ORIGINAL ARTICLE

Meta-analysis of published excess relative risk estimates

David B. Richardson[1] · Kossi Abalo[2] · Marie-Odile Bernier[2] · Estelle Rage[2] · Klervi Leuraud[2] · Dominique Laurier[2] · Alexander P. Keil[1] · Mark P. Little[3]

Abstract
A meta-analytic summary effect estimate often is calculated as an inverse-variance-weighted average of study-specific estimates of association. The variances of published estimates of association often are derived from their associated confidence intervals under assumptions typical of Wald-type statistics, such as normality of the parameter. However, in some research areas, such as radiation epidemiology, epidemiological results typically are obtained by fitting linear relative risk models, and associated likelihood-based confidence intervals are often asymmetric; consequently, reasonable estimates of variances associated with study-specific estimates of association may be difficult to infer from the standard approach based on the assumption of a Wald-type interval. Here, a novel method is described for meta-analysis of published results from linear relative risk models that uses a parametric transformation of published results to improve on the normal approximation used to assess confidence intervals. Using simulations, it is illustrated that the meta-analytic summary obtained using the proposed approach yields less biased summary estimates, with better confidence interval coverage, than the summary obtained using the more classical approach to meta-analysis. The proposed approach is illustrated using a previously published example of meta-analysis of epidemiological findings regarding circulatory disease following exposure to low-level ionizing radiation.

Keywords Meta-analysis · Cohort studies · Excess relative risk · Cancer

Abbreviations
RR   Relative rate
CI   Confidence interval
ERR   Excess relative risk
Sv   Sievert

✉ David B. Richardson
david.richardson@unc.edu

1 Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

2 Institut de Radioprotection et de Sûreté Nucléaire, Paris, France

3 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Introduction

In a meta-analysis of epidemiological study results, a summary effect estimate is obtained by combining information from a set of study-specific estimates. A common approach is to calculate an inverse-variance-weighted average of the study-specific estimates of association [e.g., Sutton et al. (2000), United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (2018)]. This approach assigns more weight to studies with more precise study-specific estimates of association. In the context of unbiased linear sums of estimates, this approach is justified by the Gauss–Markov theorem (Plackett 1949).

For epidemiological results obtained from fitting log-linear regression models, it is easy to recover reasonable estimates of the variances associated with study-specific estimates of association using the information encoded in the confidence intervals. Standard meta-analytic techniques typically proceed by assuming that, given reported-effect measures and associated confidence intervals, one can derive the variances of estimates of association based on the assumptions typical of Wald-type statistics; these estimates of study-specific variances are used to calculate the

inverse-variance-weighted average estimate of association which is reported as the summary effect estimate (Sutton et al. 2000).

However, such an approach is not straightforward for some estimators for which variances are rarely reported. In some application areas, effect measures are typically obtained from fitting linear relative risk regression models. For example, in epidemiological studies of a variety of carcinogens, including asbestos (Hein et al. 2007), benzene (Rinsky et al. 2002), radon progeny (National Research Council (U.S.) et al. 1988; Lubin et al. 1995; Darby et al. 2005), and external ionizing radiation (Boice et al. 1987; National Research Council (U.S.) et al. 1990; Preston et al. 2003), investigators have modeled the relative risk per unit exposure as a linear function of exposure rather than an exponential function of exposure. In radiation research, this convention follows from a long history of use of the linear relative risk model in analyses of the Life Span Study of Japanese atomic bomb survivors (Preston et al. 2007; Pawel et al. 2008) for which there is a biophysical basis (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) 1993; Little et al. 2009, Little 2010); and, in the contemporary epidemiological literature, the linear relative risk model has been applied in analyses of many radiation-exposed populations (Gilbert et al. 1993; Cardis et al. 1995, 2005; Muirhead et al. 2009; Metz-Flamant et al. 2013). The widespread use of the same model form has the advantage that it may facilitate comparison of results between studies. Unfortunately, a quantitative meta-analytic summarization of epidemiological results that have been quantified using linear relative risk models is more challenging than doing so with the results that have been quantified using standard log-linear regression models.

An important challenge in meta-analyses of results that have been quantified using a linear relative risk model is deriving reasonable estimates of study-specific variances. The methodology developed for quantitative summaries of epidemiological findings has largely focused on log-linear model forms, where symmetric Wald-type confidence intervals are routinely reported (DerSimonian and Laird 1986). In contrast, likelihood-based confidence intervals are commonly reported for estimates of association derived from linear relative risk models, and often these intervals are asymmetric (Cox and Hinkley 1974, Meeker and Escobar 1995). Consequently, a reasonable estimate of the variance associated with a point estimate may be difficult to infer from the information encoded in the likelihood-based confidence bounds by simply leveraging the assumptions typical of Wald-type statistics.

In this paper, we describe a method to address these challenges to meta-analysis of published studies that report estimates of association derived from linear relative risk models. The approach is based on an algebraic transformation of published results to yield an estimator with a more symmetrical distribution than that reported in the literature, and then derive an expression of variance of this transformed estimator assuming that the reported profile likelihood bounds for the estimate of association in the original scale conform well to a re-expression of Wald-type bounds of the transformed estimator. The effect on meta-analyses of non-normality in study-specific estimates has been recognized by prior authors (Jackson and White 2018); and, the Cochrane Handbook, for example, discusses transformation of results as an approach to reduce skew (Higgins and Cochrane Collaboration 2020). A meta-analytic summary and associated confidence interval are constructed and back transformed to the original scale. We address fixed-effect and random-effects meta-analyses; these approaches employ different assumptions (i.e., under a fixed-effect model, it is assumed that there is one true association that underlies all the studies in the analysis; and, under the random-effects model, it is assumed that there is an underlying distribution of true associations across studies). For illustration, the proposed methodology is implemented using an empirical example.

