Likelihood-Based Inference in Control Risk Regression with Study-Specific Covariates

Phuc Thien Tran [1] Annamaria Guolo [2]

[1]University Medical Center Göttingen [2]University of Padova

Background

- Meta-analysis is a well-established tool to combine, synthesize and summarize results from independent studies to address one or many questions of interest.
- Heterogeneity between studies can be due to design, patients' characteristics or comparison intervention.
- Control risk regression is a meta-analysis about the effectiveness of a treatment including a measure of risk for controls as covariate.
- Let n be the number of studies, η_i and ξ_i denote unobserved treatment risk and baseline risk, resp., $i = 1, \cdots, n$. A basic control risk regression model is (Arends et al., 2000)

$$\eta_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i, \;\; \xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \;\; \varepsilon_i \sim N\left(0, \tau^2\right),$$

  where τ^2 is the residual variance and σ_ξ^2 is the variance of ξ_i.
- Observed measures of treatment risk and baseline risk for dichotomous outcome are

$$\hat{\eta}_i = \log\left(\frac{y_i}{n_{i1} - y_i}\right), \;\; \hat{\xi}_i = \log\left(\frac{x_i}{n_{i0} - x_i}\right),$$

  where y_i and x_i denote the outcomes in the treatment group and in the control group, resp.
- $\hat{\eta}_i$ and $\hat{\xi}_i$ are observed (summary) versions of η_i and ξ_i, so they prone to measurement errors. Ignoring measurement errors can affect inference, e.g. with the risk of regression dilution (Carroll et al., 2006).
- Ignoring measurement error of $\hat{\eta}_i$ and $\hat{\xi}_i$ may result in more homogeneous treatment effects than really exist, or regression dilution.
- An approximate measurement error model accounts for the measurement error by assuming normality for observed measures given the true ones

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \middle| \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim N_2\left\{ \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, \begin{pmatrix} s_{\eta_i}^2 & 0 \\ 0 & s_{\xi_i}^2 \end{pmatrix} \right\},$$

  where $s_{\eta_i}^2 = 1/y_i + 1/\left(n_{i1} - y_i\right)$ and $s_{\xi_i}^2 = 1/x_i + 1/\left(n_{i0} - x_i\right)$ are the within-study variances of η_i and ξ_i, resp.

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2\left\{ \begin{pmatrix} \beta_0 + \beta_1\mu_\xi \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} \beta_1^2\sigma_\xi^2 + \tau^2 + s_{\eta_i}^2 & \beta_1\sigma_\xi^2 \\ \beta_1\sigma_\xi^2 & \sigma_\xi^2 + s_{\xi_i}^2 \end{pmatrix} \right\}.$$

Objective

- Including covariates other than control risk measure in order to better explain between-study heterogeneity

$$\eta_i = \beta_0 + \beta_1\xi_i + \beta_2\zeta_i + \varepsilon_i,$$
$$\xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \;\; \zeta_i \sim N\left(\mu_\zeta, \sigma_\zeta^2\right), \;\; \varepsilon_i \sim N\left(0, \tau^2\right),$$

  where ζ_i denotes true values of covariate, μ_ζ and σ_ζ^2 are mean and between-studies variance of ζ_i, resp. ζ_i and ξ_i are assumed independent
- Error-free or error-affected covariates
- Likelihood and pseudo-likelihood inference

Model with Error-Free Covariates

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2\left\{ \begin{pmatrix} \beta_0 + \beta_1\mu_\xi + \beta_2\zeta_i \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} \beta_1^2\sigma_\xi^2 + \tau^2 + s_{\eta_i}^2 & \beta_1\sigma_\xi^2 \\ \beta_1\sigma_\xi^2 & \sigma_\xi^2 + s_{\xi_i}^2 \end{pmatrix} \right\}.$$

Model with Error-Affected Covariates

- When covariates are summary information, the relationship between covariates and response at the study level might be different from the one at the individual level (aggregation bias, Schmid et al., 2004).
- Covariates can be mismeasured due to quality of instruments or included studies.
- $\hat{\zeta}_i$: observed measure of ζ_i. Thus, the approximate measurement error model is

