

## Article

# A Meta-Analysis for Simultaneously Estimating Individual Means with Shrinkage, Isotonic Regression and Pretests

Nanami Taketomi <sup>1</sup>, Yoshihiko Konno <sup>2</sup>, Yuan-Tsung Chang <sup>3</sup> and Takeshi Emura <sup>1,\*</sup><sup>1</sup> Biostatistics Center, Kurume University, Kurume, Fukuoka 830-0011, Japan; nnmamikrn@gmail.com<sup>2</sup> Department of Mathematical and Physical Sciences, Japan Women's University, Tokyo 112-8681, Japan; konno@fc.jwu.ac.jp<sup>3</sup> Department of Social Information, Mejiro University, Tokyo 161-8539, Japan; chogenso@gmail.com

\* Correspondence: takeshiemura@gmail.com

**Abstract:** Meta-analyses combine the estimators of individual means to estimate the common mean of a population. However, the common mean could be undefined or uninformative in some scenarios where individual means are “ordered” or “sparse”. Hence, assessments of individual means become relevant, rather than the common mean. In this article, we propose simultaneous estimation of individual means using the James–Stein shrinkage estimators, which improve upon individual studies' estimators. We also propose isotonic regression estimators for ordered means, and pretest estimators for sparse means. We provide theoretical explanations and simulation results demonstrating the superiority of the proposed estimators over the individual studies' estimators. The proposed methods are illustrated by two datasets: one comes from gastric cancer patients and the other from COVID-19 patients.

**Keywords:** statistical decision theory; isotonic regression; meta-analysis; pretest estimator; restricted parameters; shrinkage estimation

**Citation:** Taketomi, N.; Konno, Y.; Chang, Y.T.; Emura, T. A Meta-Analysis for Simultaneously Estimating Individual Means with Shrinkage, Isotonic Regression and Pretests. *Axioms* **2021**, *10*, 267. <https://doi.org/10.3390/axioms10040267>

Academic Editor:  
Hari Mohan Srivastava

Received: 23 August 2021  
Accepted: 12 October 2021  
Published: 20 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Meta-analysis is a statistical method used to summarize results from published studies [1]. While it has been applied to all areas of science, meta-analysis has been a particularly important tool in educational studies [2,3] and medical studies [4,5]. Especially, meta-analysis has played an important role in studies on the impact of COVID-19 [6–8].

The most basic method in meta-analyses uses a fixed-effect model, where all of the studies have a common mean [5,9]. This model has long been recognized in mathematical statistics and stratified sampling designs, where the common mean estimator has rigorously been investigated (pp. 55–103 of [10]; [11–13]).

Another basic model is the random-effects model [14], where the studies have different means. This model imposes normally distributed random-effects to account for between-study deviations from the common mean [15]. In either model, the goal of meta-analyses is to estimate the common mean by combining the estimators of individual means.

However, in some scenarios, the common mean is undefined or uninformative. In these scenarios, the focus of meta-analyses can be on individual studies' means. While individual means are estimated by published studies, they could be *improved* by statistical techniques. For instance, Raudenbush and Bryk [16] pointed out the possibility for improving individual studies' estimators in meta-analyses by using empirical Bayes estimators under the normal random-effects. Schmid [17] also considered similar Bayes estimators that improve upon individual studies' estimators. Röver and Friede [18,19] tried to improve an individual estimator by another individual estimator. All these Bayes estimators are defined by posterior means under the normal priors (random-effects). Evidently, the posterior means are related to but different from the individual studies' estimates. However, there are some cases where

the normal priors (random-effects) do not apply (Section 2.2). This motivates us to develop non-Bayesian estimation.

In this article, we consider some improved estimation methods of individual means without normal random-effects. First, we will explain how to employ *shrinkage estimators* to obtain improved estimators of individual means. Unlike the existing Bayes estimators, our frequentist shrinkage estimators do not involve any prior distribution. Next, we will provide theoretical and numerical results to show the superiority of the proposed estimators over the individual studies' estimators. Finally, we will analyze two datasets to illustrate the proposed methods.

This article is organized as follows. Section 2 provides a background including a review of meta-analyses. Section 3 proposes the estimators of individual means and studies their theoretical properties. Section 4 conducts simulation studies, and Section 5 analyzes real datasets. Section 6 concludes the article.

## 2. Background

### 2.1. Meta-Analysis

This subsection reviews the basic models for meta-analysis under the normal distribution, including the fixed-effect model and random-effects model [1,5].

