption that the underlying effect for the RE model is taken to be a normal distribution, as, however, in practice, is not always the case[18]. It is usually considered that the heterogeneity, in the RE model which involves a normal assumption, refers to the variability in the intervention effects evaluated in different studies, and is described as statistical heterogeneity which solely accounts for the unexplainable differences. Since any kind of variability among studies in a systematic review may be termed heterogeneity, comprised of the clinical heterogeneity, methodological heterogeneity as well as statistical het-

erogeneity, it is acceptable for some researchers to consider as much heterogeneity as possible rather than to consider parts of the heterogeneity. Furthermore, if the assumption is inappropriate, some statistical heterogeneity may be introduced. In sum, it is, in practice, of interest to develop more flexible approaches using a broader family of distributions.

On the basis of the problems aforementioned, we are trying to develop an alternative to allowing for the variance by virtue of the variation in underlying effect size as well as variance modeled by quality scores in meta-analysis. In this article, we make use of quality as a covariation in each study weight based on the random effects weight, which is the inverse of the sampling variance plus the random effects variance component τ^2. However, the variance of the true effect size for each study is probably not τ^2, owning to the skewed study effect sizes, Thus, the problem we face is urgent. In statistical literature, the distribution of the random effects and model errors have been some heated arguments[19–21], and skewness has also long been used as a descriptive quantity for the asymmetry of a distribution[22], such as the extensive application in econometrics employed to characterize the leptokurtosis of the securities' return[23], but they are quite novel in the literature of meta-analysis. Navigated by analyses above, it is therefore proposed that the distributions of the true effects across studies are presupposed to follow skew-normal distributions with the same mean and variance but different skewness parameters, through which the asymmetry of the collected study results is well-portrayed and the magnitudes of the meta-analysis results are thus adjusted to be closer to the truth as well under certain circumstances.

The following sections of this paper are arranged below. Section 2 covers "a skew-normal random effects model", "a combination of quality and variation of the true effect size for each study into study weights", as well as "the usage of WinBUGS implemented to obtain Bayesian estimates". Section 3 investigates the performance of the new method (QE-v model), compared to the QE model and the RE model with a normality assumption or a skewed normality assumption using Markov chain Monte Carlo (MCMC) simulations. A real example on the efficacy of radioactive iodine for the ablation of thyroid remnants after surgery for thyroid cancer is presented to be in comparison to the results obtained from the old and new methods applied to a summary of clinical in Section 4. Section 5 provides a discussion of the limitations and strengths of the proposed model.

2 Incorporating Variation and Quality of the Underlying Effects for Random Effects

2.1 Skew-Normal Assumption for Random Effects

A normal assumption for random effect model is usually utilized, which, however, can not adequately fit the observed data. Hence, we might seek a parsimonious model to account for the specification. One choice of the model is to specify a skew-normal distribution, the skewness parameter of which is used to describe the asymmetry of data. The details for the model are as follows.

The basic idea is to parcel out some measure of the observed treatment effect in each study,

say y_i, into two additive components: The true effect size ζ_i, and sample error, ε_i. The variance of ε_i is the sample variance, δ_i^2, and it is usually calculated from the data of the ith observed sample. Because of the patient characteristics, experimental design, measurement of outcome and execution of the study, we assume that the true effect size for the ith study follows a skew-normal distribution, rather than the conventional normal distribution, i.e., $\zeta_i \sim \mathrm{SN}(\mu, \tau^2, \gamma_i), i = 1, 2, \cdots, n$, where τ^2 is the between-study variance and γ_i is the shape parameter, representing the magnitude of the asymmetry of the effect sizes of included studies. To explicitly account for the variation in true treatment effects, the true effect is given by the grand mean, μ, plus the deviation of the study's true effect from the grand mean, ξ_i. The model is:

$$\begin{aligned} y_i &= \zeta_i + \varepsilon_i \\ &= \mu + \xi_i + \varepsilon_i, \quad i = 1, 2, \cdots, n, \end{aligned}$$

where $\xi_i \sim \mathrm{SN}(0, \tau^2, \gamma_i)$ and $\varepsilon_i \sim N(0, \delta_i^2)$. ξ_i and ε_i are assumed independent. Efficiently to be computed, $y_i \sim \mathrm{SN}(\mu, \tau^2 + \delta_i^2, \frac{\tau\gamma_i}{\sqrt{\tau^2+\delta_i^2(1+\gamma_i^2)}})$ (see Appendix B for computation), and it is called "RE-s model". Note that when γ_i is the same, the RE model has been discussed in [18]. when $\gamma = 0$, ξ_i is reduced to be $N(0, \tau^2)$, which indicates it is the RE model with a normal assumption.

2.2 Differences in Weighting Between Models

One problem with meta-analysis today is these differences between the studies involved. There are multiple sources for these inevitable differences encompassing chance, study quality, and study design, and so on, all of which may falsify the conclusion of the pooled estimate[24]. It is warned that estimates from meta-analysis that do not consider study quality will be biased and could lead to an inflated Type I error, and, at the same time, some researchers are aware of that conclusions drawn from meta-analyses using flawed studies will be invalid. Therefore, results from better quality studies should in some sense be more valid or accurate. Fortunately, based on the additional variance contribution from internal biases (ϕ_i) viewed as a bias adjustment of the combined estimate[25, 26], the latest QE model tactfully settles out the quality of trails[14]. The rescaled score is

$$Q_i = \frac{q_i}{q_{i\max}} = \frac{\phi^2 + \phi_{i\min}^2}{\phi^2 + \phi_i^2},$$

where ϕ is between study bias and intra-class correlation $q_i = \frac{\phi^2}{\phi^2+\phi_i^2}$ can be considered the "quality-weight" for each study, reflecting the proportion of total bias variability not related to variability from internal study bias[16,27,28]. The final QE model weight ($\widehat{\omega_i}''$) for each study is given by

$$\widehat{\omega_i}'' = \frac{Q_i}{\delta_i^2} + \widehat{\kappa_i},$$

and weights that sum to 1 are given by

$$\omega_i'' = \frac{\frac{Q_i}{\delta_i^2} + \widehat{\kappa_i}}{\sum_{i=1}^{n}(\frac{Q_i}{\delta_i^2} + \widehat{\kappa_i})}.$$

The quantity, $\widehat{\kappa}_i$, is described in the document[14].