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \\ \hat{\zeta}_i \end{pmatrix} \middle| \begin{pmatrix} \eta_i \\ \xi_i \\ \zeta_i \end{pmatrix} \sim N_3\left\{ \begin{pmatrix} \eta_i \\ \xi_i \\ \zeta_i \end{pmatrix}, \Gamma_i = \begin{pmatrix} s_{\eta_i}^2 & 0 & s_{\eta_i\zeta_i} \\ 0 & s_{\xi_i}^2 & s_{\xi_i\zeta_i} \\ s_{\eta_i\zeta_i} & s_{\xi_i\zeta_i} & s_{\zeta_i}^2 \end{pmatrix} \right\}.$$

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \\ \hat{\zeta}_i \end{pmatrix} \sim N_3\left\{ \begin{pmatrix} \beta_0 + \beta_1\mu_\xi + \beta_2\mu_\zeta \\ \mu_\xi \\ \mu_\zeta \end{pmatrix}, \begin{pmatrix} \beta_1^2\sigma_\xi^2 + \beta_2^2\sigma_\zeta^2 + \tau^2 + s_{\eta_i}^2 & \beta_1\sigma_\xi^2 & \beta_2\sigma_\zeta^2 + s_{\eta_i\zeta_i} \\ \beta_1\sigma_\xi^2 & \sigma_\xi^2 + s_{\xi_i}^2 & s_{\xi_i\zeta_i} \\ \beta_2\sigma_\zeta^2 + s_{\eta_i\zeta_i} & s_{\xi_i\zeta_i} & \sigma_\zeta^2 + s_{\zeta_i}^2 \end{pmatrix} \right\}.$$

Proposed Likelihood and Pseudo-Likelihood Approaches

- Let $f_i\left(s_{\zeta_i}^2, s_{\eta_i\zeta_i}, s_{\xi_i\zeta_i}\right)$ denote the density function of $\left(\hat{\eta}_i, \hat{\xi}_i, \hat{\zeta}_i\right)^\top$.
- Parameter of interest is $\theta = \left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$.
- Components of Γ_i except for $s_{\eta_i}^2$ and $s_{\xi_i}^2$ are often unavailable or unable to compute due to limited information.
- If $\hat{\zeta}_i = \zeta_i + \delta_i, \delta_i \sim N\left(0, \sigma_\delta^2\right)$, only σ_δ^2 is known and a pseudo-likelihood approach is proposed

$$pL\left(\theta\right) = \prod_{i=1}^{n} f_i\left(\sigma_\delta^2, 0, 0\right).$$

- If $\hat{\zeta}_i$ is summary information, let k be indicator of event, j be indicator of control group, i be study number, $\left(x_{ijk}, y_{ijk}, z_{ijk}\right)^\top$ be outcome corresponding to $\left(\xi, \eta, \zeta\right)^\top$.
- If ζ_i is mean, $s_{\zeta_i}^2$ is known and
  - if $\hat{\zeta}_{ijk}$ is unavailable, pseudo-likelihood approach is proposed since $s_{\eta_i\zeta_i}$ and $s_{\xi_i\zeta_i}$ are uncomputable

$$pL\left(\theta\right) = \prod_{i=1}^{n} f_i\left(s_{\zeta_i}^2, 0, 0\right).$$

  - otherwise, likelihood approach can be used based on approximation suggested by Bagos (2012)

$$L\left(\theta\right) = \prod_{i=1}^{n} f_i\left(s_{\zeta_i}^2, \frac{1}{n_i}\left(\hat{\zeta}_{i11} - \hat{\zeta}_{i10}\right), \frac{1}{n_i}\left(\hat{\zeta}_{i01} - \hat{\zeta}_{i00}\right)\right).$$

- If ζ_i is log odds, $s_{\zeta_i}^2 = 1/z_i + 1/\left(n_{i1} + n_{i0} - z_i\right)$ and
  - if $\hat{\zeta}_{ijk}$ is unavailable, pseudo-likelihood approach is proposed since $s_{\eta_i\zeta_i}$ and $s_{\xi_i\zeta_i}$ are uncomputable