First, we introduce some mathematical notations and basic assumptions for meta-analyses. Let  $G > 0$  be the number of studies ( $G$  stands for Groups) in a meta-analysis. For  $i = 1, 2, \dots, G$ , the  $i$ -th study provides an estimate  $Y_i$  for an unknown mean  $\mu_i$  with a known variance  $\sigma_i^2 > 0$ . This implies that each study provides  $\sigma_i$  as the standard error (SE) of the estimate  $Y_i$ . Hence, in meta-analyses, the variance  $\sigma_i^2$  is known. We note that there could be a different setting, where the variance is unknown [20,21].

Next, we further assume the normal model  $Y_i \sim N(\mu_i, \sigma_i^2)$ . This model's special case, the fixed-effect model, imposes the common mean assumption. That is,  $\mu \equiv \mu_1 = \dots = \mu_G$ . The random-effects model imposes normally distributed means given by  $\mu_i \sim N(\mu, \tau^2)$  for  $i = 1, 2, \dots, G$ . In either case, the main goal of meta-analyses is to estimate  $\mu$  by optimally combining  $Y_i$ s [1,5,9].

In a meta-analysis, observed data consist of  $\{Y_i; i = 1, 2, \dots, G\}$ . An example is the two-sample  $t$ -test for a continuous outcome, where  $Y_i$  is the (standardized) mean difference estimator from the  $i$ -th study, and  $\sigma_i^2$  is the pooled variance estimator [5]. Another example is the  $2 \times 2$  table analysis for association between an event and a risk factor, where  $Y_i$  is the logarithm of the risk ratio (or odds ratio), and  $\sigma_i^2$  is estimated variance [5]. Another example is a survival analysis for censored data, where  $Y_i$  can be the logarithm of an estimate of relative risks from the Cox model. Without loss of generality, we assume that  $\mu_j = 0$  corresponds to the null value.

Finally, we describe the main idea of meta-analyses. The idea is to optimally combine the results from the  $G$  studies, based on either the fixed-effect model or random-effects model. The first model leads to the *fixed-effect meta-analysis* under the assumption  $\mu \equiv \mu_1 = \dots = \mu_G$ . The second model leads to the *random-effects meta-analysis* under the assumption  $\mu_i \sim N(\mu, \tau^2)$ . In either case, the target is to estimate the common (overall) mean  $\mu$  by

$$\hat{\mu}^{\text{Fixed}} = \frac{\sum_{i=1}^G (\sigma_i^2)^{-1} Y_i}{\sum_{i=1}^G (\sigma_i^2)^{-1}},$$

$$\hat{\mu}^{\text{Rand}} = \frac{\sum_{i=1}^G (\sigma_i^2 + \hat{\tau}^2)^{-1} Y_i}{\sum_{i=1}^G (\sigma_i^2 + \hat{\tau}^2)^{-1}},$$

where the value of  $\hat{\tau}^2$  is an estimator from the data [5,14].

## 2.2. Does the Estimation of the Common Mean Make Sense?

Given the data of  $\{Y_i: i = 1, 2, \dots, G\}$ , one usually conducts the fixed-effect meta-analysis and/or random-effects meta-analysis. The  $Q$  test for homogeneity can help select between the fixed-effect and random-effects model [1,5]. If the fixed-effect model is rejected (i.e., the random-effects model is selected), one might explore a systematic source of heterogeneity by meta-regression. One might also test the goodness-of-fit of the normal random-effects model [22]. In any case, statistical inference for the common mean depends on some statistical models.

However, the above statistical models do not always fit the data at hand. If the studies have a *monotonic-effect*, e.g.,  $\mu_1 \leq \dots \leq \mu_G$ , the model is neither fixed nor random. Thus, there is no general way to estimate the common parameter unless some specific models are imposed. If a covariate is available to explain some structure among  $\mu_i$ s, a meta-regression may be used (Chapter 20 of [1]). However, we do not wish to assume the presence of covariates to conduct meta-regression in this article.

While the common mean estimate can be an informative summary of individual estimates, this single value rarely describes the whole results. Thus, meta-analysis typically displays individual estimates  $(y_1, \dots, y_G)$ , providing an opportunity for learning the data. For instance, forest plots show the individual estimates  $(y_1, \dots, y_G)$  along with their 95% confidence intervals (CIs):  $(y_1, \dots, y_G) \pm 1.96 \times (\sigma_1, \dots, \sigma_G)$ . Funnel plots show  $(y_1, \dots, y_G)$  against  $(\sigma_1, \dots, \sigma_G)$ ; see [1,5,23,24] for these plots. In some cases, one is interested in the  $\mu_i$  that has the largest size [16]. Therefore, reporting individual estimates is an important part of meta-analysis.