Methods

We assume that a systematic literature search was performed, study results (in terms of point estimates and associated confidence intervals) extracted, and study quality appraised. These important steps in a meta-analysis are not addressed here. Rather, here the focus is on the stage of data synthesis during which a quantitative summary of the study findings is calculated. We focus on a setting where epidemiological results have been obtained by fitting a model of the form

$$\psi = 1 + \beta D,$$

where ψ denotes the risk ratio or odds ratio, D denotes the continuous exposure of interest, and the parameter of primary interest in the meta-analysis, β, denotes the excess relative risk or excess odds ratio per unit D (e.g., the excess relative risk per sievert (Sv) in a radiation epidemiology study), and likelihood-based confidence intervals have been reported.

For simplicity we will henceforth assume 95% confidence intervals but the approach is readily adapted to other bounds. First, a standard approach to meta-analytic summarization of epidemiological study results is described below. Second, the proposed alternative approach to meta-analytic summarization of epidemiological study results is described. Third, it is addressed how to proceed with a meta-analysis of results that have been quantified using a linear risk ratio model

when a lower confidence bound was not determined for the reported estimate.

A standard approach to meta-analysis of published linear relative risk estimates

Typically, the data structure for a standard approach to summarization of epidemiological findings in a meta-analysis consists of a table of point estimates and associated confidence intervals. Let $i = 1 \ldots k$ index the k studies to be summarized in the meta-analysis. Let $\hat{\beta}_i$ denote the estimated excess relative risk or excess odds ratio per unit D for study i; and, let L_i and U_i denote the associated lower and upper confidence limits for $\hat{\beta}_i$.

For each study, i, we derive the standard error of the reported estimate of association, denoted se $(\hat{\beta}_i)$ given the reported associated confidence intervals L_i, U_i for the published results, by the following calculation: se $(\hat{\beta}_i) = (U_i - L_i)/(2 \times 1.96)$.

This approach to estimation of the study-specific standard error follows from considering the framework typical of a linear regression model fitting that yields a point estimate $(\hat{\beta}_i)$ and associated Wald-type confidence bounds (L_i, U_i). Given this information, an estimate of se $(\hat{\beta}_i)$ can be derived under the conditions typical of Wald-type statistics: $L_i = \hat{\beta}_i - 1.96 \times$ se $(\hat{\beta}_i)$ and $U_i = \hat{\beta}_i + 1.96 \times$ se $(\hat{\beta}_i)$. With simple rearrangement one gets, $L_i + 1.96 \times$ se $(\hat{\beta}_i) = U_i - 1.96 \times$ se $(\hat{\beta}_i)$, and it follows that $(U_i - L_i) = 2 \times (1.96$ se $(\hat{\beta}_i))$, leading to the above expression for se$(\hat{\beta}_i)$ as a function of (L_i, U_i).

Little et al. (2012) described an approach to deriving a fixed-effect inverse-variance-weighted estimate of the excess relative risk per unit exposure, where the meta-analytic summary is calculated as the sum of the study-specific estimates divided by its variance, over the sum of the inverse of the study-specific variances, $\hat{\beta}_{tot}^{Fixed} = \frac{\sum_{i=1}^{k} \hat{\beta}_i / \text{se}(\hat{\beta}_i)^2}{\sum_{i=1}^{k} 1 / \text{se}(\hat{\beta}_i)^2}$. This inverse-variance-weighted average estimate of association is reported as the summary estimate of association.

Confidence intervals for this fixed-effect summary estimate of association are derived by calculation of an estimate of the standard error for the meta-analytic summary association. The estimate of the standard error is simply the reciprocal of the square-root of the sum of the study-specific variances,

$$se(\hat{\beta}_{tot}^{Fixed}) = \frac{1}{\left[\sum_{i=1}^{k} / \text{se}(\hat{\beta}_i)^2\right]^{0.5}}; \text{ and,}$$

a Wald-type confidence interval for the summary estimate of association is derived as, 95% CI $(\hat{\beta}_{tot}^{Fixed}) = \hat{\beta}_{tot}^{Fixed} \pm 1.96 \times$ se $(\hat{\beta}_{tot}^{Fixed})$.

A random-effects summary estimate of association may also be derived. Little et al. (2012) described how to calculate a random-effects summary estimate of association based on the method proposed by DerSimonian and Laird (1986) for a one-step estimation of the variance of the random effect, where the summary meta-analytic estimate of the association based on a random-effects model is calculated as,

$$\hat{\beta}_{tot}^{Random} = \frac{\sum_{i=1}^{k} \hat{\beta}_i / \left[\text{se}(\hat{\beta}_i)^2 + \Delta^2\right]}{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{\beta}_i)^2 + \Delta^2\right]},$$

where

$$\Delta^2 = \max\left[0, \frac{Q - (k-1)}{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{\beta}_i)^2\right] - \frac{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{\beta}_i)^4\right]}{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{\beta}_i)^2\right]}}\right],$$

and

$$Q = \sum_{1}^{k} \left[(\hat{\beta}_i - \hat{\beta}_{tot}^{Fixed}) / \text{se}(\hat{\beta}_i)\right]^2.$$

The associated standard error is calculated as,

$$se(\hat{\beta}_{tot}^{Random}) = \frac{1}{\left[\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{\beta}_i)^2 + \Delta^2\right]\right]^{0.5}}.$$

An alternative approach to meta-analysis of published linear relative risk estimates

In most contemporary epidemiological analyses that quantify associations under a linear relative risk model, the reported confidence interval is derived from likelihood-based methods rather than calculated as a Wald-type interval (McCullagh and Nelder 1989). This is because, in a given study, the distribution of maximum likelihood estimators for the parameter β may be far from normal unless the sample size is large. When maximum likelihood estimators are not approximately normal (e.g., in small or moderate samples), Wald-type intervals may not have nominal coverage (Cox and Hinkley 1974; Meeker and Escobar 1995); for this reason, it has become common practice for published results for fittings of linear relative risk models to report likelihood-based confidence intervals rather than Wald-type intervals (Prentice and Mason 1986; Moolgavkar and Venzon 1987). By extension, meta-analytic summaries that proceed under the assumption that typical Wald-type statistical assumptions hold may not

yield an appropriately inverse-variance-weighted average estimate of association or confidence interval.

