Improving likelihood-based inference in control rate regression

Annamaria Guolo

University of Padova, Via Cesare Battisti 241/243, I-35121, Padova, Italy. Phone: +39 049 8274150. Fax: +39 049 8274170.

annamaria.guolo@unipd.it

Abstract Control rate regression is a diffuse approach to account for the heterogeneity among studies in meta-analysis by including information about the outcome risk of patients in the control condition. Correcting for the presence of measurement error affecting risk information in the treated and in the control group has been recognized as a necessary step to derive reliable inferential conclusions. Within this framework, the paper considers the problem of small sample size as an additional source of misleading inference about the slope of the control rate regression. Likelihood procedures relying on first-order approximations are shown to be substantially inaccurate, especially when dealing with increasing heterogeneity and correlated measurement errors. We suggest to address the problem by relying on higher-order asymptotics. In particular, we derive Skovgaard's statistic as an instrument to improve the accuracy of the approximation of the signed profile log-likelihood ratio statistic to the standard normal distribution. The proposal is shown to provide much more accurate results than standard likelihood solutions, with no appreciable computational effort. The advantages of Skovgaard's statistic in control rate regression are shown in a series of simulation experiments and illustrated in a real data example. `R` code for applying first- and second-order statistic for inference on the slope on the control rate regression is provided.

Keywords: control rate; higher-order asymptotics; likelihood inference; measurement error; meta-analysis

1 Introduction

Control rate regression is a diffuse instrument for taking into account the between-study heterogeneity in meta-analysis comparing a treated group and a control group ([1], [2], [3], [4]). To this aim, a measure of the outcome risk of patients in the control condition is considered, so that emerging differences among studies are due to treatment effects only. The control rate, defined as the proportion of patients with the event of interest in the control group, represents a surrogate for the true risk of patients in the control condition and thus is a measure affected by error. Correcting for the presence of measurement error is a necessary step for inference to be reliable ([5], [6]). The most famous effect of ignoring the presence of measurement error is the downwards bias of the estimate of the slope in a linear regression model with additive homoschedastic errors on the covariate. Likelihood-based inference for measurement error correction in control rate regression has received substantial attention in the literature given its limit properties, see, for example, Arends et al. [2], Ghidey et al. [7], and Guolo [8]. We show that, despite the advantages in terms of properties of the maximum likelihood estimator, likelihood inference relying on first-order approximations can be inaccurate when the sample size, i.e., the number of studies included in the meta-analysis, is small. In particular, the asymptotic χ^2 distribution for the likelihood ratio statistic is shown to be flawed, seriously affecting inferential conclusions. In this paper we suggest to overcome the problem and refine first-order likelihood inference through Skovgaard's second-order statistic [9]. The present work takes advantage of previous results about higher-order asymptotics illustrated in Guolo [10] within the classical meta-analysis framework and constitutes a step forward for developing Skovgaard's second-order statistic in the multivariate meta-analysis accounting for measurement errors. The accuracy of the results is obtained with no substantial computational effort, as Skovgaard's statistic can be derived in closed-form with components having a complexity comparable to that of evaluating the expected information matrix. Advantages over first-order results are highlighted in a series of simulation studies. The application of the method is illustrated via a real data example about the efficacy of a drug treatment against cardiovascular mortality in middle-aged patients with mild to moderate hypertension. The `R` [11] code for implementing Skovgaard's second-order statistic is made available as supplementary material and illustrated in the Appendix.

2 Likelihood inference

We consider a meta-analysis of n independent studies about the effectiveness of a treatment. Let η_i denote the risk measure in the treated group, or the treatment effect, and let ξ_i denote the outcome underlying risk measure in the control group, $i = 1, \ldots, n$. Control rate regression is typically defined as (e.g., [1], [2])

$$\eta_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i, \ \varepsilon_i \sim N(0, \tau^2), \tag{1}$$

with parameter τ^2 accounting for the heterogeneity with respect to the treatment measure in the population with the same underlying risk. The inferential interest is usually in β_1, with $\beta_1 = 0$ used to verify the constance of the treatment effect and its independence with respect to ξ_i. An alternative specification of the model considers the relationship between the treatment effect $\eta_i - \xi_i$ and ξ_i (e.g., [12]), with $(\beta_0, \beta_1)^\top = (0, 1)^\top$ representing a claim of no relationship between the treatment effect and the risk in the control condition, on average.

The simplest approach for analysis suggested by Brand and Kragt [13] is a weighted least squares regression, with weights given by the inverse of the variance of the treatment effect. This approach does not consider that the summary information from each study represents a surrogate for the true unobserved risk measure and consequently is prone to measurement error. A huge literature warns against misleading inferential conclusions due to ignoring measurement errors, see Carroll et al. [5] and Buonaccorsi [6]. In case of additive homoschedastic errors, the most evident effect is the attenuation of the least squares estimate of β_1, which is biased towards zero, a situation known as *regression dilution bias*. Let $\hat{\eta}_i$ and $\hat{\xi}_i$ denote the observed error-prone versions of η_i and ξ_i available from study i. A commonly adopted measurement error structure ([1], [2], [14]) relates $(\hat{\eta}_i, \hat{\xi}_i)^\top$ to $(\eta_i, \xi_i)^\top$ through the bivariate Normal distribution

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, \Gamma_i \right), \tag{2}$$

where the within-study variance/covariance matrix Γ_i is specified using single study information. Likelihood-based inference requires the specification of the distribution for the underlying risk ξ_i. Typically, a normal specification is adopted for computational convenience, $\xi_i \sim N(\mu, \sigma^2)$ (e.g., [3]). Given the above distributional assumptions, the likelihood function for the whole parameter vector $\theta = (\beta_0, \beta_1, \tau^2, \mu, \sigma^2)^\top$ is easily obtained with a closed-form considering that, marginally,

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \beta_0 + \beta_1 \mu \\ \mu \end{pmatrix}, \Gamma_i + \begin{pmatrix} \tau^2 + \beta_1^2 \sigma^2 & \beta_1 \sigma^2 \\ \beta_1 \sigma^2 & \sigma^2 \end{pmatrix} \right). \tag{3}$$

The computational convenience of the closed-form for the likelihood function is a practical justification for the choice of the normal specification for the measurement error model and for the underlying risk distribution. Alternative specifications for both the models, however, have been examined in the literature. Specification (2) is often an approximation of the exact measurement error structure, which can be defined case by case [2], although at the price of computational complications. Alternatives to the normal model for the underlying risk examined in the literature include flexible solutions based on mixture of normals [2], semiparametric specification [7] and the skew-normal distribution [8].

