MATHEMATICS IN ENGINEERING, SCIENCE AND AEROSPACE
MESA - www.journalmesa.com
Vol. 2, No. 3, pp. 251-265, 2011

Validity of regression meta-analyses versus pooled analyses of mixed effect linear models

Eric V. Slud [1⋆], Mathewos DeMissie [1]

[1] Mathematics Dept., Univ. of Maryland College Park, USA

[⋆] *Corresponding Author*. E-mail: evs@math.umd.edu

Abstract. Meta-analysis is a class of techniques for combining the results of many small studies to reach a unified inference for a parameter of interest, for example a parameter quantifying treatment effectiveness. Often, the statistician has from each individual study only an estimator of the parameter of interest, together with an estimator of standard error, and the separate estimates are treated as data and fitted within a so-called 'meta-regression' model, with study as a categorical predictor and the parameter of interest regarded as a common mean, and with error term consisting of a constant-variance error plus an independent study effect with standard deviation equal to the separate estimated study standard error. Occasionally, patient-level data including covariates are available from all of the component studies, in which case a pooled patient-level analysis can be conducted including fixed covariate effects, plus study effects as random intercepts, and possibly also including random treatment-by-study interactions. Comparisons are sometimes published in biomedical settings between the meta-analysis results and the results of pooled mixed (generalized) linear model analyses. This paper reports theoretical and simulation results regarding the biases of meta-analysis estimates versus pooled model estimates when the latter are correct, showing that in *linear* models, these biases are generally small, especially when treatment-allocation is balanced and covariate and error distributions are symmetric.

1 Introduction

Meta-analysis is a large field of statistical methodology (Hartung, Knapp and Sinha 2008), with extensive applications in biomedical statistics, educational statistics and social sciences. The unifying thread in this field is the use of summary statistical point estimates and standard errors for a parameter which is shared across a number of independent studies to provide a combined inference for the common parameter. Generally, the point estimates from the separate studies are treated as data within a combined model, often with random study effect standard deviations assumed to be given or estimated by the estimated standard error from the separate studies. Such an aggregate level analysis combining

[2010] **Mathematics Subject Classification**: Primary: 62F10, Secondary: 62F12, 62J05.
Keywords: Bias of ML estimators, Kullback-Leibler minimization, treatment by covariate interaction.

multiple studies can usually not rely on access to unit- or patient- level data from the separate studies. However, sometimes such access is possible, in which case one could confront the meta-analysis with a unified multilevel statistical model at unit level, with 'study' as a clustering or group variable which may require modeling with group random effects and possibly interaction random effects with any fixed-effect covariates in the separate studies. Examples of such comparative analyses are Jiao et al. 2010, and Murphy et al. 2009. When such comparisons have occasionally been made in applied settings, the results (estimates and confidence intervals for combined treatment effects) of the meta- and unified analyses are often quite similar. Nevertheless, use of meta-analytic techniques would seem to require strong assumptions about the ways in which treatments, covariates, and centers interact. To learn about the empirical agreement among meta- analytic and unified unit-level model analyses of treatment effect, one could also regard *center* within a large multi-center clinical trial as the study cluster variable and perform both types of analyses along with simulation experiments based on realistic parameters for various models of the combined patient-level data. This approach was followed by DeMissie (2009) using the ECOG EST 1582 lung cancer clinical trial data previously analyzed by Gray (1994).

The purpose of the present paper is to examine necessary conditions for agreement and examples where such agreement can fail, within the important context of mixed effect, not necessarily Gaussian, linear models for treatment effect within randomized clinical trials. To narrow the focus of this research, we restrict attention first to the case of trials with quantitative responses, typically either log survival times or the results of a continuous bioassay (size of tumor, level of chemical in the blood, etc.) For the survival-time setting, we assume for simplicity that censoring can be ignored, either because survival rates are very small or because almost all censoring is administrative, occurring after a time-on-test interval that is the same for all patients. Next, we restrict attention to randomized-allocation patient-level models which are linear, with terms which may include fixed-effect interaction terms between treatment and covariates and random effects and interactions among center and treatment and covariates.

Thus, the underlying models we consider — as being general enough to accommodate many but not all phenomena of interest — are of patient-level quantitative responses at the level of patient $j = 1,\ldots,n_i$ within study-center $i = 1,\ldots,m,$ as expressed by

$$Y_{ij} = \mu + u_i + \xi_{ij}(\vartheta + v_i) + X'_{ij}(\beta + b_i) + \xi_{ij} Z'_{ij}(\gamma + \rho_i) + \varepsilon_{ij} \tag{1.1}$$

where Y_{ij} denotes the real-valued quantitative response such as log survival time; X_{ij} and Z_{ij} respectively denote d-dimensional patient-level vectors of observed baseline covariates (including possible interaction terms among them); ξ_{ij} denotes the randomized binary treatment indicator; u_i, v_i, b_i, ρ_i are respectively center random-effects, center-by-treatment random effects, treatment random effects, and treatment-by-covariate-by-center random effects; and ε_{ij} are random error-terms. The unknown fixed-effect parameter ϑ is taken to be the common parameter of interest, at least in all cases where $\gamma \equiv \mathbf{0}$, but β and the variance scale-parameters of the random effects and errors are unknown nuisance parameters. Within model (1.1), the columns of Z_{ij} are generally also included among those of X_{ij}, and for interpretability of the model coefficients, the X_{ij} and Z_{ij} vectors are generally centered at 0 when averaged over i. The different random-effect terms and errors ε_{ij} are assumed to be jointly independent, and each of these error-terms is assumed *iid* across its index values. At least initially, the random-effect terms u_i, v_i, b_i (and ρ_i when present) are assumed to be normally distributed with mean 0. The distribution of ε_{ij} is assumed to be known up to a scale parameter but not necessarily normal.

Most treatments of meta-analysis within the context just outlined have restricted attention to normally distributed mean-0 response errors ε_{ij} (Hartung, Knapp, and Sinha 2008, Chapters 5–8). A different example, the one studied by DeMissie (2009) following Yamaguchi and Ohashi (1999), concerned logarithms of survival times assumed to follow a Weibull distribution with the logarithm scale-parameter following a linear fixed-effects model in terms of covariates along with normally distributed mean-0 'frailty' terms; but in such an example, the reciprocal of the Weibull shape parameter acts as an unknown scale parameter for ε_{ij}, which otherwise has the known distribution with extreme-value survival function $\exp(-e^x)$.