We describe an alternative approach to meta-analytic summarization of epidemiological study results that have been obtained from fitting of linear relative risk regression models. The data structure for the proposed approach includes a table of point estimates, associated confidence intervals, and maximum observed doses. Letting $i = 1 \ldots k$ index study, $\hat{\beta}_i$ denotes the point estimate of interest, and L_i and U_i denote the associated lower and upper confidence bounds for the point estimate reported for study i. Further, let x_i denote the maximum value of the dose reported in published study i, noting that the value of x_i is often known and reported in an epidemiological study (or may be obtained from the authors).

For each study, we derive a transformed metric of the estimate of association (Eq. 1),

$$\hat{A}_i = \ln(c\hat{\beta}_i + 1), \tag{1}$$

and associated standard error, denoted $se(\hat{A}_i)$, as a function of reported values L_i and U_i (Eq. 2)

$$se(\hat{A}_i) = \ln\left(\frac{cU_i + 1}{cL_i + 1}\right)/(2 \times 1.96), \tag{2}$$

where $c = \min[x_i : 1 \le i \le k]$, to ensure that $\left(\frac{cU_i+1}{cL_i+1}\right) > 0$ for $1 \le i \le k$ and therefore, $se(\hat{A}_i)$ can be calculated for all studies in the meta-analysis. The proposed approach derives a variance estimate for this transformed quantity that is based on the reported likelihood-based confidence interval for the estimate of association on its original scale; however, estimates of the transformed quantity $\hat{A}_i$ will tend to more closely approximate a normal distribution than $\hat{\beta}_i$.

The justification for the proposed approach follows from considering that parameter transformations can improve asymptotic distributional approximations, as discussed in the context of the linear relative risk model by Barlow (1985a, b) and by Prentice and Mason (1986). Criteria for selecting such transformations have been discussed previously (Sprott 1974) and include removal of range restriction on $\hat{\beta}_i$ and reduction of the asymmetry of the log-likelihood about $\hat{\beta}_i$.

The transformation $\hat{A}_i = \ln(c\,\hat{\beta}_i + 1)$ can remove the range restriction on the excess relative risk parameters $\hat{\beta}_i$. Consider study i in which the dose variable, D_i, has compact support C_i, for which $x_i = \sup[C_i]$. Consequently, the possible range of the estimate of dose–response association, $\hat{\beta}_i$, under a model $RR_i = 1 + \beta_i D_i$, is $1/-x_i$, infinity). Prentice and Mason (1986) proposed the simple transformation $\alpha = \ln(\beta + \beta_0)$, where $\beta_0 = \frac{1}{x_i}$ to remove the range restriction; when $c = x_i$, that simple transformation is equivalent to our proposed expression $\ln(c\beta + 1)$.

The transformation $\hat{A}_i = \ln(c\,\hat{\beta}_i + 1)$ also may reduce log-likelihood skewness, and, therefore, improve symmetry of confidence bounds on the transformed metric. The standard approach assumes that typical Wald-type statistical assumptions hold. Prentice and Mason illustrated that the simple transformation $A = \ln(\beta + \beta_0)$ yields nearly complete symmetry about the transformed metric (Prentice and Mason 1986); and, when $c = x_i$, that simple transformation is equivalent to our proposed expression $A = \ln(c\beta + 1)$. Using Sprott's index as a measure of the normality of the likelihood function, Barlow demonstrated that the transformation $A = \ln(\beta + 1)$ improves the normality of estimates (Barlow 1985a, b). Our proposed transformation is equivalent to the transformation proposed by Barlow (1985a, b) when $c = 1$. It follows that the proposed transformation will tend to improve the symmetry of the likelihood-based confidence bounds on the transformed scale, and upon applying this transformation to the reported bounds, L and U, one can better approximate the variance by employing assumptions of Wald-type intervals to the likelihood-based bounds on this transformed scale than when applied to these bounds on their original scale.

A fixed-effect inverse-variance-weighted summary of this estimated quantity is calculated as follows, $\hat{A}_{tot}^{Fixed} = \frac{\sum_{i=1}^{k} \hat{A}_i / se(\hat{A}_i)^2}{\sum_{i=1}^{k} 1 / se(\hat{A}_i)^2}$.

The standard error of $\hat{A}_{tot}^{Fixed}$ is given by $se(\hat{A}_{tot}^{Fixed}) = \frac{1}{\left[\sum_{i=1}^{k} 1 / se(\hat{A}_i)^2\right]^{0.5}}$.

We then re-transform to the original scale and obtain the summary fixed-effect meta-analytic estimate of the association $\hat{\beta}_{tot}^{Fixed}$, and associated confidence interval. This summary estimate is calculated as:

$$\hat{\beta}_{tot}^{Fixed} = \frac{\exp(\hat{A}_{tot}^{Fixed}) - 1}{c},$$

and it is this form of the inverse-variance-weighted average estimate of association that is reported as the meta-analytic summary estimate of association.

A Wald-type confidence interval for the summary estimate of association is derived based on the estimate of se $(\hat{A}_{tot}^{Fixed})$,

$$95\% \text{ CI } (\hat{\beta}_{tot}^{Fixed}) = \frac{\exp(\hat{A}_{tot}^{Fixed} \mp 1.96 se(\hat{A}_{tot}^{Fixed})) - 1}{c}.$$

A similar approach can be used to calculate a random-effects summary meta-analytic estimate and associated confidence interval. Let $\hat{A}_{tot}^{Random}$ denote the random-effects inverse-variance-weighted summary, calculated as follows,

$$\hat{A}_{\text{tot}}^{\text{Random}} = \frac{\sum_{i=1}^{k} \hat{A}_i / \left[\text{se}(\hat{A}_i)^2 + \Delta^2 \right]}{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{A}_i)^2 + \Delta^2 \right]}, \text{ where}$$

$$\Delta^2 = \max \left[0, \frac{Q - (k-1)}{\sum_{i=1}^{k} 1 \bigg/ \left[\text{se}(\hat{A}_i)^2 \right] - \frac{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{A}_i)^4 \right]}{\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{A}_i)^2 \right]}} \right], \text{ and}$$

$$Q = \sum_{1}^{k} \left[(\hat{A}_i - \hat{A}_{\text{tot}}^{\text{Fixed}}) / \text{se}(\hat{A}_i) \right]^2.$$