2.1 First-order inference

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ, based on (3) and consider the inferential interest directed to the slope parameter β_1. Denote, for simplicity, $\lambda = (\beta_0, \mu, \tau^2, \sigma^2)^\top$ the vector of the remaining nuisance parameters and let $\tilde{\theta} = (\beta_1, \lambda_{\beta_1})^\top$ be the constrained maximum likelihood estimate of θ for fixed β_1. Inference on β_1 can rely on the profile log-likelihood function $\ell_P(\beta_1) = \ell(\beta_1; \lambda_{\beta_1})$. Hypothesis testing and construction of confidence intervals can be based on the signed (square root of the) profile log-likelihood ratio statistic

$$r_P(\beta_1) = \mathrm{sign}\left(\hat{\beta}_1 - \beta_1\right)\left\{\ell_{\mathrm{P}}(\hat{\beta}_1) - \ell_{\mathrm{P}}(\beta_1)\right\}^{1/2}, \tag{4}$$

which, under mild regularity conditions, has an approximate standard normal distribution up to first-order error, see Section 4.4 in Severini [15]. The use of $r_P(\beta_1)$ is preferable to the commonly adopted Wald-type statistic as the inferential procedures are invariant to reparameterization and confidence intervals based on $r_P(\beta_1)$ are not forced to be symmetric. Nevertheless, first-order asymptotic results are known to provide misleading conclusions when the sample size is small, see, for example, Brazzale et al. [16]. The problem in the meta-analysis context has been previously investigated in Guolo [10], Guolo and Varin [17], Bellio and Guolo [18].

2.2 Skovgaard's statistic

Several modifications of r_P have been proposed in the literature, which are aimed at reducing the order of the error in approximating the standard normal ([15], [19]). In this paper, we consider the refinement given by Skovgaard's statistic [9] which has an approximate standard normal distribution up to second-order error. The choice is motivated by the fact that the statistic is well-defined for a wide class of regular problems and is computationally feasible. In fact, the evaluation of Skovgaard's statistic components requires an effort similar to that of computing the expected information matrix. The invariance with respect to interest-respecting reparameterizations is maintained. Guolo [10] investigated the use of Skovgaard's statistic in meta-analysis and meta-regression, under the classical random-effects formulation [20]. This paper takes advantage of the starting results in Guolo [10] to extend the usage of Skovgaard's statistic to the multivariate meta-analysis represented by control rate regression. Measurement errors on $\hat{\eta}_i$ and $\hat{\xi}_i$ are taken into account but they do not substantially affect the feasibility of the approach.

Skovgaard's statistic is defined as a modification of r_P

$$\bar{r}_P(\beta_1) = r_P(\beta_1) + \frac{1}{r_P(\beta_1)} \log \frac{u(\beta_1)}{r_P(\beta_1)}, \tag{5}$$

where $u(\beta_1)$ represents the correction term

$$u(\beta_1) = [S^{-1}q]_{\beta_1} |\hat{j}|^{1/2} |\hat{i}|^{-1} |S| |\tilde{j}_{\lambda\lambda}|^{-1/2}.$$

In the above expression, symbol $|\cdot|$ denotes the determinant, $\hat{i}$ and $\hat{j}$ are the expected information matrix and the observed information matrix, respectively, both evaluated at the maximum likelihood estimate $\hat{\theta}$ and $\tilde{j}_{\lambda\lambda}$ represent the sub-block of j corresponding to the parameter vector λ evaluated at the constrained maximum likelihood estimate $\tilde{\theta}$. Similarly, $[S^{-1}q]_{\beta_1}$ is the component of the vector $S^{-1}q$ corresponding to β_1, with S and q covariances of likelihood terms, namely,

$$S = \text{cov}_{\theta_1} \left\{ \frac{\partial \ell(\theta_1)}{\partial \theta}, \frac{\partial \ell(\theta_2)}{\partial \theta} \right\} \Bigg|_{\theta_1 = \hat{\theta}, \theta_2 = \tilde{\theta}}$$

and

$$q = \text{cov}_{\theta_1} \left\{ \frac{\partial \ell(\theta_1)}{\partial \theta}, \ell(\theta_1) - \ell(\theta_2) \right\} \Bigg|_{\theta_1 = \hat{\theta}, \theta_2 = \tilde{\theta}}.$$

Within the control rate regression, the derivation of the Skovgaard's statistic components gives rise to the following expression. Let f_i denote the mean vector of $(\hat{\eta}_i, \hat{\xi}_i)^\top$ and V_i denote the associated variance/covariance matrix in (3). A subfix indicates the derivation with respect to each component of θ. A "hat" and a "tilde" indicates the evaluation of a vector or a matrix with respect to $\hat{\theta}$ and $\tilde{\theta}$, respectively. Accordingly, S is a 5×5 matrix with components

$$S_{\beta_j,\beta_k} = \sum_{i=1}^{n} \left\{ \frac{1}{2} \text{trace} \left(\hat{V}_{\beta_j}^{-1} \hat{V}_i \tilde{V}_{\beta_k}^{-1} \hat{V}_i \right) + \hat{f}_{i,\beta_j}^\top \tilde{V}_{\beta_k}^{-1} \left(\tilde{f}_i - \hat{f}_i \right) + \hat{f}_{\beta_k} \tilde{V}_i^{-1} \tilde{f}_{\beta_k} \right\}, \; j,k = 0,1,$$

$$S_{\beta_j,\mu} = \sum_{i=1}^{n} \hat{f}_{i,\beta_j}^\top \tilde{V}_i^{-1} \tilde{f}_{i,\mu}, \; j = 0,1,$$

$$S_{\mu\mu} = \sum_{i=1}^{n} \hat{f}_{i,\mu}^\top \tilde{V}_i^{-1} \tilde{f}_{i,\mu},$$

$$S_{\beta_j,\psi_k} = \sum_{i=1}^{n} \left\{ \frac{1}{2} \text{trace} \left(\hat{V}_{i,\beta_j}^{-1} \hat{V}_i \tilde{V}_{i,\psi_k}^{-1} \hat{V}_i \right) + \hat{f}_{i,\beta_j}^\top \tilde{V}_{i,\psi_k}^{-1} \left(\tilde{f}_i - \hat{f}_i \right) \right\}, \; j = 0,1, \; \psi_k \in \{\tau^2, \sigma^2\},$$

$$S_{\mu,\psi_k} = \sum_{i=1}^{n} \left\{ \frac{1}{2} \text{trace} \left(\hat{V}_{i,\mu}^{-1} \hat{V}_i \tilde{V}_{i,\psi_k}^{-1} \hat{V}_i \right) + \hat{f}_{i,\mu}^\top \tilde{V}_{i,\psi_k}^{-1} \left(\tilde{f}_i - \hat{f}_i \right) \right\}, \; \psi_k \in \{\tau^2, \sigma^2\}, \; \psi_k \in \{\tau^2, \sigma^2\},$$

$$S_{\psi_j,\psi_k} = \frac{1}{2} \sum_{i=1}^{n} \text{trace} \left(\hat{V}_{i,\psi_j}^{-1} \hat{V}_i \tilde{V}_{i,\psi_k}^{-1} \hat{V}_i \right), \; \psi_j, \psi_k \in \{\tau^2, \sigma^2\},$$