In the QE model and IVhet model[14, 15], one way to prevent the coverage probability of the confidence interval below the nominal level is to use a scale parameter, ψ_i, appropriately inflating the variance and defined by interpreting the multiplicative factor as an intra-class correlation (ICC) as described by Kulinskaya and Olkin[29]. The scale parameter is defined as

$$\psi_i = \frac{1}{1 - ICC_i}, \tag{1}$$

where ICC_i is $\frac{\tau^2}{\tau^2+\delta_i^2}$, the estimator of the between-study heterogeneity variance (τ^2) from which is derived from the RE model under the normal setting by adopting the Bayesian estimation described in Subsection 2.3. Based on (1), the variance of the weighted mean estimator, $\widehat{\mu}_{QE}$, is given by

$$\mathrm{var}(\widehat{\mu}_{QE}) = \sum_{i=1}^{n} \omega_i'' \mathrm{var}(y_i)\psi_i = \sum_{i=1}^{n} (\omega_i'')^2(\tau^2 + \delta_i^2).$$

Nevertheless, some different arguments that quality weighting can virtually produce biased estimates emerged[16, 17], and quality is just one of the explainable differences. To make reasonable use of quality information and get a summary result estimation closer to the true value, we thus propose a model called QE-v model amalgamating the heterogeneity in quality scores and some unrelated-quality heterogeneity into study weights. Specifically, the proposed weight considers statical heterogeneity by estimating the between-study variance when the true effects differ between studies, as well as the methodological heterogeneity caused by explained difference (quality) between studies. The weights hinging on the variance of the observed effect and quality scores are

$$\widetilde{\omega}_i' = \frac{Q_i}{\tau^2(1 - \frac{2\gamma_i^2}{\pi(1+\gamma_i^2)}) + \delta_i^2}, \tag{2}$$

and weights that sum to 1 are given by

$$\omega_i{}' = \frac{\widetilde{\omega}_i'}{\sum_{i=1}^{n} \widetilde{\omega}_i'},$$

where, like the QE model, parameter estimates are stemmed from the random effect model which involves a skew-normal distribution by using the Bayesian estimation showed in Subsection 2.3. If the between-study heterogeneity variance is too large, leading to equal weights for each study, our proposed weight, Equation (2), would be capable of redistributing the weights by using the inverse-variance weights of RE-s model to multiply the quality scores of the corresponding studies. In other words, the proposed model can get rid of the deficiency from the random effect that a statistical adjustment for heterogeneity will still produce invalid estimates when used in a meta-analysis of poorly designed studies, and overcome the shortcoming of the QE model that the variation between the results of studies not covered by the term "quality" is not taken into account. Additionally, the intra-class correlation (ICC_i) is $\frac{\tau^2(1-\frac{2\gamma_i^2}{\pi(1+\gamma_i^2)})}{\tau^2(1-\frac{2\gamma_i^2}{\pi(1+\gamma_i^2)})+\delta_i^2}$, as

accounts for the relative importance of between batch variation versus variation due to sampling errors.

2.3 Bayesian Approach to Random Effects Meta-Analysis

When it refers to the parameter inference, the classical estimate — Maximum Likelihood Estimation is always preferred. In fact, inference about the parameters of the skew-normal distribution based on the maximum likelihood approach has some problems, such as

(a) The maximum likelihood (ML) estimator for the skewness parameter can be infinite;

(b) The Fisher information matrix is singular when $\gamma = 0$;

(c) Existence of local maximum.

To solve the singularity problem of the Fisher information matrix, Pewsey proposed a reparameterization method[30]. As far as the first problem is concerned, the utilization of the Bayesian analysis can be a good alternative to making inference under the skewness parameter[31]. Besides, in the context of meta-analysis, the prior distribution will describe uncertainty regarding the particular effect measure, and the whole process can be easily implemented, due to the flexibility of the WinBUGS software. For these reasons, the Bayesian method is naturally harnessed.

The hierarchical model for the random effect meta-analysis is:

$$y_i \mid \zeta_i \sim N(\zeta_i, \delta_i^2),$$
$$\zeta_i \sim SN(\mu, \tau^2, \gamma_i).$$

Because the density of a skew-normal distribution is not specified directly in WinBUGS software, we consider the stochastic representation given by Property D in Appendix A. The true effect hence has the hierarchical representation below:

$$\zeta_i \mid U_i, \quad \gamma_i,\ \mu,\ \tau \sim N\left(\mu + \frac{\tau\gamma_i}{\sqrt{1+\gamma_i^2}}U_i, \frac{\tau^2}{\sqrt{1+\gamma_i^2}}\right), \quad U_i \sim HN(0,1),$$

where HN is half normal distribution which, in general, defined as follows: If $X \sim N(0, \Delta^2)$, then $U = |X|$ follows a half normal distribution with $E[U] = \frac{\Delta\sqrt{2}}{\sqrt{\pi}}$.

When a Bayesian MCMC approach in WinBUGS is carried out, it naturally comes to the choices of priors. Here, we pick out vague priors for μ and τ, that's to say, Normal$(0, 100)$ for μ and Uniform$(0, 100)$ for τ[32]. The Jeffreys prior $t(0, \pi^2/4; 1/2)$ is used for the γ_i and γ in the skew-normal distributions[31].