The associated standard error for this meta-analytic summary of the transformed estimates is,

$$\text{se}\left(\hat{A}_{\text{tot}}^{\text{Random}} \right) = \frac{1}{\left[\sum_{i=1}^{k} 1 / \left[\text{se}(\hat{A}_i)^2 + \Delta^2 \right] \right]^{0.5}}.$$

We then re-transform to the original scale, and obtain the summary random-effects meta-analytic estimate of association and associated confidence interval, calculated as:

$$\hat{\beta}_{\text{tot}}^{\text{Random}} = \frac{\exp\left(\hat{A}_{\text{tot}}^{\text{Random}} \right) - 1}{c},$$

with a Wald-type confidence interval for this random-effects summary estimate of association is derived,

$$95\% \text{ CI } (\hat{\beta}_{\text{tot}}^{\text{Random}}) = \frac{\exp(\hat{A}_{\text{tot}}^{\text{Random}} \mp 1.96 \text{se}(\hat{A}_{\text{tot}}^{\text{Random}})) - 1}{c}.$$

A simple computer code written for the SAS and R statistical packages is provided that calculates fixed and random-effects meta-analytic summary estimates as well as associated confidence intervals (Electronic Supplementary Material).

Meta-analysis when a lower confiden e bound was not determined in a published report

Sometimes, a likelihood-based lower confidence bound, L_i, is not determined in a particular analysis because, at the lower constraint on the parameter range (i.e., $1/-x_i$), the likelihood-based statistic that defines the lower confidence bound has not reached the specified critical value. In such instances, a lower bound is typically not reported; rather, authors may indicate that the bound is simply $< -1/x_i$.

To date, practice for how to address this has not been well described in the literature. It appears that what is done in standard meta-analyses of linear relative risk estimates is to impute a lower bound by assuming that the

confidence bounds are symmetrical on the original scale (Little et al. 2012), such that the imputed lower bound is $L_i' = \hat{\beta}_i - (U_i - \hat{\beta}_i)$. The standard approach then proceeds using L_i' in place of the missing L_i.

When using the proposed alternative method, a lower bound may be imputed by assuming symmetry on the transformed scale. If the transformation improves normality as compared to the original scale, this should be advantageous. For study i with no reported lower bound L_i, impute the value $L_i' = \frac{\exp(\hat{A}_i - (\ln(cU_i + 1) - \hat{A}_i)) - 1}{c}$. The data set can proceed with analysis described above using the imputed lower bound.

Sensitivity to observed exposure ranges

The proposed transformation involves selection of a constant, c, to ensure that one can calculate se $(\hat{A}_i)$ for all studies in the meta-analysis. The proposed approach defines $c = \min[x_i : 1 \le i \le k]$. The sensitivity of results to choice of c can be assessed by recalculating the meta-analytic summary measure under an alternative value, c', under the constraint $0 < c' \le \min [x_i : 1 \le i \le k]$, to ensure calculation of se $(\hat{A}_i)$. This permits investigation of sensitivity of results, for example, to outliers or extreme values of the exposure variable in study samples. We suggest such sensitivity analyses proceed by calculating results under a value such as, $c' = 0.9c$.

Simulations

The proposed approach is compared here to meta-analytic summaries to the standard fixed-effect and random-effects approaches in simulated data examples. For this, we simulated 1000 meta-analyses under scenarios in which the number of individuals in each study was small (1000–1500), moderate (2000–2500), large (4000–4500), or variable size (1000–4500); additionally, examples were considered in which the number of studies in a simulated meta-analysis was set to 5, 10, or 15 studies. In each simulation, the number of people in a study was drawn from a uniform distribution over the specified range of study size; for each cohort member, we generated an independent standard normal covariate Z. An exposure, E, was generated by sampling from a uniform distribution (0, 5). We generated a binary outcome, Y, with dependence of Y on Z and E encoded by specifying that Y took a value of 1 with odds $= \exp(\log(\alpha) + 0.1Z) \times (1 + \eta E)$, where the parameter describing the baseline odds of the outcome, α, was set to 0.15, 0.2, or 0.25, and the parameter describing the excess odds ratio per unit E, η was set to 1.0, 0.75, or 0.50. The excess odds ratio model was used for data generation to avoid numerical issues with generating data from a linear relative risk model. The coefficients associated with E for

each study were estimated using maximum likelihood and profile likelihood confidence bounds were obtained. We calculated fixed-effect and random-effects meta-analyses using the standard approach; and, we calculated fixed-effect and random-effects meta-analyses using the proposed approach described above. To summarize the results, the average meta-analyzed estimate was calculated as well as the percentage of associated confidence intervals that cover the specified true effect.

Empirical examples

The calculation of a meta-analytic summary is illustrated here using both the standard approach and the proposed alternative approach in an empirical data example based on the data reported in a prior systematic review and meta-analysis of ischemic heart disease following exposure to low-level ionizing radiation (Little et al. 2012).

Results

Simulations: fixed-effect meta-analyses

Table 1 reports the results of simulations in which a meta-analytic summary estimate of the excess odds ratio per unit exposure was estimated using the standard fixed-effect meta-analytic approach and our proposed approach; in Table 1, the baseline odds of the outcome was set to 0.2 and the excess relative odds of the outcome was set to 1.0. In all simulation scenarios, the standard fixed-effect meta-analytic approach yielded summary effect measures that were null biased and had less than nominal confidence interval coverage. In simulations of meta-analyses of large cohorts (i.e., 4000–4500 people per study), the standard fixed-effect meta-analysis yielded a slightly biased meta-analytic summary result with 95% confidence interval coverage that was closest to nominal. In simulation scenarios involving meta-analyses of smaller cohorts, the standard fixed-effect meta-analysis exhibited greater null bias and the summary effect measure had less than nominal confidence interval coverage. As the number of studies per meta-analysis increased from 5 to 15 studies per meta-analysis, and other parameters remained unchanged, the bias in the standard fixed-effect meta-analysis remained unchanged; however, the confidence interval coverage for the standard meta-analytic summary worsened and was substantially less than the nominal level.