$$S_{\mu,\beta_j} = \sum_{i=1}^{n} \left(\hat{f}_{i,\mu}^{\top} \tilde{V}_i^{-1} \tilde{f}_{i,\beta_j} + \hat{f}_{i,\mu}^{\top} \tilde{V}_{i,\beta_j}^{-1} \tilde{f}_i - \hat{f}_i^{\top} \tilde{V}_{i,\beta_j}^{-1} \hat{f}_{i,\mu} \right), \; j = 0, 1$$

$$S_{\psi_j,\beta_k} = \frac{1}{2} \sum_{i=1}^{n} \text{trace} \left(\hat{V}_{i,\psi_j}^{-1} \hat{V}_i \tilde{V}_{i,\beta_k}^{-1} \hat{V}_i \right), \; \psi_j \in \{\tau^2, \sigma^2\}, k = 0, 1$$

$$S_{\psi_j,\mu} = 0, \; \psi_j \in \{\tau^2, \sigma^2\}.$$

Similarly, q is vector of 5 components

$$q_{\beta_j} = \sum_{i=1}^{n} \left[\frac{1}{2} \text{trace} \left\{ \hat{V}_{i,\beta_j}^{-1} \hat{V}_i \left(\hat{V}_i^{-1} - \tilde{V}_i^{-1} \right) \hat{V}_i \right\} + \hat{f}_{i,\beta_j}^{\top} \tilde{V}_i^{-1} \left(\hat{f}_i - \tilde{f}_i \right) \right], \; j = 0, 1,$$

$$q_{\mu} = \sum_{i=1}^{n} \left[\frac{1}{2} \text{trace} \left\{ \hat{V}_{i,\mu}^{-1} \hat{V}_i \left(\hat{V}_i^{-1} - \tilde{V}_i^{-1} \right) \hat{V}_i \right\} + \hat{f}_{i,\mu}^{\top} \tilde{V}_i^{-1} (\hat{f}_i - \tilde{f}_i) \right]$$

and

$$q_{\psi_j} = \frac{1}{2} \sum_{i=1}^{n} \left\{ \text{trace} \left(\hat{V}_{\psi_j}^{-1} \hat{V}_i \right) - \text{trace} \left(\hat{V}_{\psi_j}^{-1} \hat{V}_i \tilde{V}_i^{-1} \hat{V}_i \right) \right\}, \; \psi_j \in \{\tau^2, \sigma^2\}.$$

The covariances of the likelihood terms S and q that give rise to the improvement of $r_P(\beta_1)$ include the measurement error correction, as the error components are taken into account both in the mean f_i and in the variance/covariance matrix V_i, see expression (3). Unfortunately, such a structure does not allow to write Skovgaard's components by separating higher-order terms and measurement error correction terms. Details about how to compute the components of $\overline{r}_P(\beta_1)$ are provided in the Appendix.

3 Simulation studies

Several simulation studies have been conducted to investigate the performance of Skovgaard's statistic $\overline{r}_P$ with respect to the signed profile log-likelihood ratio statistic r_P in terms of accuracy of inferential results about β_1. Both the approaches are compared to the usual weighted least squares regression. Data have been simulated by first generating the number of events within each study included in the meta-analysis and then using them to produce the outcome measure of interest in the treated and in the control group. Consider each study providing the number of events y_i and the total number of person-years n_i, $i = 1, \ldots, n$, in the treatment group and the corresponding quantities x_i and m_i in the control group. For fixed n, the number of events y_i and x_i are simulated from the distributions $Y_i \sim \text{Poisson}(n_i e^{\eta_i})$ and $X_i \sim \text{Poisson}(m_i e^{\xi_i})$ [2]. Quantities n_i and m_i are generated from a Uniform variable on $[100, 5000]$. Two main scenarios are considered, by distinguishing two specifications of the outcome of interest η_i. A first scenario considers η_i and

ξ_i being the log event rate in the treatment and in the control group, respectively, with the observed versions evaluated as $\hat{\eta}_i = \log(y_i/n_i)$ and $\hat{\xi}_i = \log(x_i/m_i)$. The variance/covariance matrix is

$$\Gamma_i = \begin{bmatrix} y_i^{-1} & 0 \\ 0 & x_i^{-1} \end{bmatrix}. \tag{6}$$

In a second scenario, we consider η_i and ξ_i being the log rate ratio and the log event rate, respectively, with the observed versions evaluated as $\hat{\eta}_i = \log(y_i/n_i) - \log(x_i/m_i)$ and $\hat{\xi}_i = \log(x_i/m_i)$. The variance/covariance matrix is

$$\Gamma_i = \begin{bmatrix} y_i^{-1} + x_i^{-1} & -x_i^{-1} \\ -x_i^{-1} & x_i^{-1} \end{bmatrix}.$$

For each simulation scheme, we set $(\beta_0, \beta_1)^\top = (0, 1)^\top$ and τ assuming value in $\{0.1, 0.3, 0.5, 0.7, 0.9, 1.2, 1.5, 2\}$. Increasing values for the number of studies included in the meta-analysis are considered, $n \in \{5, 10, 20\}$. The simulation experiment has been repeated 1,000 times for each scenario and for each combination of between-study heterogeneity τ^2 and sample size n. The methods are compared in terms of empirical coverage probabilities of confidence intervals for β_1 at nominal level 0.95. When using the weighted least squares regression, the Wald-type confidence interval is considered. Likelihood maximisation, based on the Nelder and Mead algorithm [21], employs the weighted least squares estimates as starting values.

Simulation results are reported in Figure 1 for scenario one and in Figure 22 for scenario two.

Skovgaard's statistic provides empirical coverages of confidence intervals very close to the nominal level, for every examined scenario, independently of the sample size n and the amount of between-study heterogeneity τ^2. The improvement provided by the method over alternative approaches is pronounced and more evident in case of small sample size, as well as large values of τ^2. See, for example, the results for $n = 5$ and for $\tau = 2.0$ under both the scenarios. Relying on first-order likelihood inference turns out in confidence intervals with empirical coverage probabilities substantially lower than the nominal level when the sample size is small. Differences with respect to Skovgaard's statistic are more evident in presence of correlation between the measurement errors, see Figure 2. Globally, differences reduce as the sample size increases. Unsurprisingly, the weighted least squares regression shows a pronounced unsatisfactory behaviour, as a consequence of not accounting for measurement errors. The empirical coverages probabilities notably underestimate the nominal 95% level, more seriously as the amount of between-study heterogeneity increases and when η_i is the log event rate ratio (scenario two).