3 Simulation Study

3.1 Data-Generating Mechanism

In this section, we employ simulated normal and non-normal random effects data to examine the performance of the proposed overall effect size estimators by the Bayesian MCMC method. The characteristics of estimators for the four models, i.e., RE model, RE-s model, QE model, and QE-v model, are summarized in Table 1. The log odds ratio is used as the effect size. We

first fix the true heterogeneity variance τ^2, 1.69, and the true effect size as an OR from 0.3 to 2.7 with increment 0.3. The simulation data is generated as follows.

Table 1 Summary of the four methods of estimation

Model	Weights	Weights that sum to 1	Pooled effects	Variance of pooled effects
RE	$\widehat{\omega}_i^* = \frac{1}{\tau^2+\delta^2}$	$\omega_i^* = \frac{\widehat{\omega}_i^*}{\sum_{i=1}^n \widehat{\omega}_i^*}$	$\widehat{\mu}_{RE} = \frac{\sum_{i=1}^n \widehat{\omega}_i^* y_i}{\sum_{i=1}^n \widehat{\omega}_i^*}$	$\mathrm{var}(\widehat{\mu}_{RE}) = \frac{1}{\sum_{i=1}^n \widehat{\omega}_i^*}$
RE-s	$\widehat{\omega}_i^\gamma = \frac{1}{\eta_i^2}$	$\omega_i^\gamma = \frac{\widehat{\omega}_i^\gamma}{\sum_{i=1}^n \widehat{\omega}_i^\gamma}$	$\widehat{\mu}_{RE-s} = \frac{\sum_{i=1}^n \widehat{\omega}_i^\gamma y_i}{\sum_{i=1}^n \widehat{\omega}_i^\gamma}$	$\mathrm{var}(\widehat{\mu}_{RE-s}) = \frac{1}{\sum_{i=1}^n \widehat{\omega}_i^\gamma}$
QE	$\widehat{\omega}_i^{'} = \frac{Q_i}{\delta_i^2} + \widehat{\kappa}_i$	$\omega_i'' = \frac{\widehat{\omega}_i^{'}}{\sum_{i=1}^n \widehat{\omega}_i^{'}}$	$\widehat{\mu}_{QE} = \frac{\sum_{i=1}^n \widehat{\omega}_i^{'} y_i}{\sum_{i=1}^n \widehat{\omega}_i^{'}}$	$\mathrm{var}(\widehat{\mu}_{QE}) = \sum_{i=1}^n (\omega_i'')^2(\tau^2+\delta_i^2)$
$QE-v^a$	$\widetilde{\omega}_i' = \frac{Q_i}{\tau^2+\delta_i^2}$	$\omega_i' = \frac{\widetilde{\omega}_i'}{\sum_{i=1}^n \widetilde{\omega}_i'}$	$\widehat{\mu}_{QE-v} = \frac{\sum_{i=1}^n \widetilde{\omega}_i' y_i}{\sum_{i=1}^n \widetilde{\omega}_i'}$	$\mathrm{var}(\widehat{\mu}_{QE-v^a}) = \sum_{i=1}^n (\omega_i')^2(\tau^2+\delta_i^2)$
$QE-v^b$	$\widetilde{\omega}_i' = \frac{Q_i}{\eta_i^2}$	$\omega_i' = \frac{\widetilde{\omega}_i'}{\sum_{i=1}^n \widetilde{\omega}_i'}$	$\widehat{\mu}_{QE-v} = \frac{\sum_{i=1}^n \widetilde{\omega}_i' y_i}{\sum_{i=1}^n \widetilde{\omega}_i'}$	$\mathrm{var}(\widehat{\mu}_{QE-v^b}) = \sum_{i=1}^n (\omega_i')^2(\eta_i^2)$

Note: $\eta_i^2 = \tau^2(1 - \frac{2\gamma_i^2}{\pi(1+\gamma_i^2)}) + \delta_i^2$; "$a$" means the between-study heterogeneity variance τ^2 is obtained based on RE model; "b" means the between-study heterogeneity variance τ^2 is obtained based on RE-s model.

Step 1 Generate the number of patients N_i in the ith study from $Uniform(25, 238)$, and the number of studies in each meta-analysis is 100.

Step 2 Total survivors n_{2i} ($p_i N_i = n_{2i}$) and dead n_{1i} ($n_{1i} = N_i - n_{2i}$) are then determined by allocating a proportion p_i from a uniform distribution with parameters $(0.650, 0.800)$.

Step 3 The survivors n_{2i} are distributed between treated, e_{2i}, and untreated, c_{2i} groups by this allocation in a uniform distribution $Uniform(0.464, 0.545)$ to the treated group and the rest allocated to the untreated control groups. The numbers of treated and untreated survivors are used as parameters in beta distribution to generate the proportion of survivors who are treated, i.e., $b_i \sim Beta(e_{2i}, c_{2i})$, and then the proportion a_i of the dead who are treated is:

$$a_i = \frac{b_i \exp(\mu)}{1 - b_i + b_i \exp(\mu)}.$$

In addition, the conditional sampling variance s_i^2 for the ith study is:

$$s_i^2 = \frac{1}{n_{1i}a_i + 0.5} + \frac{1}{n_{1i}(1-a_i) + 0.5} + \frac{1}{n_{2i}b_i + 0.5} + \frac{1}{n_{2i}(1-b_i) + 0.5}.$$

Step 4 Both data sets consist of the 100 final observed effect sizes, y_i, from the normal distribution $N(\mu, s_i^2 + \tau^2)$ and the skew-normal distribution SN(μ, $s_i^2 + \tau^2$, $\frac{\tau\gamma_i}{\sqrt{\tau^2 + s_i^2(1+\gamma_i^2)}}$) with $\gamma_i \sim N(1, 2)$, respectively, each of which is used to re-compute a new proportion of treated non-survivors, a_i', namely,

$$a_i' = \frac{b_i \exp(y_i)}{1 - b_i + b_i \exp(y_i)}.$$

Step 5 These proportions are used to obtain the numbers of treated non-survivors (dead) and treated survivors denoted e_{1i} and e_{2i} respectively. Thus, the final four-fold cell counts

could be reconstructed to arrive at the variance of the study effect estimate:

$$\delta_i^2 = \frac{1}{e_{1i}+0.5} + \frac{1}{n_{1i}-e_{1i}+0.5} + \frac{1}{e_{2i}+0.5} + \frac{1}{n_{2i}-e_{2i}+0.5}.$$

To account for zero outcomes in one of the studies, a continuity correction is applied by adding 0.5 to all cell counts.