In all simulation scenarios, the proposed approach yielded fixed-effect meta-analytic summary results that were approximately unbiased and the confidence interval coverage for the proposed fixed-effect meta-analytic summary measure was close to the nominal 95% value in all simulation scenarios. Even in simulation scenarios involving meta-analyses

Table 1 Results for simulations, reporting the average of the standard fixed-effect and random-effects meta-analytic results, proposed fixed-effect and random-effects meta-analytic results, and confidence interval coverages

η	Simulation setup		Fixed-effect meta-analysis				Random-effects meta-analyses			
	Studies per meta-analysis	Subjects per study	Standard approach		Proposed approach		Standard approach		Proposed approach	
			Mean estimate	CI coverage %	Mean estimate	CI coverage %	Mean estimate	CI coverage %	Mean estimate	CI coverage %
1	15	4000–4500	0.95	0.78	0.99	0.95	0.96	0.86	0.99	0.95
		2000–2500	0.92	0.70	0.98	0.94	0.92	0.76	0.98	0.96
		1000–1500	0.85	0.56	0.96	0.92	0.86	0.58	0.97	0.94
	10	4000–4500	0.96	0.85	0.99	0.94	0.96	0.90	0.99	0.95
		2000–2500	0.92	0.77	0.98	0.94	0.93	0.84	0.98	0.96
		1000–1500	0.86	0.70	0.97	0.94	0.87	0.73	0.98	0.95
	5	4000–4500	0.96	0.90	0.99	0.94	0.97	0.93	1.00	0.95
		2000–2500	0.93	0.87	0.99	0.95	0.95	0.93	1.00	0.96
		1000–1500	0.89	0.83	0.98	0.96	0.90	0.88	0.99	0.97

The baseline odds of the outcome, α, was set to 0.2

of small cohorts ($N = 1000$–1500), the proposed meta-analytic summary was approximately unbiased and associated confidence interval coverage was close to the nominal 95% value. Figure 1 illustrates that the transformed metric of the estimate of association, $\hat{A}_i$, appears more normally distributed than the excess odd ratio estimates.

Simulations: random-effects meta-analyses

Table 1 also reports the results of simulations in which a meta-analytic summary estimate of the excess odds ratio per unit exposure was estimated using the standard random-effects meta-analytic approach and our proposed approach. Similar to the conclusions drawn for the fixed-effect meta-analyses, the standard random-effects meta-analytic approach yielded summary effect measures that were null biased and tended to have less than nominal confidence interval coverage. In simulations of meta-analyses of large cohorts (i.e., 4000–4500 people per study), the standard random-effects meta-analysis yielded a slightly biased meta-analytic summary result with 95% confidence interval coverage that was close to nominal. In simulation scenarios involving meta-analyses of smaller cohorts, the standard random-effects meta-analysis exhibited greater null bias and the summary effect measure had less than nominal confidence interval coverage. As the number of studies per meta-analysis increased from 5 to 15 studies per meta-analysis, and other parameters remained unchanged, the confidence interval coverage for the standard random-effects meta-analytic summary worsened and was substantially less than the nominal level.

The proposed approach yielded random-effects meta-analytic summary results that were approximately unbiased and the confidence interval coverage for the proposed
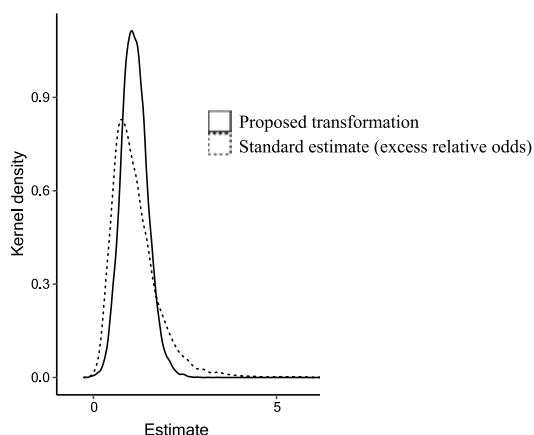

Fig. 1 Probability density functions of excess relative odds and the proposed transformed metric, *A*, across simulated studies with 500–750 subjects per study and 15 studies per meta-analysis

meta-analytic summary effect measure was slightly conservative, yet very close to the nominal 95%.

Simulation results: sensitivity analyses

Simulations also were conducted under scenarios in which baseline odds of the outcome, α, was 0.15 and 0.25; again, under those simulation scenarios, the proposed approach yielded fixed-effect and random-effects meta-analytic summary results that were approximately unbiased and the confidence interval coverage for the proposed fixed-effect and random-effects meta-analytic summary measures were close to the nominal 95% value (Appendix Table A1). In addition, simulations were conducted in which the excess relative odds of the outcome per unit exposure was 0.75 and 0.5; under those simulation scenarios it was also observed that the proposed approach yielded fixed-effect and random-effects meta-analytic summary results that were approximately unbiased and the confidence interval coverage for the proposed fixed-effect and random-effects meta-analytic summary measures were close to the nominal 95% value (Appendix Table A2). Finally, simulations were conducted in which the number of subjects per study varied from 1000 to 2500 and from 1000 to 4500; under those simulation scenarios, it was also observed that the proposed approach yielded fixed-effect and random-effects meta-analytic summary results that were approximately unbiased and the confidence interval coverage for the proposed fixed-effect and random-effects meta-analytic summary measures was close to the nominal 95% value (Appendix Table A3).

Sensitivity of the simulation results to the value *c* was assessed by re-calculating values under the condition $c' = 0.9c$ and $c' = 0.8c$; these sensitivity analyses yielded essentially very similar quantitative values (Appendix Table A4).