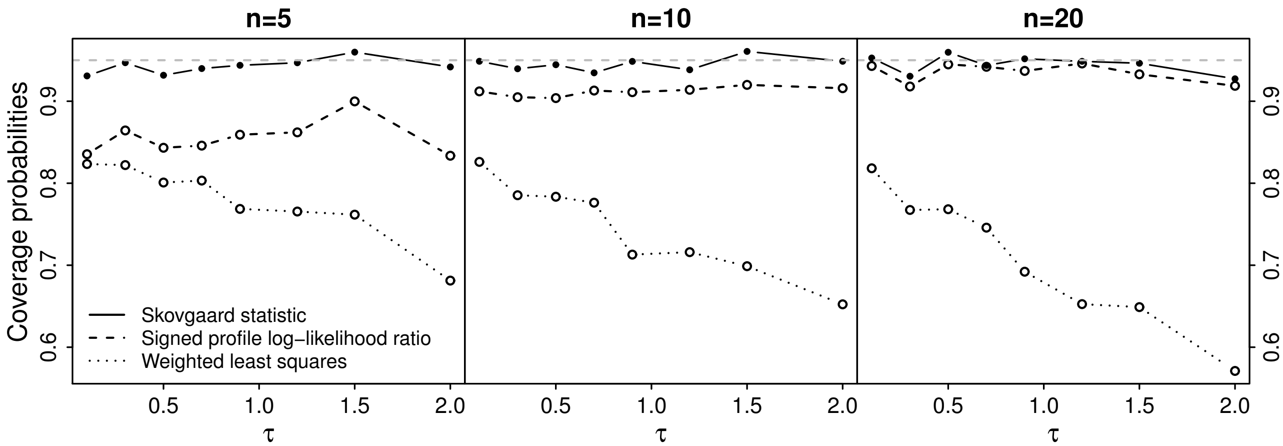

Figure 1: Empirical coverage probabilities of the nominally 95% confidence interval for β_1 when η is the log event rate, under increasing sample size n and square root of between-study variance τ. The plotted curves correspond to Skovgaard's statistic (solid), the signed profile log-likelihood ratio statistic (dashed), the weighted least squares approach (dotted). The dashed, grey horizontal line is the nominal level.

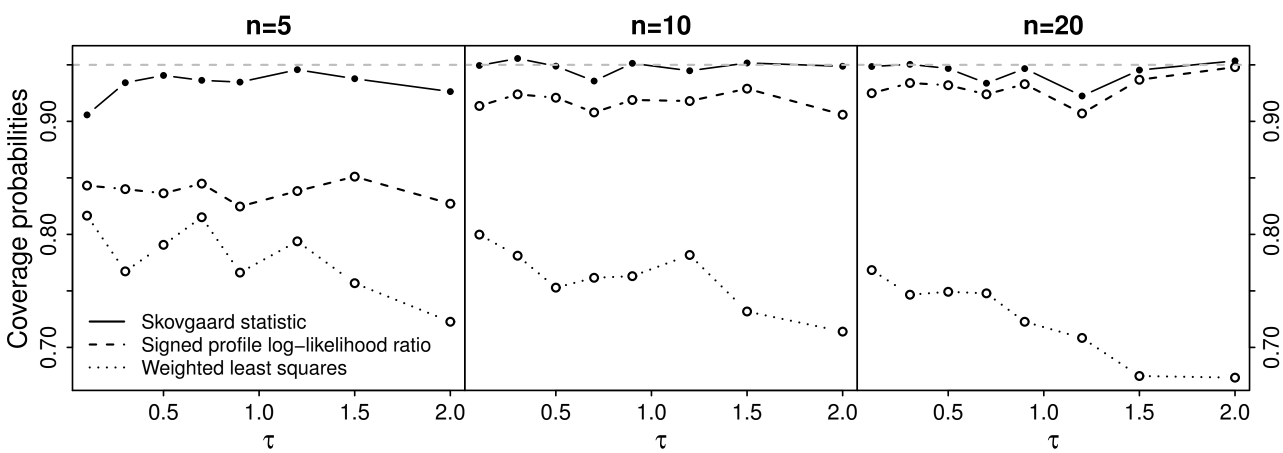

Figure 2: Empirical coverage probabilities of the nominally 95% confidence interval for β_1 when η is the log event rate ratio, under increasing sample size n and square root of between-study variance τ. The plotted curves correspond to Skovgaard's statistic (solid), the signed profile log-likelihood ratio statistic (dashed), the weighted least squares approach (dotted). The dashed, grey horizontal line is the nominal level.

4 Example

Hoes et al. [22] consider a meta-analysis of 12 studies about the efficacy of a drug treatment compared to placebo or no treatment to prevent death for cardiovascular reasons in middle-aged patients with mild to moderate hypertension. The available information is in terms of the number of events and the total number of person-years per group, as reported in Table 1. See also Arends et al. [2] who analyze the data through a Bayesian approach.

Study	Treatment group		Control group	
	Deaths	Person-years	Deaths	Person-years
1	10	595.2	21	640.2
2	2	762.0	0	756.5
3	54	5635.0	70	5600.0
4	47	5135.0	63	4960.0
5	53	3760.0	62	4210.0
6	10	2233.0	9	2084.5
7	25	7056.1	35	6824.0
8	47	8099.0	31	8267.0
9	43	5810.0	39	5922.0
10	25	5397.0	45	5173.0
11	157	22162.7	182	22172.5
12	92	20885.0	72	20645.0

Table 1: Number of deaths and total number of person-years in the treatment and control group of mild to moderate hypertension middle-aged patients in the meta-analysis of Hoes et al. [22].

Let η_i and ξ_i denote the log event rate for the i-th treatment group and control group, respectively. The observed error-prone $\hat{\eta}_i$ and $\hat{\xi}_i$ are evaluated as the logarithm of the number of deaths over the total number of person-years in the treatment and in the control group, respectively. The associated variance/covariance matrix Γ_i follows expression (6) in the first simulation study. Arends et al. [2] focus the inferential interest on β_1 in model (1) and test for $\beta_1 = 1$ as the slope of the no-effect line. When considering the same hypothesis test, then the use of the signed profile log-likelihood ratio statistic results in an effect of the drug treatment, as $r_P(\beta_1) = -2.34$, with an associated p-value equal to 0.02. Skovgaard's statistic accounting for the small number of studies included in the meta-analysis concludes for no effect of the treatment, as $\overline{r}_P(\beta_1) = -1.27$, with an associated p-value equal to 0.20.

5 Concluding remarks

This paper considered likelihood inference in control rate regression accounting for the presence of measurement error affecting the outcome risk measure of both the treatment and the control group. Attention has been paid to situations with a small number of studies, where first-order results based on the log-likelihood ratio statistic can be substantially inaccurate. In order to avoid misleading inferential conclusions, we suggested to base inference on Skovgaard's statistic, which improves to the second-order the accuracy of approximating the standard normal distribution. The simulation experiments show that the empirical coverage probabilities of confidence intervals for β_1 based on Skovgaard's statistic are closer to the nominal level than those derived from the log-likelihood ratio statistic. The improvements are more evident when the number of studies included in the meta-analysis is small, e.g., $n = 5$, and with increasing between-study heterogeneity. Correlated measurement error structure also represents a framework where advantages of Skovgaard's statistic in place of the likelihood ratio statistic emerge. The gain in accuracy is reached with no appreciable computational effort, as the evaluation of Skovgaard's statistic components has a complexity comparable to that of computing the expected information matrix.