Finally, the study rank Q_i is given by a beta distribution around q_i, a uniform distribution with parameters $(0.35, 0.7)$, which simulates a quality scale:

$$Q_i \sim Beta(q_i \times 10, 10 \times (\max(q_i) - q_i) + 0.1).$$

The whole simulation procedure is repeated 1000 times.

3.2 Simulation Results

The above models that are applied to fit the artificial normal data sets show some differences in the overall effects estimators (Table 2). The first observation from these simulations is that two random effect model estimators (RE model and RE-s model) irrespective of the distribution, have clearly lower MSEs than the QE model estimators, but rule out the two cases: $OR = 0.9$ and $OR = 1.2$, which is a little different from the conclusion that QE model estimator is more efficient than the RE estimator[14] (in this document, τ^2 is the methods of moment based between-study variance estimate[5]). The second observation is to confirm that normal and skew-normal distribution random effect model estimators have more or less similar MSEs, and if we apply the heterogeneity variance estimators obtained from the corresponding random effect models to the proposed QE-v model, the results are still similar. The third interesting observation is that the QE model and QE-v model perform differently, with various magnitudes of the effect sizes simulated. The following three sub-scenarios of OR are included:

(a) ($OR = 0.3, 0.6$) QE-v estimators have good finite sample performances, as we expect since the MSE of QE-v estimator is lower than that of QE model.

(b) ($OR = 0.9, 1.2$) In this case, the QE-v estimators perform not as well as the QE model estimator.

(c) ($OR \geq 1.5$) The MSE is lower for the QE-v estimator under increasing of OR, reflecting that the QE-v model estimator is more efficient than the QE estimator.

The simulated data generated from the skew-normal distributions is fitted in the same way. Seen from Table 3, a phenomenon that the MSE of the RE-s model estimator is obviously lower than the MSE of RE model indicates that the skew-normal distribution is more appropriate, which further mirrors the fact that the skewness of study effects can affect the overall effect to a certain degree if the distributions of study effect sizes are approximately skewed normal. Besides, the MSEs of the QE-v model employing heterogeneity variance derived from RE-s model are smaller than the MSEs of the QE-v model using heterogeneity variance derived from the RE model except for $OR = 0.3$ and $OR = 0.6$. It is notable that the performance of the proposed QE-v model, compared to the QE model, also relies on the value of OR: When $OR = 0.3, 0.6$, the corresponding results are consistent with a; but when $OR \geq 0.9$, the QE-v model estimators have unsatisfactory performances, compared to the QE model.

Table 2 Comparisons for QE-v model based on data simulated from normal random effects. The MSEs reported are multiplied by 100

OR	RE model	RE-s model	QE model	QE-v model	
	MSE	MSE	MSE	MSE	MSE
$OR = 0.3$	2.7035	2.6891	9.9982	2.8422	3.1626
$OR = 0.6$	1.8262	1.8126	3.0052	2.0044	2.0850
$OR = 0.9$	1.9258	1.9263	1.6430	2.0825	2.0721
$OR = 1.2$	1.7138	1.7380	1.6663	1.8354	1.8511
$OR = 1.5$	1.7545	1.7657	2.5707	1.8076	1.8348
$OR = 1.8$	1.9195	1.9327	3.6692	1.9672	2.0667
$OR = 2.1$	1.8980	1.9122	4.4935	2.0288	2.1371
$OR = 2.4$	2.3100	2.3246	6.3975	2.3686	2.5181
$OR = 2.7$	2.5592	2.5966	7.9779	2.6580	2.8473

Note: The weights used to compute the MSE of the last two columns are $\widetilde{\omega_i}' = \frac{Q_i}{\tau^2+\delta_i^2}$ and $\widetilde{\omega_i}' = \frac{Q_i}{\tau^2(1-\frac{2\gamma_i^2}{\pi(1+\gamma_i^2)})+\delta_i^2}$, respectively.

Table 3 Comparisons for QE-v model based on data simulated from skew-normal random effects. The MSEs are multiplied by 100

OR	RE model	RE-s model	QE model	QE-v model	
	MSE	MSE	MSE	MSE	MSE
$OR = 0.3$	20.7035	20.6566	28.9315	21.6329	23.3260
$OR = 0.6$	15.7390	15.7093	17.8432	16.6897	16.9366
$OR = 0.9$	12.7656	12.7577	11.6170	13.1136	12.6314
$OR = 1.2$	11.2434	11.1629	8.8952	11.7368	10.9900
$OR = 1.5$	9.5112	9.5007	6.3725	9.8859	8.8415
$OR = 1.8$	8.5393	8.4774	5.3463	9.2685	8.1381
$OR = 2.1$	7.7854	7.7504	4.4600	8.5156	7.2649
$OR = 2.4$	7.8018	7.7356	3.9886	8.4545	7.0431
$OR = 2.7$	7.3809	7.3174	3.5383	8.0674	6.6167

Note: The weights used to compute the MSE of the last two columns are $\widetilde{\omega_i}' = \frac{Q_i}{\tau^2+\delta_i^2}$ and $\widetilde{\omega_i}' = \frac{Q_i}{\tau^2(1-\frac{2\gamma_i^2}{\pi(1+\gamma_i^2)})+\delta_i^2}$, respectively.