Fixed-effect treatment-by-covariate interactions are a topic of considerable interest in studies of the 'treatment of choice' problem (Byar 1985, Gail and Simon 1985, Schemper 1988, Russek-Cohen and Gail 1993, and many others), but it is not very common to validate them strongly in clinical studies. This is because most clinical studies are powered primarily to detect treatment effects designed to answer direct clinical questions. Moreover, when results of many small to moderate-sized separate trials are published on the same treatment for the same disease, the separate studies may have insufficient power to answer even the direct clinical questions definitively — a deficiency that is often the strongest motivation for meta-analysis — so that the possibility of treatment by covariate interactions can hardly be addressed at the single-trial level. In the next Section of this paper, we consider first the case $\gamma = \rho_i = 0$ where all such interactions are absent. This is the setting most likely to support meta-analyses, due to the absence of first-order biases. In later sections, we address the impact of such interactions on the interpretability of meta-analyses.

In a meta-analytic setting, each of the separate studies i would be analyzed for treatment-effectiveness under the specialization of model (1.1) (ordinarily with $\gamma = \mathbf{0}$) to responses with fixed study-index i, i.e., with respect to

$$Y_{ij} = \mu_i + \vartheta_i \xi_{ij} + X'_{ij}\beta^{(i)} + \varepsilon_{ij} \quad , \qquad j = 1, \ldots n_i \tag{1.2}$$

and the estimated value $\tilde{\vartheta}_i$ for ϑ_i would be published along with its standard error $\tilde{\sigma}_i$. The parameter ϑ_i can be viewed as combining fixed and random-effects $\vartheta + v_i$ just as μ_i corresponds to $\mu + u_i$. Then the published point estimators $\tilde{\vartheta}_i$ would be analyzed (DerSimonian and Laird 1986; Normand 1999) as data within the (random-effect) Meta Analysis Model,

$$\tilde{\vartheta}_i \; = \; \vartheta + v_i^* + e_i \tag{1.3}$$

where $\tilde{\sigma}_i$ is treated as though known and non-random, where $v_i^* \sim \mathcal{N}(0, \tilde{\sigma}_i^2)$ and the independent error-term e_i has mean 0 and unknown scale-parameter σ_e^2. The point and interval estimators of the unknown common treatment-effect parameter ϑ are then used to draw combined-study conclusions about treatment effectiveness.

The present paper is organized as follows. The next Section collects the available theoretical results, including new ones, concerning the consistency of meta-analyses with the true parameters under a properly specified unified model of the type (1.1). The theme of that Section is that in the linear-model context without treatment-by-covariate interactions, at least in the absence of censoring, meta-analyses are generally highly reliable. Then Section 3 shows, by theoretical calculation and parametric examples, that treatment-by-covariate interaction effects which differ among study centers can ruin this reliable performance. Several simulations, following the general plan of DeMissie (2009), illustrate these points – both those favorable and those unfavorable to meta-analysis, using parameters chosen from the multi-center ECOG EST 1582 lung cancer clinical trial analyzed in Gray (1994).

2 Consistency

Suppose that the distribution of the model errors ε_{ij} in (1.1) is arbitrary, subject to the usual regularity conditions for maximum likelihood (ML) estimation of location and scale parameters (van der Vaart 1998, pp. 65, 68, 95). Then the regularity conditions are also satisfied for the full model (1.1) and also for the separate center-level models (1.2) when the latter are properly specified in the sense that $\gamma \equiv \rho_i \equiv 0$ in (1.1), and the random effects $u_i, v_i, b_i,$ are Gaussian with mean 0. One might estimate parameters including ϑ using least-squares instead of ML estimation, but in more complicated settings, e.g., those with censoring, estimating-equation or ML estimation methods will lead to consistent and asymptotically normal estimates with $\hat{\vartheta}_i - \vartheta_i = O_P(1/\sqrt{n_i})$, where all center sample-sizes n_i are assumed large, and the asymptotic standard deviations of $\sqrt{n_i}(\hat{\vartheta}_i - \vartheta_i)$ are estimated consistently within each study.

In the setting of the previous paragraph, the meta-analysis model (1.3) is properly specified. If also the number of centers gets large, but at a rate of smaller order of magnitude than (nearly all of) the center sample sizes, the meta-analysis leads to consistent estimation of ϑ. In other words, the two-stage estimation within models (1.2) and (1.3) leads asymptotically to results consistent with, but not identical to, those that would be produced by analysis of the unified patient-level model (1.1).

2.1 A rectangular-array formulation of meta-analysis

Recent literature has reached near consensus on the need for a random-effects formulation of meta-analysis. However, meta-analysis is an area like many others where one might also formulate cluster (here, study-center) differences as a large set of nuisance parameters, whose collective dimension grows proportionately to the number m of clusters. If we view the design variables X_{ij}, Z_{ij} as being randomly generated, with (u_i, v_i, ρ_i) **non**random, then the model (1.1) falls within the **rectangular-array** parametric structure

$$(Y_{ij}, X_{ij}, Z_{ij}, \xi_{ij}) \overset{\text{indep.}}{\sim} f(y \,|\, (x,z,\xi); \psi, \lambda_i)\, g_i(x,z)\, h(\xi) \tag{2.1}$$

where (g_i, n_i) are chosen non-identically but with large-scale stability (as though *iid* from some population of study-sizes and predictor-variable densities) as $i = 1, \ldots, m,$ and

$$\psi = (\mu, \vartheta, \beta, \gamma)\ , \quad \lambda_i = (u_i, v_i, b_i, \rho_i)$$

This definition slightly generalizes the terminology of Li, Lindsay and Waterman (2003), who did not allow non-*iid* regression terms.

The 'rectangular array asymptotics' of Li et al. (2003), show (in their univariate-parameter case, with all $n_i = n$ equal) essentially that the ML parameter estimates of ψ are consistent when m, n both get large in such a way that $m = O(n),$ but that the ML estimates are not efficient when m and n are of the same order. In the context of meta-analysis, the consistency result is of primary interest. Unfortunately, in real biostatistical meta-analyses, the center or study sample sizes may be comparable to or even smaller than the number of centers. Thus, to keep the nuisance parameter dimension from growing prohibitively fast, meta-analyses must rely on random-effect formulations of center variation among model coefficients, even though the random-effect distributions may be completely conjectural.