The simulation study and the data analysis have been implemented using the `R` programming language [11]. The `R` code for computing Skovgaard's statistic is provided as supplementary material. The Appendix includes an illustration about how to use the software in order to implement Skovgaard's statistic in control rate regression.

Likelihood inference performed in this paper, using either first-order or higher-order solutions, considers the approximate normal distribution for the measurement error, which implies the likelihood function being in closed-form. In case the exact measurement error structure is considered and the likelihood function is not in closed-form, then Skovgaard's statistic is still evaluable. In this case, however, closed-form the order of the approximation to the standard normal is not known, as a consequence of the numerical integration. Nevertheless, experimental studies in Guolo et al. [23] shows that a good performance of Skovgaard's statistic with respect to the first-order solution is maintained in random-effects models.

Supporting information

The R code for applying Skovgaard's statistic in control rate regression is provided as supplementary material.

Acknowledgments

This work was supported by a grant from the University of Padova (Progetti di Ricerca di Ateneo 2015, CPDA153257).

References

[1] Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine* 1998; **17**: 1923–1942. DOI: 10.1002/(SICI)1097-0258(19980915)17:17<1923::AID-SIM874>3.0.CO;2-6

[2] Arends LR, Hoes AW, Lubsen J, Grobbee DE, Stijnen T. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* 2000; **19**: 3497–3518. DOI: 10.1002/1097-0258(20001230)19:24<3497::AID-SIM830>3.0.CO;2-H

[3] van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta- analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**: 589–624. DOI: 10.1002/sim.1040

[4] Chaimani A. Accounting for baseline differences in meta-analysis. *Evidence-Based Mental Health* 2015; **18**: 23–26. DOI:10.1136/eb-2014-102035

[5] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective.* Chapman & Hall, CRC Press: Boca Raton, 2006. ISBN: 1584886331

[6] Buonaccorsi JP. *Measurement Error: Models, Methods and Applications.* Chapman & Hall, CRC Press: Boca Raton, 2010. ISBN: 9781420066562

[7] Ghidey W, Lesaffre E, Stijnen T. Semi-parametric modelling of the distribution of the baseline risk in meta-analysis. *Statistics in Medicine* 2007, **26**: 5434–5444. DOI: 10.1177/0962280211412244

[8] Guolo A. Flexibly modeling the baseline risk in meta-analysis. *Statistics in Medicine* 2013; **32**: 40–50. DOI: 10.1002/sim.5506

[9] Skovgaard IM. An explicit large-deviation approximation to one-parameter tests. *Bernoulli* 1996; **2**: 145–165. DOI: 10.2307/3318548

[10] Guolo A. Higher-order likelihood inference in meta-analysis and meta-regression. *Statistics in Medicine* 2012; **31**: 313–327. DOI: 10.1002/sim.4451

[11] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. ISBN 3-900051-07-0, URL http://www.R-project.org.

[12] Ghidey W, Stijnen T, van Houwelingen HC. Modelling the effect of baseline risk in meta-analysis: A review from the perspective of errors-in-variables regression. *Statistical Methods in Medical Research* 2013; **22**: 307–323. DOI: 10.1177/0962280211412244

[13] Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* 1992; **11**: 2077–2082.

[14] Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology* 2004; **57**: 683–697. DOI: http://dx.doi.org/10.1016/j.jclinepi.2003.12.001

[15] Severini TA. *Likelihood Methods in Statistics*. Oxford University Press: Oxford, 2000. ISBN: 9780198506508

[16] Brazzale AR, Davison AC, Reid N. *Applied Asymptotics: Case Studies in Small Sample Statistics*. Cambridge University Press: Cambridge, 2007. ISBN: 9780521847032

[17] Guolo A, Varin C. Random-effects meta-analysis: The number of studies matters. *Statistical Methods in Medical Research*, to appear. DOI: 10.1177/0962280215583568

[18] Bellio R, Guolo A. Integrated likelihood inference in small sample meta-analysis for continuous outcomes. *Scandinavian Journal of Statistics* 2016; **43**: 191–201. DOI: 10.1111/sjos.12172

[19] Reid N. Asymptotics and the theory of inference. *The Annals of Statistics* 2003; **31**: 1695–1731.

[20] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**: 177–188.

[21] Nelder JA, Mead R. A simplex algorithm for function minimization. *Scandinavian Journal of Statistics* 1965; **7**: 308–313.

[22] Hoes AW, Grobbee DE, Lubsen J. Does drug treatment improve survival? Reconciling the trials in mild-to-moderate hypertension. *Journal of Hypertension* 1995; **13**: 805–811.

[23] Guolo A, Brazzale AR, Salvan A. Improved inference on a scalar fixed effect of interest in nonlinear mixed-effects models. *Computational Statistics & Data Analysis* 2006; **51**: 1602–1613. DOI:10.1016/j.csda.2006.05.016

A Derivation of Skovgaard's statistic

Given the framework described in Section 2, the log-likelihood function $\ell(\theta)$ for the whole parameter vector θ is

$$\ell(\theta) \propto -\frac{1}{2}\sum_{i=1}^{n}\log|V_i| - \frac{1}{2}\sum_{i=1}^{n}(y_i - f_i)^\top V_i^{-1}(y_i - f_i),$$

where $y_i = (\hat{\eta}_i, \hat{\xi}_i)^\top$ is the observed value of the random vector Y_i in (3), with mean vector f_i and variance/covariance matrix V_i, following the notation in Section 2.2. The score vector

$$\ell_\theta(\theta) = \begin{bmatrix} \ell_{\beta_0}(\theta) \\ \ell_{\beta_1}(\theta) \\ \ell_\mu(\theta) \\ \ell_{\tau^2}(\theta) \\ \ell_{\sigma^2}(\theta) \end{bmatrix}$$

has components

$$\begin{aligned}
\ell_{\beta_j}(\theta) \quad &= \quad -\frac{1}{2}\sum_{i=1}^{n}\operatorname{trace}\left(V_i^{-1}V_{i,\beta_j}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{n}\left(y_i^\top V_{i,\beta_j}^{-1}y_i - 2f_{i,\beta_j}^\top V_i^{-1}y_i - 2f_i^\top V_{i,\beta_j}^{-1}y_i + 2f_{i,\beta_j}V_i^{-1}f_i + f_i^\top V_{i,\beta_j}^{-1}f_i\right), j = 0, 1,
\end{aligned}$$

$$\ell_\mu(\theta) = \sum_{i=1}^{n} f_{i,\mu}^\top V_i^{-1}(y_i - f_i)$$

and

$$\ell_{\psi_j}(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\operatorname{trace}\left(V_i^{-1}V_{i,\psi_j}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(y_i^\top V_{i,\psi_j}^{-1}y_i - 2f_i^\top V_{i,\psi_j}^{-1}y_i + f_i^\top V_{i,\psi_j}^{-1}f_i\right),\ \psi_j \in \{\tau^2, \sigma^2\}.$$