4 Empirical Example

In this section, we take an example of 22 studies for the non-ablation of thyroid remnants[33] to illustrate the performance of the recommended approach. The suggested Quality Scoring

System and quality score of each study have been reported previously[15]. Before computing the meta-analysis results, the following steps are required:

(a): Conversion of the univariate score to Q_i by dividing each quality score by the maximum score in the list of studies.

(b): Adoption of software WinBUGS implemented to attain the parameter estimates and meta-analytic estimates.

A succinct description is provided in Figure 1 that the observed effect sizes approximately follow a little left-skewed and left-tailed distribution, which, hence, guides us to choose a skew-normal distribution to fit the real data. Alternatively, the results are listed in Table 5.

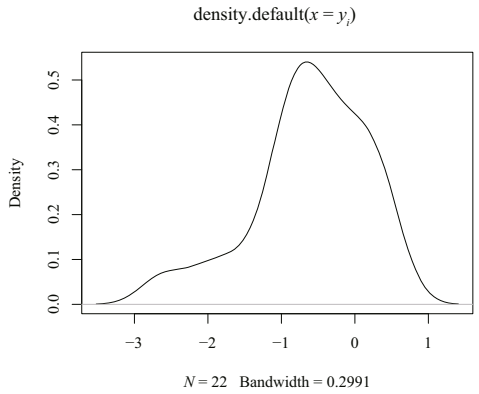


Figure 1 Distribution of the effect size of each study

We first test the performance of models without incorporation of quality, i.e., RE model and RE-s model. In general, the effect size heterogeneity is of great interest. The conventional analysis assuming normality gives a heterogeneity estimate of 0.2308 (Table 5). Relaxing the normality assumption and presuming that the underlying effects themselves sample from skew-normal distributions with distinct shape parameters, a bit different effect size heterogeneity variance ($\widehat{\tau}^2$=0.1398) is obtained, and the precision of the heterogeneity variance estimate is a slightly smaller than that in the RE model, indicating that an appropriate distribution would facilitate the increase the precision of heterogeneity. Seen from Table 5, under the non-normal assumption, the pooled effect estimate for the RE-s model is larger than that for the RE model, and the uncertainty in the location of the overall mean of systematically different effects in the different studies declines (i.e., its confidence interval). Moreover, the DIC is reduced in the skewed model highlighting the importance of the skewing and confirming the skew-normal distribution is the preferred distribution in this example.

Table 4 Characteristics of studies included in meta-analysis.

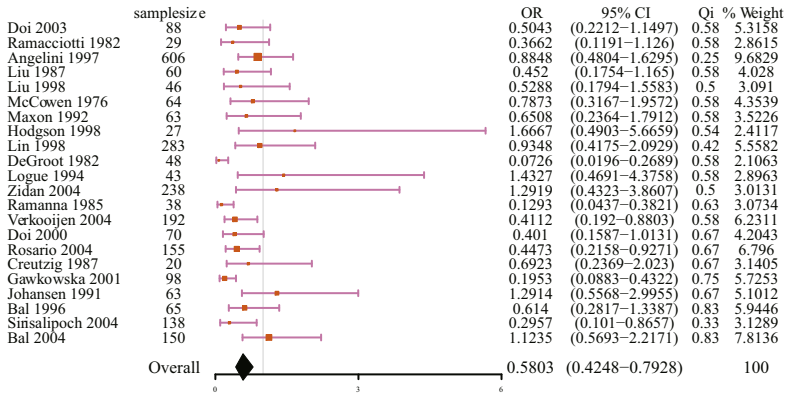
study	Total T	Failures T	Total C	Failures C	Quality score Q_i	OR
Cohorts with mixed surgical status						
Doi 2003	49	23	39	25	0.58	0.5043
Ramacciotti 1982	9	3	20	12	0.58	0.3662
Angelini 1997	426	226	180	101	0.25	0.8848
Liu 1987	40	14	20	11	0.58	0.4520
Liu 1998	21	6	25	11	0.50	0.5288
McCowen 1976	28	10	36	15	0.58	0.7873
Maxon 1992	37	6	26	6	0.58	0.6508
Hodgson 1998	7	2	20	4	0.54	1.6667
Lin 1998	89	15	194	35	0.42	0.9348
DeGroot 1982	30	0	18	3	0.58	0.0726
Logue 1994	21	4	22	3	0.58	1.4327
Zidan 2004	172	11	66	3	0.50	1.2919
Cohorts with near total surgery						
Ramanna 1985	19	9	19	17	0.63	0.1293
Verkooijen 2004	159	61	33	20	0.58	0.4112
Doi 2000	22	4	48	18	0.67	0.4010
Rosario 2004	90	13	65	18	0.67	0.4473
RCTs with mixed surgical status						
Creutzig 1987	10	4	10	5	0.67	0.6923
Gawkowska-Suwinska 2001	44	6	54	25	0.75	0.1953
Johansen 1991	27	13	36	15	0.67	0.1294
Bal 1996	38	10	27	10	0.83	0.6140
Sirisalipoch 2004	75	10	63	22	0.33	0.2957
Bal 2004	77	4	73	12	0.83	1.1235

Table 5 Parameter estimates using various methods mentioned above from a meta-analysis of 22 trails of stimate of the association of radioactive iodine dosage for the ablation of thyroid remnants