Empirical example

While high doses of ionizing radiation have a fairly well-established association with circulatory disease, evidence for an association at lower doses (e.g., < 0.5 Sv) remains more controversial. Little et al. reported on a meta-analysis of epidemiological findings of association between radiation exposure and circulatory disease involving moderate- or low-dose whole-body exposure to ionizing radiation (Little et al. 2012). Table 2 shows the point estimates and associated lower and upper confidence bounds for each study considered in the meta-analysis of radiation and ischemic heart disease, where the estimates represent the estimated excess relative rate per Sv whole-body dose (noting that the studies expressed radiation dose in Sv). The standard approach yields a fixed-effect estimate $\hat{\beta}_{tot}^{Fixed} = 0.10$ (95% CI 0.05, 0.15). The proposed alternative approach yields a

Table 2 Meta-analysis of epidemiological findings from eight individual studies regarding associations between ischemic heart disease and exposure to low-level radiation

i	β_i	L_i	U_i	x_i	Persons
1	0.12	0.051	0.186	5.92	12,210
2	0.41	0.05	0.78	0.50	61,017
3	0.15	−0.14	0.58	0.12	16,236
4	4.10	−2.9	13.7	0.60	22,393
5	0.26	−0.05	0.61	0.40	174,541
6	0.02	−0.1	0.15	4.00	86,611
7	-0.01	−0.59	0.69	0.50	275,312
8	0.05	−0.05	0.16	4.00	10,399

Meta-analytic summary	Standard approach		Proposed approach	
	Fixed effect	Random effects	Fixed effect	Random effects
Estimate (95% CI)	0.10 (0.05, 0.15)	0.10 (0.04, 0.15)	0.10 (0.05, 0.15)	0.10 (0.04, 0.16)

Reported coefficients (β_i) refer to estimates of excess relative risk per Sv and are indicated along with associated lower (L_i) and upper (U_i) 95% confidence intervals, and maximum dose (x_i). Illustration of data and fixed-effect, and random-effects, estimates derived in a standard approach to calculation and in the proposed alternative approach

Study citations, in numerical order: Azizova et al. (2010), Ivanov et al. (2006), Lane et al. (2010), Laurent et al. (2010), Muirhead et al. (2009), Shimizu et al. (2010), Vrijheid et al. (2007), Yamada et al. (2004), as reported in Little et al. (2012)

fixed-effect estimate $\hat{\beta}_{tot}^{Fixed} = 0.10$ (95% CI 0.05, 0.15), equivalent to the results obtained using the standard approach to calculation of a fixed-effect estimate. These results correspond to the fixed-effect meta-analytical result reported by Little et al. (2012) (0.10; 95% CI 0.05, 0.15). The similarity of the results is expected given the large sample sizes of the studies included in this meta-analysis, demonstrating that the proposed approach is not influential in the case when conditions suggest the normality assumption is tenable (Table 2).

We also report results derived under a random-effects meta-analysis. A standard approach to estimation of a random-effects estimate $\hat{\beta}_{tot}^{Random} = 0.10$ (95% CI 0.04, 0.15). The proposed alternative approach yields a random-effects estimate $\hat{\beta}_{tot}^{Random} = 0.10$ (95% CI 0.04, 0.16). Again, these results correspond closely to the random-effects meta-analytical result reported by Little et al. (0.10; 95% CI 0.04, 0.15).

A sensitivity analysis to the value c' was performed by recalculating values under the condition $c' = 0.9c$ and $c' = 0.8c$. These sensitivity analyses yielded essentially equivalent quantitative values under these conditions, the proposed fixed-effect meta-analytic summary was $\hat{\beta}_{tot}^{Fixed} = 0.10$ (95% CI 0.05, 0.15) and the proposed random-effects meta-analytic summary was $\hat{\beta}_{tot}^{Random} = 0.10$ (95% CI 0.04, 0.16).

Discussion

Non-linear regression models fitted via maximum likelihood methods are known to suffer problems when data are sparse (Greenland et al. 2016). For example, the commonly used logistic and Cox regression models are susceptible to bias in small samples (Greenland et al. 2000). These biases translate into bias of meta-analyses based on them (Greenland et al. 2016). Maximum likelihood estimates of the linear odds ratio or linear risk ratio per unit exposure are much more prone to bias in small samples than standard log-linear regression models (Prentice and Mason 1986); and, unless the study size is very large, the resultant parameter estimates may have a profile likelihood-based confidence intervals that differ substantially from Wald-type intervals (Moolgavkar and Venzon 1987).

In the current paper, we focus on the implications for meta-analytic summarization of epidemiological study results derived from maximum likelihood fittings of linear relative risk models. Using simulations, we illustrate the potential for bias and lack of appropriate confidence interval coverage in meta-analyses of linear odds ratio models that employ a standard approach to fixed-effect meta-analysis. It was observed that bias increased as the size of the studies included in meta-analyses diminished. We further noted that as the number of studies included in meta-analyses increased from 5 to 15, while other parameters remained unchanged, the bias in the meta-analytic summary remained similar but the confidence interval coverage for the meta-analytic summary decreased. This is likely because confidence intervals for meta-analytic summary estimates become tighter as the number of studies in a meta-analysis increases while the bias remains; a similar phenomenon has been reported in simulations of sparse data bias in ordinary (i.e., loglinear) logistic regression (Lin 2018).

Barlow (1985a, b), and Prentice and Mason (1986), proposed re-parameterizations of linear relative risk models that substantially reduced bias and improved approximations of confidence intervals to those predicted by the asymptotic normal distribution. This prior body of work suggested that a transformation applied for conducting a meta-analysis of published study results should lead to a distribution of estimators for *A* that tend to more closely approximate a normal density than the maximum likelihood estimators for the excess relative risk per unit *D* (Prentice and Mason 1986). A similar transformation was applied in the present study to existing estimates of excess relative risk to better approximate the normality of the meta-analyzed parameter, which is a necessary assumption underlying typical inverse-variance-weighted meta-analyses. The proposed approach is based on the known improved symmetry of the confidence interval for the transformed metric relative to the untransformed (Prentice and Mason 1986). When $\hat{A}_i$ and se $(\hat{A}_i)$ better conform to normal distributions than $\hat{\beta}_i$ and se$(\hat{\beta}_i)$ the proposed analytical approach to deriving a meta-analytic summary based on inverse-variance weighting of the transformed quantities should improve estimation as the underlying assumptions will be better approximated (Barlow 1985a, b; Prentice and Mason 1986). In principle, this should provide improved approximations of parameters with asymptotic normal distributions for meta-analysis when working with the reported results from linear relative risk models than working directly with the published values, which may not conform well to underlying distributional assumptions in small and moderately sized studies (Barlow 1985a, b).