The expected information matrix

$$i(\theta) = \begin{bmatrix}
i_{\beta_0\beta_0}(\theta) & i_{\beta_0\beta_1}(\theta) & i_{\beta_0\mu}(\theta) & i_{\beta_0\tau^2}(\theta) & i_{\beta_0\sigma^2}(\theta) \\
i_{\beta_0\beta_1}(\theta) & i_{\beta_1\beta_1}(\theta) & i_{\beta_1\mu}(\theta) & i_{\beta_1\tau^2}(\theta) & i_{\beta_1\sigma^2}(\theta) \\
i_{\beta_0\mu}(\theta) & i_{\beta_1\mu}(\theta) & i_{\mu\mu}(\theta) & i_{\mu\tau^2}(\theta) & i_{\mu\sigma^2}(\theta) \\
i_{\beta_0\tau^2}(\theta) & i_{\beta_1\tau^2}(\theta) & i_{\mu\tau^2}(\theta) & i_{\tau^2\tau^2}(\theta) & i_{\tau^2\sigma^2}(\theta) \\
i_{\beta_0\sigma^2}(\theta) & i_{\beta_1\sigma^2}(\theta) & i_{\mu\sigma^2}(\theta) & i_{\tau^2\sigma^2}(\theta) & i_{\sigma^2\sigma^2}(\theta)
\end{bmatrix}$$

has generic component

$$i_{\theta_j,\theta_k} = \frac{1}{2}\sum_{i=1}^{n}\operatorname{trace}\left(V_{i,\theta_k}^{-1}V_{i,\theta_j} + V_i^{-1}V_{i,\theta_j\theta_k} - V_{i,\theta_j}^{-1}V_{i,\theta_j\theta_k}V_{i,\theta_j}^{-1}V_i\right) + \sum_{i=1}^{n} f_{i,\theta_j}V_i^{-1}f_{i,\theta_k},\ \theta_j, \theta_k \in \theta,$$

where $V_{i,\theta_j\theta_k}$ denotes the second derivative of V_i with respect to $\theta_j, \theta_k \in \theta$. In order to derive the components of S and q, consider that

$$\operatorname{cov}\left(Y_i^\top \hat{V}_{i,\theta_j}^{-1} Y_i, Y_i^\top \tilde{V}_{i,\theta_k} Y_i\right) = \operatorname{trace}\left(\hat{V}_{i,\theta_j}^{-1} \hat{V}_i \tilde{V}_{i,\theta_k}^{-1} \hat{V}_i\right) + 4\hat{f}_i^\top \hat{V}_{i,\theta_j}^{-1} \hat{V}_i \tilde{V}_{i,\theta_k}^{-1} \hat{f}_i,$$

$$\operatorname{cov}\left(Y_i^\top \hat{V}_{i,\theta_j}^{-1} Y_i, Y_i^\top \tilde{V}_{i,\theta_k} Y_i\right) = 2 f_i^\top \hat{V}_{i,\theta_j}^{-1} \hat{V}_i \tilde{V}_i^{-1} \tilde{f}_{i,\theta_j}$$

and

$$\operatorname{cov}\left(\hat{f}_{i,\theta_j} \hat{V}_i^{-1} Y_i, \tilde{f}_i^\top \tilde{V}_{i,\theta_k}^{-1} Y_i\right) = \hat{f}_{i,\theta_j}^\top \hat{V}_i^{-1} \hat{V}_i \tilde{V}_{i,\theta_k}^{-1} \tilde{f}_i,$$

for $\theta_j, \theta_k \in \theta$.

Then,

$$\begin{aligned}
S_{\beta_j,\beta_k}(\theta) &= \operatorname{cov}\left\{\ell_{\beta_j}(\theta_1), \ell_{\beta_k}(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\operatorname{cov}\sum_{i=1}^n \left(Y_i \hat{V}_{i,\beta_j}^{-1} Y_i - 2\hat{f}_{i,\beta_j}^\top \hat{V}_i^{-1} Y_i - 2\hat{f}_i \hat{V}_{i,\beta_j}^{-1} Y_i, Y_i^\top \tilde{V}_{i,\beta_k}^{-1} Y_i - 2\tilde{f}_{i,\beta_k}^\top \tilde{V}_i^{-1} Y_i - 2\tilde{f}_i^\top \tilde{V}_{i,\beta_k}^{-1} Y_i\right) \\
&= \sum_{i=1}^n \left\{\frac{1}{2}\operatorname{trace}\left(\hat{V}_{i,\beta_j}^{-1} \hat{V}_i \tilde{V}_{i,\beta_k}^{-1} \hat{V}_i\right) + \hat{f}_{i,\beta_j}^\top \tilde{V}_{i,\beta_k}^{-1}\left(\tilde{f}_i - \hat{f}_i\right) + \hat{f}_{i,\beta_k} \tilde{V}_i^{-1} \tilde{f}_{i,\beta_k}\right\},\ j,k = 0,1
\end{aligned}$$

$$\begin{aligned}
S_{\beta_j,\mu}(\theta) &= \operatorname{cov}\left\{\ell_{\beta_j}(\theta_1), \ell_\mu(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{2}\operatorname{cov}\sum_{i=1}^n \left(Y_i^\top \hat{V}_{i,\beta_j}^{-1} Y_i - \hat{f}_{i,\beta_j}^\top \hat{V}_i^{-1} Y_i - 2\hat{f}_i \hat{V}_{i,\beta_j}^{-1} Y_i, \tilde{f}_{i,\mu}^\top \tilde{V}_i^{-1} Y_i\right) \\
&= \sum_{i=1}^n \hat{f}_{i,\beta_j}^\top \tilde{V}_i^{-1} \tilde{f}_{i,\mu},\ j = 0,1
\end{aligned}$$

$$\begin{aligned}
S_{\mu,\mu}(\theta) &= \operatorname{cov}\left\{\ell_\mu(\theta_1), \ell_\mu(\theta_2)\right\}\big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \operatorname{cov}\sum_{i=1}^n \left(\hat{f}_{i,\mu}^\top \hat{V}_i^{-1} Y_i, \tilde{f}_{i,\mu}^\top \tilde{V}_i^{-1} Y_i\right) \\
&= \sum_{i=1}^n \hat{f}_{i,\mu}^\top \tilde{V}_i^{-1} \tilde{f}_{i,\mu}
\end{aligned}$$