$$pL\left(\theta\right) = \prod_{i=1}^{n} f_i\left(s_{\zeta_i}^2, 0, 0\right).$$

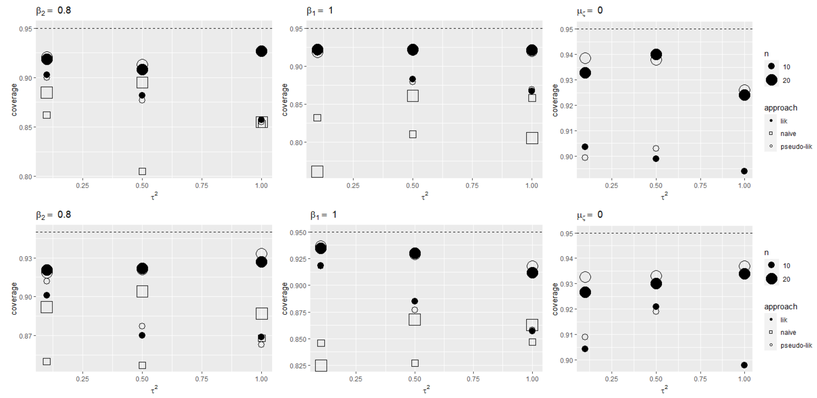
  - otherwise, likelihood approach can be used based on approximation by Bagos (2012)

$$L\left(\theta\right) = \prod_{i=1}^{n} f_i\left(s_{\zeta_i}^2, \frac{1}{z_i}\left(\frac{z_{i11}}{y_i} - \frac{z_{i10}}{n_{i1} - y_i}\right) - \frac{1}{n_i - z_i}\left(\frac{n_{i11} - z_{i11}}{y_i} - \frac{n_{i10} - z_{i10}}{n_{i1} - y_i}\right),$$
$$\frac{1}{z_i}\left(\frac{z_{i01}}{x_i} - \frac{z_{i00}}{n_{i0} - x_i}\right) - \frac{1}{n_i - z_i}\left(\frac{z_{i01} - z_{i01}}{x_i} - \frac{n_{iC0} - z_{i00}}{n_{i0} - x_i}\right)\right).$$

- Sandwich matrix is proposed when using pseudo likelihood approach to account for the risk of misspecification.

Simulation Study

No. of replicates	$1,000$
Scenario	$n \in \{10, 20\}$, $\tau^2 \in \{0.1, 0.5, 1.0\}$, $\xi_i \sim N(0,1), SN(0,1,-5)$
Parameter	$\left(\beta_0, \beta_1, \beta_2, \mu_\zeta, \sigma_\zeta^2\right)^\top = (0, 1, 0.8, 0, 1)^\top$
Group size	$n_{i1} \sim n_{i0} \sim U(15, 200)$
Metric	Bias, standard error, standard deviation, empirical coverage probability, convergence
Aim	Evaluate performance of likelihood and pseudo likelihood approaches



Schizophrenia Dataset (Pardamean et al., 2022)

- Consider a meta-analysis of 10 studies, we want to assess the relationship between the risk of mortality in patients getting COVID-19 and schizophrenia, to the baseline risk and other several other factors.
- Control risk regression with additional covariates is fitted using the naive analysis, likelihood approach and pseudo-likelihood approach.

Additional covariate	Approach	β_1	β_2	τ^2
No	likelihood	**0.761 (0.091)**	-	0.028 (0.049)
	naive analysis	**0.709 (0.056)**	-	0.527 (0.211)
Scaled mean age	pseudo-likelihood	**1.079 (0.186)**	**-0.570 (0.270)**	2.5e-05 (0.001)
	naive analysis	**1.004 (0.149)**	**-0.492 (0.236)**	0.466 (0.174)
Log odds of male	pseudo-likelihood	**0.755 (0.063)**	-0.612 (0.418)	0.002 (0.008)
	naive analysis	**0.724 (0.057)**	-0.422 (0.407)	0.492 (0.184)
Log odds of diabetes	pseudo-likelihood	**0.766 (0.137)**	-0.014 (0.269)	0.028 (0.061)
	naive analysis	**0.718 (0.110)**	-0.025 (0.251)	0.524 (0.196)

References

[1] Arends LR et al. (2000). Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Stat Med*, **19** (24), 3497–3518.

[2] Bagos PG et al. (2012). On the covariance of two correlated log-odds ratios. *Stat Med*, **31** (14), 1418-1431.

[3] Carroll RJ et al. (2006). Measurement error in nonlinear models: a modern perspective. *CRC*, **2nd Ed.**

[4] Pardamean E et al. (2022). Mortality from coronavirus disease 2019 (COVID-19) in patients with schizophrenia: A systematic review, meta-analysis and meta-regression. *Gen. Hosp. Psychiatry*, **31** (75), 61-67.

[5] Schmid CH et al. (2004). Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol*, **57** (7), 683-697.

thien.tran@med.uni-goettingen.de