2.2 Meta- vs. Unified analysis in moderate samples

There are not many theoretical results directly comparing the meta-analytic estimates obtained from (1.2) and (1.3) with those obtained from (1.1). One very clean and simple result was obtained by Olkin and Sampson (1998), in a special case (although they allowed more than two treatments). Consider the model (1.1) without regression terms ($\beta = b_i = \mathbf{0}$, $\gamma = \rho_i = 0$), under the further restriction that $v_i \equiv 0$ (i.e., $\sigma_v = 0$), **and with the terms u_i treated as fixed effects**. Then the least-squares estimator of ϑ_i (which is also the ML estimator if ε_{ij} are *iid* normally distributed), whether ξ_{ij} are fixed or random, is

$$\tilde{\vartheta}_i = \bar{Y}_{i\cdot}^{(t)} - \bar{Y}_{i\cdot}^{(c)}, \quad \text{where} \quad \bar{Y}_{i\cdot}^{(t)} = \frac{\sum_{j=1}^{n_i} \xi_{ij} Y_{ij}}{\sum_{j=1}^{n_i} \xi_{ij}}, \quad \bar{Y}_{i\cdot}^{(c)} = \frac{\sum_{j=1}^{n_i} (1-\xi_{ij}) Y_{ij}}{\sum_{j=1}^{n_i} (1-\xi_{ij})}$$

Now, under the further restriction that the variances of these estimators $\tilde{\vartheta}_i$ are all assumed to be proportional to the known (i.e., reported) numbers

$$\Sigma_i \equiv 1/\sum_{j=1}^{n_i} \xi_{ij} + 1/\sum_{j=1}^{n_i} (1-\xi_{ij})$$

the meta-analysis least-squared estimator (also the MLE when the errors in (1.3) are normal) is

$$\hat{\theta} = \sum_{i=1}^{m} (\tilde{\vartheta}_i/\Sigma_i) \Big/ \sum_{i=1}^{m} (1/\Sigma_i) \tag{2.2}$$

Olkin and Sampson (1998) show by applying the Gauss-Markov theorem that the least-squares estimator of ϑ in (1.1) coincides exactly with $\hat{\vartheta}$ defined in (2.2). This is not obvious due to the random effect v_i remaining in the unbalanced ANOVA model (1.1), but does follow because both $\hat{\vartheta}$ and the ML estimator are unbiased linear combinations of the terms $\bar{Y}_{i\cdot}^{(t)}, \bar{Y}_{i\cdot}^{(c)}$, and their variances can be calculated to be identical to $\sigma^2/\sum_{i=1}^{m} (1/\Sigma_i)$.

This exact finite-sample result does not seem to generalize to models involving random center-effects u_i and especially to models with nonzero random treatment-effects v_i.

2.3 First-order unbiasedness of ML estimates in (1.2)

Since study-center sample sizes n_i are often rather small (of the order of dozens rather than hundreds), it becomes important to know whether the moderate-sample biases of Maximum Likelihood Estimators of ϑ_i in linear models (1.2) could cause noticeable biases in meta-analysis estimates (1.3). In this subsection, we present theoretical calculations to the effect that this will not be a problem in equal-allocation studies ($p = E(\xi_{ij}) = 1/2$) with properly specified models ($\gamma = \rho_i = 0$) in the absence of censoring.

The special feature of the linear models studied here is that we are interested in the coefficient of a single predictor ξ_{ij} which by design is independent of all others. Our primary tool in analyzing bias is the asymptotic expansion via Taylor series of the log-likelihood to higher order than is done in standard ML theory. This has been done rigorously in a one-dimensional parameter setting by Pfanzagl (1973a), with some further regularity conditions and relevant results (but not a convenient formula) in Pfanzagl (1973b). The necessary extensions for multivariate parameters are sketched below. The appropriate regularity conditions require smoothness to third and fourth order of the density f_e of the errors ε_{ij}, with boundedness on compact sets of expressions like

$\int \sup_{t:|t|\le K} |(d^k/dt^k)\log f_e(t)|\, f_e(t)\, dt$ for finite K and $1 \le k \le 4$. In addition, a fully rigorous justification requires assumptions on the predictors (X_{ij}, Z_{ij}) so that empirical averages of certain functions of them over i, j satisfy laws of large numbers.

The idea of the asymptotic expansion for bias of maximum likelihood estimators in a multivariate setting is as follows. Denote *iid* data-vectors by $\mathbf{W}_j$ for $j = 1, \ldots, n$, with r-dimensional parameter η, and log-likelihood $\mathrm{LLk}(\eta) = \sum_{j=1}^n \log f(\mathbf{W}_j, \eta)$. Under the assumption that the MLE $\hat\eta$ is consistent for the true value η_0 in the interior of its parameter space, Taylor expansion of $\nabla \mathrm{LLk}(\hat\eta)$ around η_0 yields

$$-\nabla^{\otimes 2}\mathrm{LLk}(\eta_0)\,(\hat\eta - \eta_0) \approx \nabla \mathrm{LLk}(\eta_0) + \frac{1}{2}\left((\hat\eta-\eta_0)'\frac{\partial}{\partial \eta_k}\nabla^{\otimes 2}\mathrm{LLk}(\eta_0)\,(\hat\eta-\eta_0)\right)_{k=1}^r$$

up to a remainder of order $o_P(1)$ as n gets large. Here and throughout the paper, we use the notation $\nabla^{\otimes 2}$ to denote the Hessian matrix-valued operator of mixed partial derivatives, and for any vector $\mathbf{v}$, the notation $\mathbf{v}^{\otimes 2} \equiv \mathbf{v}\mathbf{v}'$. Substitute

$$U = \nabla \mathrm{LLk}(\eta_0)\,, \quad I = n^{-1} E(-\nabla^{\otimes 2}\mathrm{LLk}(\eta_0))\,, \qquad R = \nabla^{\otimes 2}\mathrm{LLk}(\eta_0) + nI$$

into this expansion, along with $\hat\eta - \eta_0 = (nI)^{-1}U + o_P(n^{-1/2})$, to find

$$\begin{aligned}\hat\eta - \eta_0 &= (nI)^{-1}U + (nI)^{-1}R\,(nI)^{-1}U \\ &\quad + \frac{1}{2}\left((nI)^{-1}U'(nI)^{-1}\nabla^{\otimes 2}\frac{\partial}{\partial\eta_k}\mathrm{LLk}(\eta_0)(nI)^{-1}U\right)_{k=1}^r + o_P(n^{-1})\end{aligned}$$

Since $E(U) = 0$, the quantity $\hat\eta - \eta_0$ differs by $o_P(n^{-1})$ from a random variable with expectation

$$n^{-1}E_{\eta_0}\Big\{ I^{-1}\,\nabla^{\otimes 2}\log f(\mathbf{W},\eta_0)\, I^{-1}\nabla \log f(\mathbf{W},\eta_0) \tag{2.3}$$

$$+\frac{1}{2}I^{-1}\left((\nabla \log f(\mathbf{W},\eta_0))' I^{-1} E\Big[\frac{\partial}{\partial\eta_k}\nabla^{\otimes 2}\log f(\mathbf{W}_*,\eta_0)\Big] I^{-1}\nabla \log f(\mathbf{W},\eta_0)\right)_{k=1}^r\Big\}$$

where $\mathbf{W}_*, \mathbf{W}$ are independent and identically distributed.