In contrast to standard meta-analytic approaches, the proposed approach requires information on the maximum exposure range in each study included in the meta-analysis. Often this value is reported; if not, often it can be ascertained from the study authors or inferred (e.g., from the range of observed data in an exposure–response plot, or from substantive knowledge about exposure conditions).

We have illustrated the proposed approach with an empirical example. In the example, the study sizes are very large and the published likelihood-based confidence intervals are highly symmetrical. The empirical example illustrates the important point that when the assumption of normality approximately holds, the proposed approach yields essentially equivalent results to those of the standard approach. Our simulations using larger sample sizes support the finding from the empirical example that this transformation does not distort estimates when it is used in settings where normality may be a tenable assumption. The situations under which the proposed approach is likely to perform much better than the standard approach will tend to be meta-analyses that encompass many small studies, as opposed to a few large studies. The simulations illustrate that when data are sparse the approaches may yield somewhat different

meta-analytic summaries of a set of estimates of excess relative risk per unit exposure, with substantially different means and associated confidence intervals (Table 1). Our simulations demonstrate that the proposed transformation improved performance of meta-analyses in terms of bias and confidence interval coverage.

Often, in epidemiological analyses that use a linear relative risk model, the lower likelihood-based confidence bound is not determined. This poses a challenge for meta-analysis in which the published results of point estimates and confidence bounds are the basis for deriving estimates of variance that underpin inverse-variance-weighted meta-analytic summaries. It appears that a standard practice in meta-analysis of excess relative risk per unit dose estimates has been to impute a lower bound by assuming that the bounds are symmetrical on the original scale, so that given just the reported upper bound and point estimate, a lower bound is imputed. In the proposed approach, we suggest that a lower bound may be imputed by assuming symmetry on the transformed scale. If the transformation does improve normality as compared to the original scale, as illustrated in Fig. 1, this should be advantageous and improve practice for meta-analysis of radiation epidemiology results in settings where a profile likelihood-based lower confidence interval is not defined.

Conclusion

The simple approach we describe, that follows from the transformation proposed in Barlow (1985a, b) and Prentice and Mason (Prentice and Mason 1986), may offer a useful complement to standard methods that can be employed when undertaking meta-analysis of reported results based on the linear relative risk models.

Compliance with ethical standards

References

Azizova TV, Muirhead CR, Druzhinina MB, Grigoryeva ES, Vlasenko EV, Sumina MV, O'Hagan JA, Zhang W, Haylock RG, Hunter

N (2010) Cardiovascular diseases in the cohort of workers first employed at Mayak PA in 1948–1958. Radiat Res 174(2):155–168

Barlow WE (1985a) General relative risk models in stratified epidemiologic studies. J R Stat Soc Ser C (Appl Stat) 34(3):246–257

Barlow WE (1985b) General relative risk models in stratified epidemiologic studies. Appl Stat 34(3):246–257

Boice JD Jr, Blettner M, Kleinerman RA, Stovall M, Moloney WC, Engholm G, Austin DF, Bosch A, Cookfair DL, Krementz ET et al (1987) Radiation dose and leukemia risk in patients treated for cancer of the cervix. J Natl Cancer Inst 79(6):1295–1311

Cardis E, Gilbert ES, Carpenter L, Howe G, Kato I, Armstrong BK, Beral V, Cowper G, Douglas A, Fix J et al (1995) Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. Radiat Res 142(2):117–132

Cardis E, Vrijheid M, Blettner M, Gilbert E, Hakama M, Hill C, Howe G, Kaldor J, Muirhead CR, Schubauer-Berigan M, Yoshimura T, Bermann F, Cowper G, Fix J, Hacker C, Heinmiller B, Marshall M, Thierry-Chef I, Utterback D, Ahn YO, Amoros E, Ashmore P, Auvinen A, Bae JM, Solano JB, Biau A, Combalot E, Deboodt P, Diez Sacristan A, Eklof M, Engels H, Engholm G, Gulis G, Habib R, Holan K, Hyvonen H, Kerekes A, Kurtinaitis J, Malker H, Martuzzi M, Mastauskas A, Monnet A, Moser M, Pearce MS, Richardson DB, Rodriguez-Artalejo F, Rogel A, Tardy H, Telle-Lamberton M, Turai I, Usel M, Veress K (2005) Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries. BMJ 331(7508):77

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London, pp 1–511

Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, Deo H, Falk R, Forastiere F, Hakama M, Heid I, Kreienbrock L, Kreuzer M, Lagarde F, Mèakelèainen I, Muirhead C, Oberaigner W, Pershagen G, Ruano-Ravina A, Ruosteenoja E, Rosario AS, Tirmarche M, Tomâasek L, Whitley E, Wichmann HE, Doll R (2005) Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ 330(7485):223

DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7(3):177–188

Gilbert ES, Cragle DL, Wiggs LD (1993) Updated analyses of combined mortality data for workers at the Hanford Site, Oak Ridge National Laboratory, and Rocky Flats Weapons Plant. Radiat Res 136(3):408–421

Greenland S, Schwartzbaum JA, Finkle WD (2000) Problems due to small samples and sparse data in conditional logistic regression analysis. Am J Epidemiol 151(5):531–539

Greenland S, Mansournia MA, Altman DG (2016) Sparse data bias: a problem hiding in plain sight. BMJ 352:i1981

Hein MJ, Stayner LT, Lehman E, Dement JM (2007) Follow-up study of chrysotile textile workers: cohort mortality and exposure-response. Occup Environ Med 64(9):616–625

Higgins JPT, Cochrane Collaboration (2020) Cochrane handbook for systematic reviews of interventions. Wiley-Blackwell, Hoboken