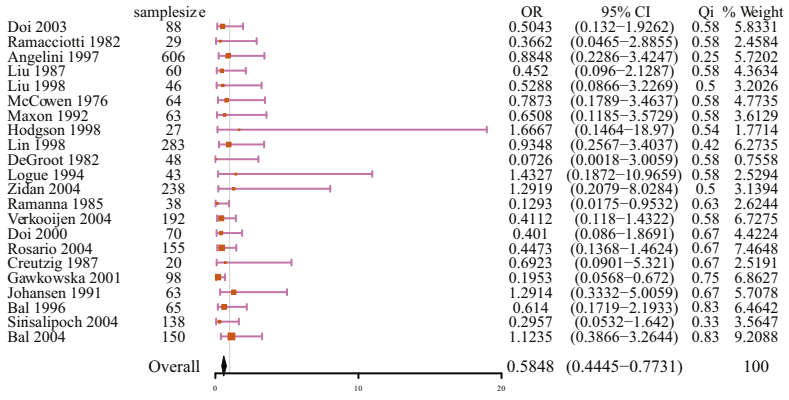
Model	heterogeneity between studies	OR	Two-sided 95%	width	Variance	DIC
Model RE	$\widehat{\tau}^2$=0.2308 (0.1684)	0.5887	[0.4342, 0.7980]	0.3638	0.0241	47.44
Model QE		0.5803	[0.4248, 0.7928]	0.3680	0.0253	–
Model RE-s	$\widehat{\tau}^2$=0.1398 (0.1294)	0.6009	[0.4622, 0.7846]	0.3224	0.0182	40.814
Model QE-v		0.5848	[0.4445, 0.7731]	0.3286	0.0199	–

Note: "-" denotes that DIC is absent in the corresponding model; Variance means the variance of OR.

It is also appealing for investigators to know to what extent the quality influences the overall treatments. When the meta-analysis linked with quality is implemented, there are less extreme values for both QE model estimator and QE-v model estimator ($OR = 0.5803$; $95\%CI$ $0.4248 - 0.7928$ and $OR = 0.5848$; $95\%CI$ $0.4445 - 0.7731$). What's more, because of the utilization of the scale parameter, ψ_i, the variances of the overall effect estimates go up as expected. If, however, the meta-analysis estimate is calculated using the previous QE model which only considers the design-related heterogeneity (quality), there are a more conservative confidence interval and larger variance of the overall effect estimate than that in our proposed QE-v model, which consequently suggests that incorporation of the quality factor and some detectable variations between study effects can produce a more precise estimate of the effect size. The picture the QE-v estimate depicts (Figure 2) supports for the results that the weights in Formula (2) make for the decline of study weights for the larger but poorer quality study



(a) QE model



(b) QE-v model

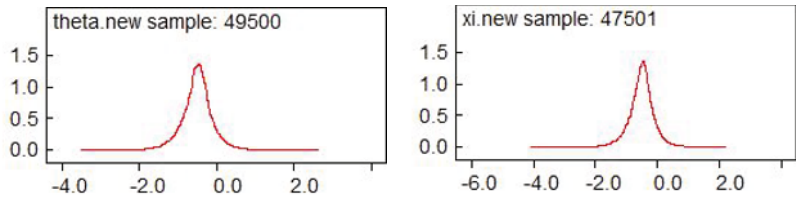
Figure 2 Forest plot of failure of Ablation of the thyroid remnant in a comparison of high-dose radioiodine (treatment group) with low dose radioiodine (control group) using QE model (a) and QE-v model (b). QE-v model does not simply favor smaller studies or larger studies of low quality

(Angelini), and some small studies, with the except of DeGroot study as there was no failure

in the high-dose group. Furthermore, it is easily noted that the QE-v estimator progressively increases the weight in Bal study (2004) which is of high quality based on the additional inclusion of design-unrelated study heterogeneity. In summary, our analysis suggests that high dose of 75 to 100 mCi (2775-3700 MBq) of patients has about forty percent, less failure of non-ablation than low-dose patients.

5 Discussion

We have first explored another flexible distribution for random effect in hierarchical models when the data is not approximately normally distributed. Here, a skew-normal distribution for random effect is chosen even if the random effects are truly normal, and it does not misrepresent the data because the normal distribution is a special case of the skew-normal distribution (Table 2). However, if the normal assumption is still selected for the skewed data, the inappropriate statistical model could result in a biased result (Table 3). Apart from the factor analyzed above, as shown in Figure 3, the predictive distribution for the treatment effect is slightly skewed under the skew assumption, which, consequently, provides evidence that the skew-normal distribution favors the empirical example data.



(a) Normal distribution (b) Skew-normal distribution

Figure 3 Predictive distributions for the treatment effect in a new trial in the effectiveness of radioactive iodine dosage for the ablation of thyroid remnants: (a) normal distribution; (b) skew-normal distribution

Note: To simply analyze the effects of the predictive distribution for the mean outcome, we take $\gamma_i = \gamma, i = 1, 2, \cdots, 22$, which illustrates the underlying effects are exchangeable.

The QE-v model estimate differs from the QE model estimate in the following perspectives. In simulation studies, the performance of the QE-v model is determined by the magnitude of effect sizes. When the outcomes of the studies approximately follow a normal distribution, our proposed model works better than the QE model in most cases, except for $OR = 0.9$ and $OR = 1.2$, while if the effects are skewed, the QE-v model should be adopted at $OR = 0.3$ and $OR = 0.6$. In the practical example, it is obvious that the QE-v model yields a more precise estimate of the effect size when more types of heterogeneities are taken into consideration. Besides, the weight of the larger studies is redistributed to small studies only if their quality is deemed lower, as well as the QE-v estimator also down-weights some smaller studies whose sample sizes are less than 45 (Figure 2).

In a set of heterogeneous studies, a QE-v model makes an adjustment to the study weight according to the study quality and underlying effect variation, whilst a QE model only exploits

quality information to construct the varying weight and an RE model is intended primarily for heterogeneity that can not be explained. A key difficulty in the assessment of quality is the obstacle provided by the incomplete reporting liable to the subjectivity. Further, it is well-known that a study may be performed to the highest possible standards yet still have an important risk of bias that is suggested to be noticed, for instance, in many situations. Even so, it is impractical or impossible to blind participants or study personnel of intervention group, and it is inappropriately judgmental to describe all such studies as of low quality[34]. Beyond these issues, there is difficulty in distinguishing whether heterogeneity results from clinical or methodological diversity including study design (quality) and risk of bias, which probably leads to a double-counting of "quality" term. In consequence, more attention should be paid when research quality is applied.