$$\begin{aligned}
S_{\beta_j,\psi_k}(\theta) &= \operatorname{cov}\left\{\ell_{\beta_j}(\theta_1), \ell_{\psi_k}(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{2}\operatorname{cov}\sum_{i=1}^n \left(Y_i^\top \hat{V}_{i,\beta_j}^{-1} Y_i - 2\hat{f}_{i,\beta_j}^\top \hat{V}_i^{-1} Y_i - 2\hat{f}_i^\top \hat{V}_{i,\beta_j}^{-1} Y_i, Y_i^\top \tilde{V}_{i,\psi_k}^{-1} Y_i - 2\tilde{f}_i^\top \tilde{V}_{i,\psi_k}^{-1} Y_i\right) \\
&= \sum_{i=1}^n \left\{\frac{1}{2}\operatorname{trace}\left(\hat{V}_{i,\beta_j}^{-1} \hat{V}_i \tilde{V}_{i,\psi_k}^{-1} \hat{V}_i\right) + \hat{f}_{i,\beta_j}^\top \tilde{V}_{i,\psi_k}^{-1}\left(\tilde{f}_i - \hat{f}_i\right)\right\},\ j = 0,1,\ \psi_k \in \{\tau^2, \sigma^2\}
\end{aligned}$$

$$
\begin{aligned}
S_{\mu,\psi_k}(\theta) &= \operatorname{cov}\left\{\ell_\mu(\theta_1), \ell_{\psi_k}(\theta_2)\right\}\big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= -\frac{1}{2}\operatorname{cov}\sum_{i=1}^{n}\left(\hat{f}_{i,\mu}\hat{V}_i^{-1}Y_i, Y_i^\top \tilde{V}_i^{-1}Y_i - 2\tilde{f}_i^\top \tilde{V}_{i,\psi_k}^{-1}Y_i\right) \\
&= \sum_{i=1}^{n}\left\{\frac{1}{2}\text{trace}\left(\hat{V}_{i,\mu}^{-1}\hat{V}_i\tilde{V}_{i,\psi_k}^{-1}\hat{V}_i\right) + \hat{f}_{i,\mu}^\top \tilde{V}_{i,\psi_k}^{-1}\left(\tilde{f}_i - \hat{f}_i\right)\right\},\ \psi_k \in \{\tau^2, \sigma^2\}
\end{aligned}
$$

$$
\begin{aligned}
S_{\psi_j,\psi_k}(\theta) &= \operatorname{cov}\left\{\ell_{\psi_j}(\theta_1), \ell_{\psi_k}(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\operatorname{cov}\sum_{i=1}^{n}\left(Y_i^\top \hat{V}_{i,\psi_j}^{-1}Y_i - 2\hat{f}_i^\top \hat{V}_{i,\psi_j}^{-1}Y_i, Y_i^\top \tilde{V}_{i,\psi_k}^{-1}Y_i - 2\tilde{f}_i^\top \tilde{V}_{i,\psi_k}^{-1}Y_i\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\text{trace}\left(\hat{V}_{i,\psi_j}^{-1}\hat{V}_i\tilde{V}_{i,\psi_k}^{-1}\hat{V}_i\right),\ \psi_j,\psi_k \in \{\tau^2, \sigma^2\}
\end{aligned}
$$

$$
\begin{aligned}
S_{\mu,\beta_j}(\theta) &= \operatorname{cov}\left\{\ell_\mu(\theta_1), \ell_{\beta_k}(\theta_2)\right\}\big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= -\frac{1}{2}\operatorname{cov}\sum_{i=1}^{n}\left(\hat{f}_{i,\mu}^\top \hat{V}_i^{-1}Y_i, Y_i^\top \tilde{V}_{i,\beta_j}^{-1}Y_i - 2\tilde{f}_{i,\beta_j}^\top \tilde{V}_i^{-1}Y_i - 2\tilde{f}_i\tilde{V}_{i,\beta_j}^{-1}Y_i\right) \\
&= \sum_{i=1}^{n}\left(\hat{f}_{i,\mu}^\top \tilde{V}_i^{-1}\tilde{f}_{i,\beta_j} + \hat{f}_{i,\mu}^\top \tilde{V}_{i,\beta_j}^{-1}\tilde{f}_i - \hat{f}_i^\top \tilde{V}_{i,\beta_j}^{-1}\hat{f}_{i,\mu}\right),\ j = 0, 1
\end{aligned}
$$

$$
\begin{aligned}
S_{\psi_j,\beta_k}(\theta) &= \operatorname{cov}\left\{\ell_{\psi_j}(\theta_1), \ell_{\beta_k}(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\operatorname{cov}\sum_{i=1}^{n}\left(Y_i^\top \hat{V}_{i,\psi_j}^{-1}Y_i - 2\hat{f}_i^\top \hat{V}_{i,\psi_j}^{-1}Y_i, Y_i^\top \tilde{V}_{i,\beta_k}^{-1}Y_i - 2\tilde{f}_{i,\beta_k}^\top \tilde{V}_i^{-1}Y_i - 2\tilde{f}_i^\top \tilde{V}_{i,\beta_k}^{-1}Y_i\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\text{trace}\left(\hat{V}_{i,\psi_j}^{-1}\hat{V}_i\tilde{V}_{i,\beta_k}^{-1}\hat{V}_i\right), \psi_j \in \{\tau^2, \sigma^2\}, k = 0, 1
\end{aligned}
$$

$$
\begin{aligned}
S_{\psi_j,\mu}(\theta) &= \operatorname{cov}\left\{\ell_{\psi_j}(\theta_1), \ell_\mu(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= -\frac{1}{2}\operatorname{cov}\sum_{i=1}^{n}\left(Y_i^\top \hat{V}_{i,\psi_k}^{-1}Y_i - 2\hat{f}_i^\top \hat{V}_{i,\psi_k}^{-1}Y_i, \tilde{f}_{i,\mu}^\top \tilde{V}_i^{-1}Y_i\right) \\
&= 0,\ \psi_j \in \{\tau^2, \sigma^2\}
\end{aligned}
$$

$$
\begin{aligned}
q_{\beta_j}(\theta) &= \operatorname{cov}\left\{\ell_{\beta_\mathrm{j}}(\theta_1), \ell(\theta_1) - \ell(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\operatorname{cov}\sum_{i=1}^{n}\left(Y_i^\top \hat{V}_{i,\beta_j}^{-1}Y_i - 2\hat{f}_{i,\beta_j}^\top \hat{V}_i^{-1}Y_i - 2\hat{f}_i^\top \hat{V}_{i,\beta_j}^{-1}Y_i, Y_i^\top \hat{V}_i^{-1}Y_i - 2f_i^\top \hat{V}_i^{-1}Y_i - Y_i^\top \tilde{V}_i^{-1}Y_i + 2\tilde{f}_i\tilde{V}_i^{-1}Y_i\right) \\
&= \sum_{i=1}^{n}\left[\frac{1}{2}\text{trace}\left\{\hat{V}_{i,\beta_j}^{-1}\hat{V}_i\left(\hat{V}_i^{-1} - \tilde{V}_i^{-1}\right)\hat{V}_i\right\} + \hat{f}_{i,\beta_j}^\top \tilde{V}_i^{-1}\left(\hat{f}_i - \tilde{f}_i\right)\right],\ j = 0, 1
\end{aligned}
$$