Ivanov VK, Maksioutov MA, Chekin SY, Petrov AV, Biryukov AP, Kruglova ZG, Matyash VA, Tsyb AF, Manton KG, Kravchenko JS (2006) The risk of radiation-induced cerebrovascular disease in Chernobyl emergency workers. Health Phys 90(3):199–207

Jackson D, White IR (2018) When should meta-analysis avoid making hidden normality assumptions? Biom J 60(6):1040–1058

Lane RS, Frost SE, Howe GR, Zablotska LB (2010) Mortality (1950–1999) and cancer incidence (1969–1999) in the cohort of Eldorado uranium workers. Radiat Res 174(6):773–785

Laurent O, Metz-Flamant C, Rogel A, Hubert D, Riedel A, Garcier Y, Laurier D (2010) Relationship between occupational exposure to ionizing radiation and mortality at the French electricity company, period 1961–2003. Int Arch Occup Environ Health 83(8):935–944

Lin L (2018) Bias caused by sampling error in meta-analysis with small sample sizes. PLoS One 13(9):e0204056

Little MP (2010) Cancer models, genomic instability and somatic cellular Darwinian evolution. Biol Direct 5:19

Little MP, Wakeford R, Tawn EJ, Bouffler SD, Berrington de Gonzalez A (2009) Risks associated with low doses and low dose rates of ionizing radiation: why linearity may be (almost) the best we can do. Radiology 251(1):6–12

Little MP, Azizova TV, Bazyka D, Bouffler SD, Cardis E, Chekin S, Chumak VV, Cucinotta FA, de Vathaire F, Hall P, Harrison JD, Hildebrandt G, Ivanov V, Kashcheev VV, Klymenko SV, Kreuzer M, Laurent O, Ozasa K, Schneider T, Tapio S, Taylor AM, Tzoulaki I, Vandoolaeghe WL, Wakeford R, Zablotska LB, Zhang W, Lipshultz SE (2012) Systematic review and meta-analysis of circulatory disease from exposure to low-level ionizing radiation and estimates of potential population mortality risks. Environ Health Perspect 120(11):1503–1511

Lubin JH, Boice JD Jr, Edling C, Hornung RW, Howe G, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM et al (1995) Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. Health Phys 69(4):494–500

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, Boca Raton

Meeker WQ, Escobar LA (1995) Teaching about approximate confidence regions based on maximum likelihood estimation. Am Stat 49(1):48–53

Metz-Flamant C, Laurent O, Samson E, Caer-Lorho S, Acker A, Hubert D, Richardson DB, Laurier D (2013) Mortality associated with chronic external radiation exposure in the French combined cohort of nuclear workers. Occup Environ Med 70(9):630–638

Moolgavkar SH, Venzon DJ (1987) General relative risk regression models for epidemiologic studies. Am J Epidemiol 126(5):949–961

Muirhead CR, O'Hagan JA, Haylock RG, Phillipson MA, Willcock T, Berridge GL, Zhang W (2009) Mortality and cancer incidence following occupational radiation exposure: third analysis of the National Registry for Radiation Workers. Br J Cancer 100(1):206–212

National Research Council (U.S.). Committee on the Biological Effects of Ionizing Radiations (1990) Health effects of exposure to low levels of ionizing radiation: BEIR V. National Academy Press, Washington, D.C.

National Research Council (U.S.), Committee on the Biological Effects of Ionizing Radiations, United States, Environmental Protection Agency, U.S. Nuclear Regulatory Commission (1988) Health risks of radon and other internally deposited alpha-emitters. National Academy Press, Washington, D C

Pawel D, Preston D, Pierce D, Cologne J (2008) Improved estimates of cancer site-specific risks for A-bomb survivors. Radiat Res 169(1):87–98

Plackett RL (1949) A historical note on the method of least squares. Biometrika 36(3/4):458–460

Prentice RL, Mason MW (1986) On the application of linear relative risk regression models. Biometrics 42(1):109–120

Preston DL, Shimizu Y, Pierce DA, Suyama A, Mabuchi K (2003) Studies of mortality of atomic bomb survivors. Report 13: solid cancer and noncancer disease mortality: 1950–1997. Radiat Res 160(4):381–407

Preston DL, Ron E, Tokuoka S, Funamoto S, Nishi N, Soda M, Mabuchi K, Kodama K (2007) Solid cancer incidence in atomic bomb survivors: 1958-1998. Radiat Res 168(1):1–64

Rinsky RA, Hornung RW, Silver SR, Tseng CY (2002) Benzene exposure and hematopoietic mortality: a long-term epidemiologic risk assessment. Am J Ind Med 42(6):474–480

Shimizu Y, Kodama K, Nishi N, Kasagi F, Suyama A, Soda M, Grant EJ, Sugiyama H, Sakata R, Moriwaki H, Hayashi M, Konda M,

Shore RE (2010) Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data, 1950–2003. BMJ 340:b5349

Sprott DA (1974) Normal likelihoods and their relation to large sample theory of estimation. Biometrika 60(3):457–465

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000) Methods for meta-analysis in medical research. Wiley, Chichester

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (1993) Sources and effects of ionizing radiation. UNSCEAR 1993 report to the General Assembly, with scientific annexes. United Nations, New York. E.94.IX.2:1–922

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (2018) Sources, effects and risks of ionizing radiation. UNSCEAR 2017 report to the General Assembly. Scientific annexes A and B. United Nations, New York. E.18.IX.1:1–194

Vrijheid M, Cardis E, Ashmore P, Auvinen A, Bae JM, Engels H, Gilbert E, Gulis G, Habib R, Howe G, Kurtinaitis J, Malker H, Muirhead C, Richardson D, Rodriguez-Artalejo F, Rogel A, Schubauer-Berigan M, Tardy H, Telle-Lamberton M, Usel M, Veress K (2007) Mortality from diseases other than cancer following low doses of ionizing radiation: results from the 15-Country Study of nuclear industry workers. Int J Epidemiol 36(5):1126–1135

Yamada M, Wong FL, Fujiwara S, Akahoshi M, Suzuki G (2004) Non-cancer disease incidence in atomic bomb survivors, 1958-1998. Radiat Res 161(6):622–632

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.