We now apply the formula for bias to the ϑ component of the MLE based on the model (1.3) for a single study. For simplicity, in this subsection let ξ_{ij} take values $q, -p$ with respective probabilities p, q, where $0 < p < 1$ and $q = 1 - p$. (The mechanism of treatment allocation is known to the statistician, and usually $p = 1/2$.) Also for simplicity, let the unknown parameter be $\eta = (\vartheta, \beta, \mu, \sigma^2)$, a vector of dimension $r = d + 3$. Then the parameter of main interest, the treatment effect ϑ, has the secondary notation η_1. The density of ε_{ij} has been assumed known up to the scale parameter, so without loss of generality let ε_{ij} be centered and scaled so that it has mean 0 and scale-parameter σ, i.e., let the density be $\sigma^{-1}g(s/\sigma)$, where g a known density with mean 0 (or else median 0). Denote

$$i_g \equiv \int \frac{(g'(s))^2}{g(s)}\,ds\,, \quad a_g = \int s^2\,\frac{(g'(s))^2}{g(s)}\,ds - 1\,, \quad m_g = \int s\,\frac{(g'(s))^2}{g(s)}\,ds \tag{2.4}$$

where the a_g integral is assumed finite, so that the others are also finite.

For the rest of this subsection, we suppress the study-center index i, denote $\mathbf{W}_j = (Y_j, X_j, \xi_j)$ and $\varepsilon_j = Y_j - \mu - \vartheta_j\xi_j - X_j'\beta$, and assume that the vectors $X_{ij} = X_j$ are random and *iid* with mean 0, with $E(X_j^{\otimes 2}) \equiv \Sigma_X$. Then the density $f(\cdot,\eta)$ for $\mathbf{W}$ is $\sigma^{-1}f_e(\varepsilon_j/\sigma,\eta)$, and it is straightforward to check that

$$\nabla \log f(\mathbf{W},\eta) = \frac{1}{\sigma}\Big\{ -\frac{g'(\varepsilon/\sigma)}{g(\varepsilon/\sigma)} \begin{pmatrix} \xi \\ X \\ 1 \\ \varepsilon/\sigma \end{pmatrix} - \begin{pmatrix} 0 \\ \mathbf{0} \\ 0 \\ 1 \end{pmatrix} \Big\}\ , \quad I = \frac{1}{\sigma^2}\begin{pmatrix} pqi_g & \mathbf{0}' \\ \mathbf{0} & B \end{pmatrix}$$

where I is $r \times r$, $r = d+3$, and the $(d+2)\times(d+2)$ matrix B is given by

$$B = \begin{pmatrix} \Sigma_X & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & i_g & m_g \\ \mathbf{0}' & m_g & a_g \end{pmatrix}$$

Now we successively evaluate the first component for the two expectation terms in curly brackets in (2.3). Because of the block-diagonal form of I given above,

$$E\Big(I^{-1}\nabla^{\otimes 2}\log f(\mathbf{W},\eta_0) I^{-1}\nabla \log f(\mathbf{W},\eta_0)\Big)_1$$

$$= \frac{\sigma^2}{pqi_g} E\Big((\frac{\partial}{\partial \eta_1}\nabla \log f(\mathbf{W},\eta_0))' I^{-1}\nabla \log f(\mathbf{W},\eta_0)\Big)$$

Substituting the form found above for $\nabla \log f(\mathbf{W},\eta)$ along with its partial derivative with respect to ϑ, we find the last expression equal to

$$E\Big\{ \frac{-\xi}{\sigma pqi_g}\Big((\log g)''(\frac{\varepsilon}{\sigma}) \begin{pmatrix} \xi \\ X \\ 1 \\ \varepsilon/\sigma \end{pmatrix}' + \frac{g'(\varepsilon/\sigma)}{g(\varepsilon/\sigma)} \begin{pmatrix} 0 \\ \mathbf{0} \\ 0 \\ 1 \end{pmatrix}' \Big) I^{-1} \Big(\frac{g'(\varepsilon/\sigma)}{g(\varepsilon/\sigma)} \begin{pmatrix} \xi \\ X \\ 1 \\ \varepsilon/\sigma \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{0} \\ 0 \\ 1 \end{pmatrix} \Big) \Big\}\ .$$

Finally, using the block-diagonal form of I to eliminate the terms involving ξ^2, and recalling that ξ has mean 0 and is independent of $(X,\mathbf{W})$ to eliminate the terms linear in ξ, we conclude

$$E\Big(I^{-1}\nabla^{\otimes 2}\log f(\mathbf{W},\eta_0) I^{-1}\nabla \log f(\mathbf{W},\eta_0)\Big)_1 = \frac{-\sigma E(\xi^3)}{(pqi_g)^2}\int \Big\{ \frac{g''(s)}{g(s)} - \frac{g'(s)^2}{g^2(s)} \Big\} g'(s)\, ds\ .$$

where we have also made the change of variable $s = \varepsilon/\sigma$ in the integrals over the density $\sigma^{-1} g(\varepsilon/\sigma)$ for ε. Similarly, the first component of the second term in (2.3) is calculated as

$$E\Big(I^{-1}\Big((\nabla \log f(\mathbf{W},\eta_0))' I^{-1} E\Big[\frac{\partial}{\partial \eta_k}\nabla^{\otimes 2}\log f(\mathbf{W}_*,\eta_0)\Big] I^{-1}\nabla \log f(\mathbf{W},\eta_0)\Big)_{k=1}^{r}\Big)_1$$

$$= \frac{\sigma^2}{pqi_g} E\Big((\nabla \log f(\mathbf{W},\eta_0))' I^{-1} E\Big[\frac{\partial}{\partial \eta_1}\nabla^{\otimes 2}\log f(\mathbf{W}_*,\eta_0)\Big] I^{-1}\nabla \log f(\mathbf{W},\eta_0)\Big)\ . \qquad (2.5)$$

This expression is reduced further, again by means of the block-diagonal form of I together with the independence of ξ from (X,ε), resulting in

$$I^{-1} E\Big[\frac{\partial}{\partial \eta_1}\nabla^{\otimes 2}\log f(W_*,\eta_0)\Big] I^{-1} = \frac{-E(\xi^3)}{\sigma^3}\int (\log g)'''(s)\, g(s)\, ds \begin{pmatrix} 1 \\ \mathbf{0} \\ 0 \\ 0 \end{pmatrix}^{\otimes 2} \Big(\frac{\sigma^2}{pqi_g}\Big)^2\ .$$

Then expression (2.5) reduces to $-\sigma E(\xi^3)\,(\int (\log g)'''(s)\, g(s)\, ds)/(pqi_g)^2$, so overall the asymptotic bias for $\hat{\vartheta} - \vartheta_0$ given in expression (2.3) has been shown equal to

$$\text{Order } \frac{1}{n} \text{ Bias of } \hat{\vartheta} = \frac{-\sigma E(\xi^3)}{n(pqi_g)^2}\int \Big(\Big\{ \frac{g''(s)}{g(s)} - \frac{g'^2(s)}{g^2(s)} \Big\} g'(s) + \frac{1}{2}(\log g)'''(s)\, g(s) \Big)\, ds\ . \qquad (2.6)$$

Two applications of this result deserve special mention. First, in the case of normal errors ($g(s) = (2\pi)^{-1/2} e^{-s^2/2}$), the order $1/n$ bias is exactly 0. Second, in the case of extreme-value errors ($g(s) = \exp(s - c - e^{s-c})$, with $c = -0.5772$ defined so that g has expectation 0), $i_g = 1$ and the order $1/n$ bias is $-E(\xi^3)\sigma/(2n(pq)^2) = (p-q)\,\sigma/(2npq)$. For all g, the top-order bias is 0 when $p = q = 1/2$, which will be the case in most randomized clinical trials. Simulation examples related to the biases studied in this subsection will be given below, in a case with $p = 1/3$ such as might arise in practice in a case-control study with two controls matched to each case.