$$
\begin{aligned}
q_{\mu}(\theta) &= \text{cov}\left\{\ell_{\mu}(\theta_1), \ell(\theta_1) - \ell(\theta_2)\right\}\big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\text{cov}\sum_{i=1}^{n}\left(Y_i^{\top}\hat{V}_{i,\mu}^{-1}Y_i - 2\hat{f}_{i,\mu}^{\top}\hat{V}_i^{-1}Y_i - 2\hat{f}_i^{\top}\hat{V}_{i,\mu}^{-1}Y_i, Y_i^{\top}\hat{V}_i^{-1}Y_i - 2f_i^{\top}\hat{V}_i^{-1}Y_i - Y_i^{\top}\tilde{V}_i^{-1}Y_i + 2\tilde{f}_i\tilde{V}_i^{-1}Y_i\right) \\
&= \sum_{i=1}^{n}\left[\frac{1}{2}\text{trace}\left\{\hat{V}_{i,\mu}^{-1}\hat{V}_i\left(\hat{V}_i^{-1} - \tilde{V}_i^{-1}\right)\hat{V}_i\right\} + \hat{f}_{i,\mu}^{\top}\tilde{V}_i^{-1}\left(\hat{f}_i - \tilde{f}_i\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
q_{\psi_j}(\theta) &= \text{cov}\left\{\ell_{\psi_j}(\theta_1), \ell(\theta_1) - \ell(\theta_2)\right\}\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \\
&= \frac{1}{4}\text{cov}\sum_{i=1}^{n}\left(Y_i^{\top}\hat{V}_{i,\psi_j}^{-1}Y_i - 2\hat{f}_{i,\psi_j}^{\top}\hat{V}_i^{-1}Y_i - 2\hat{f}_i^{\top}\hat{V}_{i,\mu}^{-1}Y_i, Y_i^{\top}\hat{V}_i^{-1}Y_i - 2f_i^{\top}\hat{V}_i^{-1}Y_i - Y_i^{\top}\tilde{V}_i^{-1}Y_i + 2\tilde{f}_i\tilde{V}_i^{-1}Y_i\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\left\{\text{trace}\left(\hat{V}_{\psi_j}^{-1}\hat{V}_i\right) - \text{trace}\left(\hat{V}_{\psi_j}^{-1}\hat{V}_i\tilde{V}_i^{-1}\hat{V}_i\right)\right\}, \ \psi_j \in \{\tau^2, \sigma^2\}
\end{aligned}
$$

B Data analysis

This appendix shows how to evaluate Skovgaard's statistic for inference on the slope of the control rate regression in the `R` programming language. The illustration is based on the data of Hoes et al. [22] reported in Table 1. Functions needed to implement Skovgaard's statistic are obtained as supplementary material and they can be loaded as follows

```
R> source("control_rate_regression_LRTs.R")
```

Consider the hypothesis test $\beta_1 = 1$ against the two-sided alternative. Wald statistic, the signed profile log-likelihood ratio statistic and Skovgaard's statistic are obtained by applying function `crr.test` (control rate regression test)

```
crr.test(data, beta1.null, alternative = c("two.sided", "less",
    "greater"), maxit = 1000)
```

with arguments

- `data`: the dataset
- `beta1.null`: the value of β_1 under the null hypothesis
- `alternative`: a character string specifying the alternative hypothesis, chosen between "two.sided" (default), "greater" or "less"; just the initial letter can be specified

- `maxit`: the maximum number of iterations for the Nelder and Mead [21] optimization algorithm; default value 1,000

The dataset is composed by n rows corresponding to the studies recruited in the meta-analysis and 6 columns including the values of $\hat{\eta}_i$, $\hat{\xi}_i$, and the elements of the variance/covariance matrix Γ_i inserted by row, namely, $\text{var}(\hat{\eta}_i), \text{cov}(\hat{\eta}_i, \hat{\xi}_i), \text{cov}(\hat{\eta}_i, \hat{\xi}_i), \text{var}(\hat{\xi}_i)$. For the analysis of Hoes et al. [22] data, the object to be passed to function `crr.test` can be constructed as follows

```
R> deaths.treated <- c(10, 2, 54, 47, 53, 10, 25, 47, 43, 25, 157, 92)
R> ## number of person-years for the cases
R> py.treated <- c(595.2, 762, 5635, 5135, 3760, 2233, 7056.1, 8099, 5810, 5397,
R+                 22162.7, 20885)
R> deaths.controls <- c(21, 0, 70, 63, 62, 9, 35, 31, 39, 45, 182, 72)
R> deaths.controls[2] <- 0.5
R> ## number of person-years for the controls
R> py.controls <- c(640.2, 756, 5600, 4960, 4210, 2084.5, 6824, 8267, 5922, 5173,
R+                 22172.5, 20645)
R> py.controls[2] <- py.controls[2]+0.5
R> hoes.data.original <- data.frame(deaths.treated, py.treated,
R+                                  deaths.controls, py.controls)
## estimated log event rate for the controls
R> xi.obs <- log(hoes.data.original$deaths.treated/hoes.data.original$py.treated)
## estimated log event rate for the treated
R> eta.obs <- log(hoes.data.original$deaths.controls/hoes.data.original$py.controls)
R> n <- length(hoes.data.original$deaths.treated)
## variance/covariance matrix
R> gamma.matrix <- matrix(0.0, ncol=4, nrow=n)
R> for(i in 1:n)
R+   gamma.matrix[i,] <- c(1/hoes.data.original$deaths.treated[i], 0,
R+                         0, 1/hoes.data.original$deaths.controls[i])
R> hoes.data <- data.frame(eta.obs, xi.obs, gamma.matrix)
R> colnames(hoes.data) <- c('eta.obs', 'xi.obs', 'var.eta', 'cov.etaxi',
R+                          'cov.etaxi', 'var.xi')
```

Function `crr.test`

```
R> crr.test(data=hoes.data, beta1.null=1, alternative='two.sided')


Estimate of beta1:
     Estimate  Std.Err.
WLS  0.60973   0.10892
MLE  0.68917   0.08124


Hypothesis test for beta1:
                                              Value      P-value
Wald statistic                                -3.5830787  0.0003396
Signed profile log-likelihood ratio statistic -2.3447177  0.0190415
Skovgaard statistic                           -1.2709290  0.2037539


alternative hypothesis: parameter is different from 1
```

provides the following information:

- the weighted least squares estimate and the maximum likelihood estimate of β_1;
- the associated standard error;
- the value of Wald statistic, the value of the signed profile log-likelihood ratio statistic r_P and the value of Skovgaard's statistic $\bar{r}_P$ under the null hypothesis;
- the p-value of the test based on the three statistics for the specified alternative hypothesis.