In conclusion, though the limitations discussed above may be inevitable, it is still advisable to take on our model if quality information of studies is available, especially in a situation where the effect sizes follow an approximately normal distribution. However, if the quality information available is limited, the random effect model with a skew-normal assumption would be the preferred one.

References

[1] Sutton A J and Higgins J P T, Recent development in meta-analysis, *Statistics in Medicine*, 2008, **27**(5): 625–650.

[2] Hozo S P, Djulbegovic B, and Hozo I, Estimating the mean and variance from the median, range, and the size of a sample, *BMC Medical Research Methodology*, 2005, **5**(1): 13.

[3] Wan X, Wang W, Liu J, et al., Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range, *BMC Medical Research Methodology*, 2014, **14**(1): 135–135.

[4] Luo D, Wan X, Liu J, et al., Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range, *Statistical Methods in Medical Research*, 2018, **27**(6): 1785–1805.

[5] DerSimonian R and Laird N, Meta-analysis in clinical trials, *Controlled Clinical Trials*, 1986, **7**(3): 177–188.

[6] Senn S, Trying to be precise about vagueness, *Statistics in Medicine*, 2007, **26**(7): 1417–1430.

[7] Verhagen A P, de Vet H C W, de Bie R A, et al., The art of quality assessment of RCTs included in systematic reviews, *Journal of Clinical Epidemiology*, 2001, **54**(7): 651–654.

[8] Berard A and Bravo G, Combining studies using effect sizes and quality scores: Application to bone loss in postmenopausal women, *Journal of Clinical Epidemiology*, 1998, **51**(10): 801–807.

[9] Tritchler D L, Modelling study quality in meta-analysis, *Statistics in Medicine*, 1999, **18**(16): 2135–2145.

[10] Fleiss J L and Gross A J, Meta-analysis in epidemiology, with special reference to studies of

the association between exposure to environmental tobacco smoke and lung cancer: A critique, *Journal of Clinical Epidemiology*, 1991, **44**(2): 127–139.

[11] Shadish W R and Haddock C K, Combining estimates of effect size, Eds. by Cooper H, Hedges L V, and Valentine J C, The handbook of research synthesis and meta-analysis (p.257 C277). Russell Sage Foundation, 2009.

[12] Doi S A R and Thalib L, A quality-effects model for meta-analysis, *Epidemiology*, 2008, **19**(1): 94–100.

[13] Doi S A R, Barendregt J J, and Mozurkewich E L, Meta-analysis of heterogeneous clinical trials: An empirical example, *Contemporary Clinical Trials*, 2011, **32**(2): 288–298.

[14] Doi S A R, Barendregt J J, Khan S, et al., Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model, *Contemporary Clinical Trials*, 2015, **45**: 123–129.

[15] Doi S A R, Barendregt J J, Khan S, et al., Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model, *Contemporary Clinical Trials*, 2015, **45**: 130–138.

[16] Greenland S and O'Rourke K, On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions, *Biostatistics*, 2001, **2**(4): 463–471.

[17] Ahn S and Becker B J, Incorporating quality scores in meta-analysis, *Journal of Educational and Behavioral Statistics*, 2011, **36**(5): 555–585.

[18] Lee K J and Thompson S G, Flexible parametric models for random-effects distributions, *Statistics in Medicine*, 2008, **27**(3): 418–434.

[19] Arellano-Valle R B, Bolfarine H, and Lachos V H, Skew-normal linear mixed models, *Journal of Data Science*, 2005, **3**(4): 415–438.

[20] Lin T I and Lee J C, Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data, *Statistics in Medicine*, 2008, **27**(9): 1490–1507.

[21] Wu M, Zhao J, Wang T, et al., The ANOVA-type inference in linear mixed model with skew-normal error, *Journal of Systems Science and Complexity*, 2017, **30**(3): 710–720.

[22] Macgillivray H L, Skewness and asymmetry: Measures and orderings, *The Annals of Statistics*, 1986, **14**(3): 994–1011.

[23] Wen F and Yang X, Skewness of return distribution and coefficient of risk premium, *Journal of Systems Science and Complexity*, 2009, **22**(3): 360–371.

[24] Bailey K R, Inter-study differences: How should they influence the interpretation and analysis of results? *Statistics in Medicine*, 1987, **6**(3): 351–358.

[25] Turner R M, Spiegelhalter D J, Smith G C S, et al., Bias modelling in evidence synthesis, *Journal of the Royal Statistical Society: Series A* (*Statistics in Society*), 2009, **172**(1): 21–47.

[26] Thompson S, Ekelund U, Jebb S, et al., A proposed method of bias adjustment for meta-analyses of published observational studies, *International Journal of Epidemiology*, 2010, **40**(3): 765–777.

[27] Spiegelhalter D J and Best N G, Bayesian approaches to multiple sources of evidence and uncertainty in complex cost effectiveness modelling, *Statistics in Medicine*, 2003, **22**: 3687–3709.

[28] Doi S A R, Evidence synthesis for medical decision making and the appropriate use of quality scores, *Clinical Medicine & Research*, 2014, **12**(1–2): 40–46.

[29] Kulinskaya E and Olkin I, An overdispersion model in meta-analysis, *Statistical Modelling*, 2014, **14**(1): 49–76.

[30] Pewsey A, Problems of inference for Azzalini's skewnormal distribution, *Journal of Applied Statistics*, 2000, **27**(7): 859–870.

[31] Bayes C L and Branco E, Bayesian inference for the skewness parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability and Statistics*, 2007, **21**(2): 141–163.

[32] Gelman A, Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Analysis*, 2006, **1**(3): 515–534.

[33] Doi S A R, Woodhouse N J, Thalib L, et al., Ablation of the thyroid remnant and I-131 dose in differentiated thyroid cancer: A meta-analysis revisited, *Clinical Medicine and Research*, 2007, **5**(2): 87–90.