Although no further details are provided here, the same result (2.6) can be proved to hold also when there are treatment-by-covariate interactions with coefficients γ allowed in the model (1.2). The impact of this result, either as given above or in this extended form, is that balanced ($p = 1/2$) linear center-level models contribute biases of order no higher than $1/n_i^2$ in the meta-analysis model (1.3) whenever the models (1.2), or even the extended versions with a treatment-by-covariate interaction term, are properly specified. For this reason, we must look to the possibility of omitted treatment-by-covariate-by-center interactions to understand whether there might actually be meta-analysis biases in the linear-model context. If so, one might expect the interactions and therefore the biases to lie in the same direction for multiple centers.

3 Inconsistency

Since individual studies for which results are combined through meta-analysis are often small, they generally ignore interaction terms between treatment and covariates. This might happen because the relevant covariates enter only weakly into the predictive linear model (1.2) for survival in each center, and the covariate Z_{ij} itself is omitted from the fitted model. But even if Z_{ij} enters (1.2) among the predictor variables X_{ij}, the interaction term — the one with coefficient $\gamma_i \equiv \gamma + \rho_i$ in (1.1) — will usually not be strong enough to warrant inclusion in the center-level model. Yet the treatment-by-covariate interaction coefficients with fixed and random-effect coefficients γ_i could easily be strong enough for inclusion in the unified analysis (1.1). That is, the pooled-data model might well estimate coefficients γ_i not quite detectable in the models (1.2).

This phenomenon was seen to occur in the combined model-fitting of Gray (1994) in the multi-center clinical trial ECOG EST 1582 of two different chemotherapy regimens for treatment of small-cell lung cancer. Among the covariates were an indicator ('`bone`') of bone metastases, one of liver metastases, and one of weight-loss prior to study entry, as well as a measure of performance status at baseline. Gray (1994) found by a Bayesian proportional-hazards analysis that treatment differences were significant but that `treatment`-by-`bone` interactions were also. A re-analysis by DeMissie (2009) showed the same phenomenon by parametric Weibull survival regressions taking into account center-level random intercepts (terms u_i in (1.1)) and random treatment effects (terms v_i), finding slightly significant (random-effect) treatment by center and (fixed-effect) `treatment` by `bone` interactions.

In the clinical trial example of the previous paragraph, DeMissie (2009) determined that meta-analysis would have recovered essentially the same significant treatment effect (and p-value) found earlier by Gray (1994) and in his own analysis of model (1.1). However, had the treatment effect been weaker and the `treatment` by `bone` effect considerably stronger, meta-analysis could have resulted in a meaningful bias in estimated treatment effect, a bias which we now explore.

3.1 Limiting-ML bias via Kullback-Leibler minimization

Our method is to calculate numerically the closest model (1.2) in Kullback-Leibler sense to a specific center-level model (the 'true model') which includes an extra treatment-by-covariate term $\gamma_i \xi_{ij} Z_{ij}$. In such a 'true model', for a single center i, the treatment-effect felt by individual j is $\vartheta_i + \gamma_i Z_{ij}$. Recall that the Kullback-Leibler distance between a working model with density h and a true model with density f is $K(h,f) \equiv \int \log(f/h) f$, and that under fairly general regularity conditions (*cf.* van der Vaart 1998, pp. 44–47), the ML estimates for the parameters of a working parametric model-family $h \in \mathcal{H}$ converge in large data-samples to the parameter values defining the element h^* which minimizes $K(h,f)$ over $h \in \mathcal{H}$. In this Section, the working model is (1.2) in center i (with parameters $(\mu_i, \vartheta_i, \beta^{(i)}, \sigma_i)$), versus a true model in which

$$Y_{ij} = m_i + t_i \xi_{ij} + X'_{ij} b_i + \gamma_i \xi_{ij} Z_{ij} + \tilde{\varepsilon}_{ij} \tag{3.1}$$

with parameters $(m_i, t_i, b_i, \gamma_i, \tau_i)$, where $\tilde{\varepsilon}_{ij}$ is assumed to fall in the same centered scale family as ε_{ij} but has scale-parameter τ_i instead of σ_i.

To understand the large-sample biases in treatment effect induced by misspecifying the true model (3.1) as (1.2), we begin calculating the difference between the large-sample limit of the working-ML-estimated treatment-effect parameter ϑ_{*i} and t_i in a single center-group i, when Z_i consists only of a single column which appears also as a column X_{ij}. In this case, all of the bias $\vartheta_{*i} - t_i$ can be shown to depend on the true model parameters only through γ_i together with σ_i, in such a way that $(\vartheta_{*i} - t_i)/\sigma_i$ is a well-defined function of γ_i/σ_i. To see this, note that in the integrals

$$-\sum_{\xi=0}^{1} p^{\xi} (1-p)^{1-\xi} \int\int \log(h(y,X,\xi))\, f(y|X,\xi)\, dy\, f_X(X)\, dX \tag{3.2}$$

being minimized, $h(y,X,\xi) = h(y|X,\xi)\, f_X(x)\, p^{\xi} (1-p)^{1-\xi}$ (with p known), and

$$f(y|X,\xi) = \frac{1}{\tau_i} g((y - m_i - t_i\xi - b_i'X - \gamma_i \xi Z)/\tau_i)\ ,\ \ h(y|X,\xi) = \frac{1}{\sigma_i} g((y - \mu_i - \vartheta_i \xi - \beta_i' X)/\sigma_i).$$

The change of variable $s = (y - m_i - t_i\xi - b_i'X - \gamma_i\xi Z)/\tau_i$ in (3.2), which then can be minimized over $(\mu_i - m_i)/\sigma_i$, $(\vartheta_i - t_i)/\sigma_i$, $(\beta_i - b_i)/\sigma_i$, and τ_i/σ_i, shows that the minimizing standarized bias $(\vartheta_i - t_i)/\sigma_i$ depends on the parameters of model (3.1) only through γ_i/σ_i and p.