[34] Higgins P T and Green S, *Cochrane Handbook for Systematic Reviews of Interventions*, John Wiley and Sons, New Jersey, 2011.

[35] Arellano-Valle R B, Gmez, Hctor W, and Quintana F A, A new class of skew-normal distributions, *Communications in Statistics — Theory and Methods*, 2004, **33**(7): 1465–1480.

[36] Azzalini A, A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 1985, **12**(2): 171–178.

[37] Henze N, A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, 1986, **13**(4): 271–275.

[38] Gupta R C and Brown N, Reliability studies of the skew-normal distribution and its application to a strength-stress model, *Communications in Statistics — Theory and Methods*, 2001, **30**(11): 2427–2445.

Appendix

Appendix A Skew-Normal Distribution

The concept and role of the skew-normal distribution may be unfamiliar for some readers, therefore, some explanations, as well as details for the discussion mentioned above are added. The concise mathematical definition of skew-normal distribution is firstly introduced.

A random variable Y has a skew-normal distribution if its probability density function (pdf) is given by

$$g(y) = \frac{2}{\alpha} f\left(\frac{y-\mu}{\alpha}\right) F\left(\lambda \frac{y-\mu}{\alpha}\right), \quad -\infty < y < +\infty,$$

where f and F are the pdf and cumulated distribution function (cdf) of a standard normal, respectively, λ is the skewness parameter, μ is the location parameter and α is the scale parameter[35]. For brevity, we shall also say that $Y \sim SN(\mu, \alpha^2, \lambda)$.

If $\mu = 0$ and $\alpha = 1$, it is reduced to a standard skew-normal distribution, i.e., $Y \sim SN(\lambda)$[36]. The following properties follow immediately from the standard skew-normal distribution[35–38].

Property A The SN(0) density is the $N(0, 1)$ density.

Property B The density is strongly unimodal, i.e., $\log g(y)$ is a concave function of y.

Property C Let U and V be independent standard random variables, then

$$Y = a \mid U \mid + bV \sim SN(\lambda),$$

where $a = \frac{\lambda}{\sqrt{1+\lambda^2}}, b = \frac{1}{\sqrt{1+\lambda^2}}$.

Property D The expectation and variance of Y are:

$$E(Y) = \sqrt{\frac{2}{\pi}}a,$$
$$\mathrm{Var}(Y) = 1 - \left(\sqrt{\frac{2}{\pi}}a\right)^2.$$

The skewness index for the standard skew-normal distribution is given by:

$$\lambda^* = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\left(\frac{\lambda}{\sqrt{1+\lambda^2}}\right)^3\left(1 - \frac{2}{\pi}\frac{\lambda^2}{1+\lambda^2}\right)^{-\frac{3}{2}}.$$

Thus, it can be seen that $\lambda^* < 0.9953$.

Appendix B Computation of y_i

Using Property C, we have

$$y_i = \frac{\tau\gamma_i}{\sqrt{1+\gamma_i^2}}t_i + s_i, \quad i = 1, 2, \cdots, n,$$

where t_i is a half standard normal distribution, and $s_i \sim N(\mu, \frac{\tau^2}{1+\gamma_i^2} + \delta_i^2)$, t_i and s_i are independently mutually.

Since the cumulative distribution function of y_i is:

$$F_{Y_i}(y_i) = \int_0^{+\infty}\int_{-\infty}^{y_i - a_i t_i} f(t_i, s_i)ds_i dt_i,$$

where $a_i = \frac{\tau\gamma_i}{\sqrt{1+\gamma_i^2}}$, then the probability density function:

$$f_{Y_i}(y_i) = \int_0^{+\infty} f(t_i, y_i - a_i t_i)dt_i.$$

By Convolution formula and simple algebra, we can show that

$$\begin{aligned}
f_{Y_i}(y_i) &= \frac{1}{\pi\sigma_i}\int_0^{+\infty} \mathrm{e}^{-\frac{1}{2}[t_i^2 + \frac{(y_i - a_i t_i - \mu)^2}{\sigma_i^2}]}dt_i \\
&= \frac{1}{\pi\sigma_i}\mathrm{e}^{-\frac{1}{2}\frac{(y_i-\mu)^2}{a_i^2+\sigma_i^2}}\int_0^{+\infty}\mathrm{e}^{-\frac{(t_i - \frac{(y_i-\mu)a_i}{a_i^2+\sigma_i^2})^2}{2\frac{\sigma_i^2}{a_i^2+\sigma_i^2}}}dt_i \\
&= \frac{1}{\pi\sqrt{a_i^2+\sigma_i^2}}\mathrm{e}^{-\frac{1}{2}\frac{(y_i-\mu)^2}{a_i^2+\sigma_i^2}}\int_{-\frac{(y_i-\mu)a_i}{\sigma_i\sqrt{a_i^2+\sigma_i^2}}}^{+\infty}\mathrm{e}^{-\frac{1}{2}s^2}ds \\
&= \frac{2}{\sqrt{\pi(a_i^2+\sigma_i^2)}}\mathrm{e}^{-\frac{1}{2}\frac{(y_i-\mu)^2}{a_i^2+\sigma_i^2}}F\left(\frac{(y_i-\mu)a_i}{\sigma_i\sqrt{a_i^2+\sigma_i^2}}\right) \\
&= 2f\left(\frac{y_i-\mu}{\sqrt{\tau^2+\delta_i^2}}\right)F\left(\frac{\tau\gamma_i}{\sqrt{\tau^2+\delta_i^2(1+\gamma_i^2)}}\frac{y_i-\mu}{\sqrt{\tau^2+\delta_i^2}}\right), \qquad (3)
\end{aligned}$$

where $\sigma_i^2 = \frac{\tau^2}{1+\gamma_i^2} + \delta_i^2$.