We gain numerical insight into the dependence of the KL-minimizing standardized bias $(\vartheta_i - t_i)/\sigma_i$, in a single study i for (3.1) misspecified as (1.2), as a function $a(\gamma_i/\sigma_i)$, through a series of numerical calculations in terms of the Extreme-Value density g, for several choices of 50×2 matrices X, with first column a fixed vector of $\mathcal{N}(0,1)$ deviates (the same vector throughut), and second column defined equal to Z, such that $\mathbf{1}'X = (0,0)$. For each choice of a column Z, simulated in each case from a specified distribution and then standardized as a vector of 50 components to have sample mean 0 and standard deviation 1, we display the KL-minimized biases by graphing the calculated pairs $(\gamma_i/\sigma_i,\ (\vartheta_i - t_i)/\sigma_i)$. With the first column of $\mathbf{X}$ fixed once and for all, we obtain a function $a(\gamma_i/\sigma_i,\ p)$ to graph for each choice of a distribution from which to simulate the column Z (the second column of $\mathbf{X}$) as an *iid* sample of size 50, after which Z is standardized.

The results are displayed in Figures 1 and 2, respectively for the two cases $p = 0.333$ and 0.5, for Z columns generated either as samples of size 50 from Beta(3,1), Beta(1,3), Beta(2,2), Beta(.2, .2), or as $\{0,1\}$ sequences of length 50 respectively containing 25, 20 ,15, or 10 ones. The calculations

generally show that the biases are larger when the distribution of Z column entries is more asymmetric, or when p is farther from $1/2$. For large values of γ_i/σ_i, the biases are nearly proportional to γ_i/σ_i, but when Z consists of equally many 1's and 0's, and also $p = 0.5$, the biases are exactly 0. In other cases where $p = 0.5$, the biases may be negative; but the magnitudes of biases are generally much smaller for $p = 0.5$ than for $p = .333$.

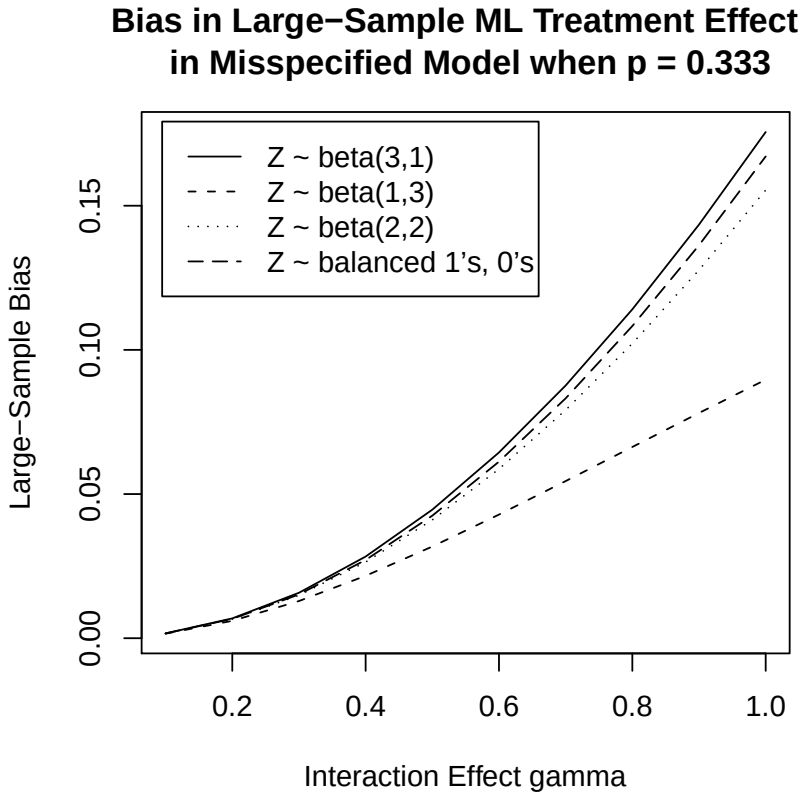

Fig. 1 Limiting large-sample ML biases as a function of interaction effect γ_i, when σ_i is set to 1, with p=0.333. Error density is centered (mean-0) extreme-value.

The results presented in these Figures, and additional calculations which are not shown, make it clear that the biases due to omitted treatment by covariate interactions cannot in general be ignored. There are certainly applied settings where biases in treatment-effect divided by the standard deviation σ_i as large as those seen in these Kullback-Leibler minimizations could seriously affect the interpretation of a meta-analysis. On the other hand, even if the error distribution is extreme-value, and especially in meta-analyses of studies with balanced allocation, the interaction-effects needed to attain biases of magnitude .03 or larger are .7 or larger for the various Z-columns that have been tried. In realistic meta-analyses, treatment-by-covariate interactions may often not be nearly that large. Moreover, with Gaussian error distributions, these misspecified ML biases can be shown to disappear completely (for model (3.1) misspecified as (1.2)), for all values p. Thus, in many meta-analyses with

nearly equal allocation, the biases studied here, in Sections 2.3 and 3, will be extremely small. We illustrate using data from the lung-cancer clinical trial mentioned in the Introduction.

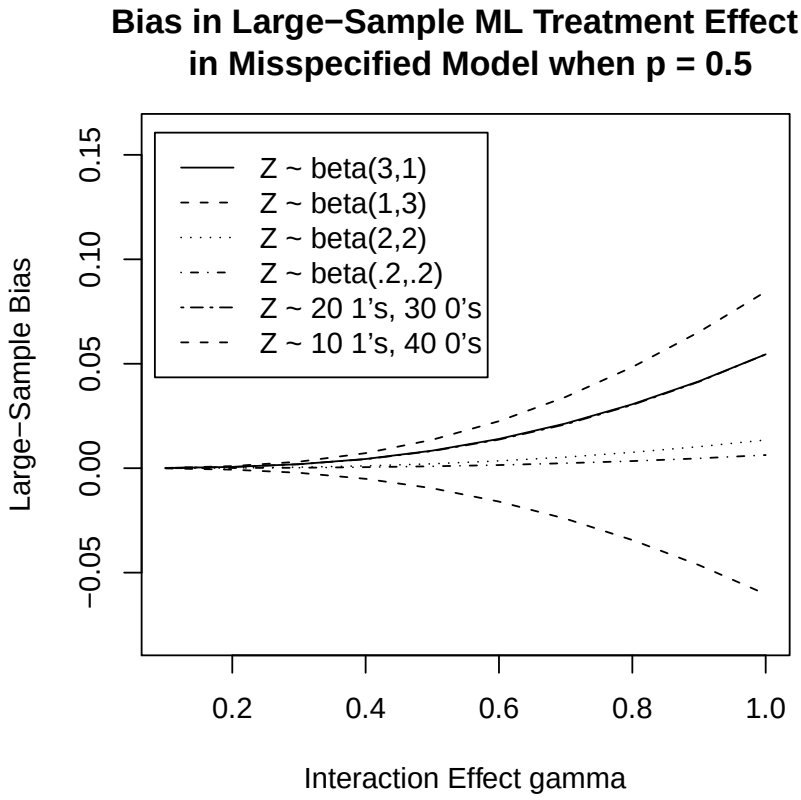

Fig. 2 Limiting large-sample ML biases as a function of interaction effect γ_i, when σ_i is set to 1, with p=0.5. Error density is centered (mean-0) extreme-value.

3.2 Numerical results based on real multi-center covariate data

We continue by applying the method of the previous subsection to assess the potential biases using the 570×4 design (covariate) matrix $(X_{ij}, 1 \le i \le 18, 1 \le j \le n_i)$ of four non-constant predictors `bone, liver, weight`, and `Perf` as in DeMissie (2009), which differed from the raw ECOG EST 1582 data only in that the original 26 centers were pooled in DeMissie (by a criterion involving similarity of the four covariates) into 18 clusters with minimum sample size > 15. DeMissie (2009) fitted a model of the form (3.1), without random effects for cluster or cluster-by-treatment, to the data Y_{ij} consisting of the log survival times in this multi-center lung cancer clinical trial, with error-term variables $\tilde{\varepsilon}_{ij}$ chosen to be extreme-value (log's of exponential) random variables corrected to have mean 0. He used $Z =$`bone` and $\xi_{ij} = \pm.5$, fitting a model (3.1) with parameters taken to be the same for all i, in the same spirit as Gray (1994). The parameters obtained by fitting the model in this way were:

$$m_0 = .033,\ b_0 = (-.18, -.23, .42, -.18),\ t_0 = .20,\ \gamma_0 = 0.22,\ \tau_0 = .8\ .$$

This was an equal-allocation clinical trial, with the proportion of ξ_{ij} indicators equal to 1 very close to $1/2$ within each center i. So after correcting the entries of X_{ij} to have mean 0 when averaged over $j = 1, \ldots, n_i$ for each fixed i, and re-expressing the model parameters in (3.1), we obtained model parameters which *did* vary with i.

We applied the calculations of Section 3 to the model just described, for each i. The result was that, with $\gamma_i \equiv 0.22,\ p = 0.5,$ the bias of the Kullback-Leibler minimizer of the misspecified model (1.2) parameters with respect to the true model with parameters just decribed was at most about 0.0001 in each cluster, and the biases increased only to a maximum of about 0.001 when γ_i was increased to 0.44. When this calculation was re-done with $\gamma_i = 0.44,\ p = 0.333,$ the largest cluster-level biases found in ϑ_i are of the order .01. Thus the parameter combinations including the actual fitted interaction effect γ would not have led to meaningful misspecified-model biases even if the treatment allocation fraction differed from 0.5, although the biases were not 0. (When $\gamma = .88$ and $p = .4,$ the largest biases range up to .026, around 8% of the corresponding cluster ϑ_i values.)

This calculation confirms that the potential biases in treatment-effect estimates when study-level treatment by covariate interactions are erroneously omitted may still in real linear-model examples (3.1) turn out to be negligibly small. Whether they do is a quantitative question relating to the magnitudes of covariate values, coefficients, and treatment-by-covariate interactions. Note that the Kullback-Leibler minimization calculations presented here relate only to *linear* models: and it remains to be seen whether omitted treatment by covariate interactions in nonlinear or generalized-linear models more easily result in important biases with realistic covariate design matrices.

4 Simulation evidence

In this section, we examine the biases which arise in center-level analyses using model (1.2) in a simulation of a patient-level clinical trial with several different treatment scenarios. We simulated the true model (3.1) with the same fixed covariate design matrix from the ECOG EST 1582 trial, as in the previous subsection, with the same parameter settings and cluster structure (570 patients in 18 hospital 'centers', some of which were combined from smaller centers).

We consider the biases in center-level estimates $\hat{\vartheta}_i$ under model (1.2) with design matrix of covariates X_{ij} **not** corrected to have center average 0, and with errors following the adjusted extreme-value density $g(w) = \exp((w - c - e^{w-c})/\sigma_e)$ where $c = -0.5772$ is defined so that g has expectation 0. Our simulations are intended to compare biases with equal random treatment allocation ($p = .5$) versus those with random allocation of two control patients to each patient receiving the experimental treatment ($p = 1/3$). The treatment variable in either case has the form as in Section 1.2 that $\xi_i = \xi_i^o - p$ where $\xi_i^o \sim \text{Binom}(1, p)$, and we consider separately the case where the treatment-allocation indicators are fixed for an entire simulation or are re-simulated in each replication of the trial.

The first set of simulations, each of 1000 replications of 570 patients' log survival times Y_{ij}, is summarized in columns 4–6 of Table 1. For these simulations, there is no variable Z_{ij} (i.e., $\gamma = \sigma_\rho = 0$). The parameters in model (1.1) — apart from the treated fraction p which is taken either equal to $1/2$ or $1/3$ — are fixed at

$$\mu = .033,\ \vartheta = 0,\ \beta = (-.18, -.23, .42, -.18),\ \sigma_u = .5,\ \sigma_v = .44,\ \sigma_e = 0.8\ .$$

In the first simulation (with empirical biases given in column 4), $p = 1/2,$ and in the second and third (col.'s 5 and 6), $p = 1/3$. In the first and second, treatment allocation was simulated only once for each

of the first and second blocks of 500 replications, for all 570 patients, while in the third simulation, treatment allocations were generated independently in each of the 1000 simulation replications. The averaged results of the center-level treatment-effect estimators $\hat{\vartheta}_i,\ 1 \le i \le 18,$ are displayed for all three simulations in Table 1. The displayed averages in columns 4–6 are biases, since the true $\vartheta = 0$. The average standard errors for these estimators (not shown), were all approximately 0.02. Finally, col. 7 contains the expected top-order bias (the expectation of column 6) found in Sec. 3 to be $2(p-q)\sigma_e/(n_i pq)$ which for $p = 1/3$ and $\sigma_e = 0.8$ is equal to $-0.6/n_i$, where n_i is the center sample size displayed in column 2. Inspection of the Table shows that there is almost no bias in the $\hat{\vartheta}_i$ estimates when $p = 1/2$. Those few cases (especially centers 7 and 14) which have average estimated biases too large to be due to chance indicate particularities of the fixed treatment-allocations simulated only once for the first and last 500 replications, and are not incompatible with expected bias of 0 when averaged over random ξ_i. Similar comments, about the different biases seen in column 5 versus 6 due to the treatment-allocation variates ξ_i fixed over blocks of 500 replications in col. 5 versus ξ_i re-simulated in each replication for col. 6. The theoretical expected top-order biases (col. 7) do correlate highly with but are not very close to the empirical numbers in column 6 for $p = 1/3$. This is not fully explained, although the moderate sample sizes mean that great reliance cannot be placed on asymptotic results.

Table 1 Biases of center treatment effects ϑ_i. Col.'s 4–6: biases in 3 simulations of 1000 replications each with $\gamma_i = 0$; col. 8 biases from simulation with $\gamma_i = .22$. Col. 7 is the theoretical top-order expectation of Col. 7 from Sec. 3. Parameter settings and structural details given in text.

			$\gamma = 0$				$\gamma = .22$
Center		info	trt fixed		trt random	Theor.	trt fixed
Ctr	Size	Bonefrac	$p = 0.5$	$p = 1/3$	$p = 1/3$	Bias$(1/3)$	$p = 0.5$
1	21	.095	.0227	-.0445	-.0308	-.0286	-.004
2	17	.000	.0054	-.0002	-.0529	-.0353	.029
3	18	.444	.0213	-.0635	-.0229	-.0333	.002
4	27	.296	.0055	-.0567	-.0411	-.0222	.003
5	46	.283	-.0287	.0012	-.0017	-.0130	-.018
6	31	.355	.0000	-.0459	.0022	-.0194	-.010
7	17	.118	.0579	-.0308	-.0464	-.0353	-.037
8	59	.339	-.0034	-.0126	-.0158	-.0102	-.013
9	56	.268	.0056	-.0281	.0058	-.0107	-.011
10	31	.290	-.0181	-.0217	-.0155	-.0194	-.014
11	22	.273	.0150	-.0477	-.0354	-.0273	-.011
12	39	.282	-.0020	-.0048	-.0063	-.0154	-.001
13	27	.333	-.0115	-.0760	-.0198	-.0222	-.013
14	53	.585	.0120	-.0083	-.0304	-.0113	.006
15	17	.176	.0275	-.0256	-.0659	-.0353	.047
16	42	.238	.0435	.0067	-.0031	-.0143	-.012
17	23	.391	.0063	-.0330	-.0093	-.0261	.016
18	24	.417	-.0150	-.0521	-.0100	-.0250	-.012

A further batch of 1000 simulation-replications was done, with $p = 0.5$ and all covariates and parameters fixed as in the previous simulations (including $\sigma_\rho = 0$ in (1.1)) *except* that now $\vartheta_i = .22$ and the Z_{ij} column (the `bone` covariate) enters in (3.1) through the treatment-by-covariate interaction-term with nonzero coefficient $\gamma_i = 0.22$ (for all i). Recalling that X_{ij} and Z_{ij} used in these simulations

were not adjusted to have average 0 within each center, we find that the actual treatment effects ϑ_i experienced under model (3.1) in centers i was actually $0.22 + b' \sum_{j=1}^{n_i} Z_{ij}/n_i$.

Averaging the estimates $\hat{\vartheta}_i$ calculated under the misspecified model (1.2) across simulations, and subtracting these true treatment effects, yields the biases recorded in the 8'th column of Table 1). These biases are almost all remarkably small, according closely with the findings of Demissie (2009). Although the specific biases found in this simulation cannot be viewed as an expectation over random ξ_{ij}, since only two treatment-allocations were simulated for blocks of 500 simulation replications, these small biases are as anticipated from the findings in Section 3.2.

In both sets of simulations, care was needed to avoid nonconvergent ML estimates because of the smallness of most center sample-sizes. Occasional large estimates were re-checked with alternative initial values suppled to the `R` function `survreg`. All computations in the paper were done in the `R` (2009) statistical computing package.

5 Conclusions

After a brief review of previous work, this paper has studied two ways for biases to enter into meta-analyses of multi-center studies under linear models with additive random errors and random effects. The first, arising from $O(1/n_i)$ biases of ML estimators, causes bias only in models with non-Gaussian errors and unequal treatment allocations, but theoretical calculations and simulations (given in Table 1) show that meaningfully large biases can enter with realistic Weibull regression models due to small center sample sizes when the experimental-treatment fraction is $1/3$. The second mechanism of bias studied is due to misspecified models, which can arise because small separate studies will generally ignore moderate treatment-by-covariate interactions. Theoretical and simulation evidence in this paper shows that this source of bias will be uncommon, and the resulting biases will be small in linear models unless the omitted interaction effects are large. Future research should aim to extend to nonlinear and generalized-linear and censored-data studies this exploration of potential biases from small-sample ML estimates and from misspecification by omitting treatment-by-covariate interactions.

References

[1] Byar, D. (1985) Assessing apparent treatment-covariate interactions in randomized clinical trials. *Statist. in Medicine* **4**, 255-263.

[2] DeMissie, M. (2009), Investigating center effects in a multi-center clinical trial study using a parametric proportional hazards meta-analysis model, Univ. Maryland MA thesis, Statistics Program.

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

[4] Fay, R. and Herriot, R. (1979), Estimates of income for small places: an application of James-Stein procedures to census data. *Jour. Amer. Statist. Assoc.* **74**, 341-53.

[5] Gail, M. and Simon, R. (1985), Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41**, 361-372.

[6] Gray, R. (1994), A Bayesian analysis of institutional effects in a multicenter cancer clinical trial, *Biometrics* **50**, 244-253.

[7] Hartung, J., Knapp, G., and B. Sinha (2008), **Statistical Meta-Analysis with Applications**, Wiley, New York.

[8] Jiao, L., de Gonzalez, A., Hartge, P., Pfeiffer, R. et al. (2010), Body mass index, effect modifiers, and risk of pancreatic cancer – a pooled study of seven prospective cohorts, *Cancer Causes Control*, **21**, 1305-1314.

[9] Klein, J. and Moeschberger, M. (2003), **Survival Analysis, 2nd ed.** New York: Springer.

[10] Li, H., Lindsay, B. and Waterman, R. (2003) Efficiency of projected score methods in rectangular array asymptotics. *Jour. Roy. Statist. Soc.* Ser. B **65**, 191-208.

[11] Murphy, G., Pfeiffer, R., Constanza Camargo, A. and Rabkin, C. (2009), Meta-analysis shows that prevalence of Epstein-Barr Virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology*, **137**, 824-833.

[12] Normand, S-L. (1999), Tutorial in biostatistics meta-analysis: Formulating, evaluating, combining, and reporting, *Statistics in Medicine*, **18**, 321-359.

[13] Olkin, I. and Sampson, A. (1998), Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* **54**, 347-352.

[14] Pfanzagl, J. (1973a), Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1**, 993-1025.

[15] Pfanzagl, J. (1973b), The accuracy of the normal approximation for estimates of vector parameters. *Zeitschr. f. Wahrsch. u.verw. Geb.* **25**, 171-198.

[16] R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

[17] Russek-Cohen, E. and Simon, R. (1993), Qualitative interactions in multifactor studies. *Biometrics* **49**, 467-477.

[18] Schemper, M. (1988), Non-parametric analysis of treatment-covariate interaction in the presence of censoring. *Statist. in Medicine* **7**, 1257-1266.

[19] Van der Vaart, A. (1998) **Asymptotic Statistics.** Cambridge: Cambridge Univ. Press.

[20] Yamaguchi, T. and Ohashi Y. (1999), Investigating center effects in a multi-center clinical trial of superficial bladder cancer. *Statist. in Medicine* **18**, 1961-1971.