## 2.3. Individual Estimates Could Be Improved

Despite the potential advantages of improving the individual estimators by the shrinkage methods [11,12,25,26], the development of such estimators has rarely been discussed in the literature of meta-analyses. In other words, the goal of meta-analyses has been the estimation of the common mean by disregarding the individual means. This causes a gap between meta-analyses and shrinkage estimation methods. Therefore, there is room for improving the standard individual estimators  $(y_1, \dots, y_G)$  for  $(\mu_1, \dots, \mu_G)$  in some scenarios.

We raise scenarios where the common mean is undefined or uninformative.

The first example is the case of the *monotonic-effect*,  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_G$ , where the common mean seems to be not well-defined. In this case, an alternative is to perform meta-regression with the aid of observable covariates. We however do not assume the availability of covariates in this article.

The second example is the case of sparse normal means [27,28], where most of  $\mu_i$ s are equal to zero, e.g.,  $(\mu_1, \dots, \mu_{10}) = (-5, 0, 0, 0, 0, 0, 0, 0, 0, 5)$ . This is relevant to meta-analyses that fail to reject many null hypotheses. In this case, the interest may be in the estimation of the nonzero means rather than the common mean.

## 3. Proposed Estimators

We propose a variety of shrinkage estimators for simultaneously estimating individual means  $(\mu_1, \dots, \mu_G)$ , and discuss their properties. In particular, we theoretically prove that the proposed estimators have better precision than the individual studies' estimates  $(y_1, \dots, y_G)$  in terms of a mean squared error criterion under some conditions.

The problem setting is as follows: Let  $Y_i \sim N(\mu_i, \sigma_i^2)$  be an estimator for an unknown mean  $\mu_i$  under a *known* variance  $\sigma_i^2 > 0$  for  $i = 1, 2, \dots, G$ . As mentioned in Section 2, the variance  $\sigma_i^2$  is known from the  $i$ -th study. Observed data consist of  $\{Y_i; i = 1, 2, \dots, G\}$ . We regard  $\mathbf{Y} \equiv (Y_1, \dots, Y_G)$  as a *simultaneous estimator* of  $\boldsymbol{\mu} \equiv (\mu_1, \dots, \mu_G)$ . That is, we regard  $\mathbf{Y}$  as an observable random vector to estimate the unknown vector  $\boldsymbol{\mu}$ .

Let  $\boldsymbol{\delta}(\mathbf{Y}) \equiv (\delta_1(\mathbf{Y}), \dots, \delta_G(\mathbf{Y}))$  be a simultaneous estimator for  $\boldsymbol{\mu}$ . We state that  $\boldsymbol{\delta}(\mathbf{Y})$  improves upon  $\mathbf{Y}$  if

$$E \left[ \sum_{i=1}^G \frac{(\delta_i(\mathbf{Y}) - \mu_i)^2}{\sigma_i^2} \right] < E \left[ \sum_{i=1}^G \frac{(Y_i - \mu_i)^2}{\sigma_i^2} \right] = G, \quad \exists (\mu_1, \dots, \mu_G), \quad (1)$$

$$E \left[ \sum_{i=1}^G \frac{(\delta_i(\mathbf{Y}) - \mu_i)^2}{\sigma_i^2} \right] \leq E \left[ \sum_{i=1}^G \frac{(Y_i - \mu_i)^2}{\sigma_i^2} \right] = G, \quad \forall (\mu_1, \dots, \mu_G). \quad (2)$$

We then call  $\delta(\mathbf{Y})$  an *improved estimator* of  $\mathbf{Y}$ .

We call  $\delta(\mathbf{Y}) = \mathbf{Y}$  the *standard unbiased estimator* because of the unbiasedness  $E(\mathbf{Y}) = \boldsymbol{\mu}$ . It is a benchmark estimator for comparing different estimators. Recall that  $\mathbf{Y}$  is simply the collection of the individual studies' estimates. If " $\forall (\mu_1, \dots, \mu_G)$ " in the criterion (2) holds only for a restricted parameter space,  $\delta(\mathbf{Y})$  is only locally improved.

The criteria (1) and (2) employ the weighted mean squared errors (WMSEs). The inverse variance weights guarantee the equal contributions of all studies to the estimation error. The weights also make it convenient to apply the classical decision theory. The special case of  $\sigma_1 = \dots = \sigma_G = 1$  gives the total MSE (TMSE), defined as  $E[\sum_{i=1}^G (\delta_i(\mathbf{Y}) - \mu_i)^2]$  [11,25,29]. We use the WMSEs for theoretical convenience while the TMSE criterion is also a relevant criterion. We will use both the WMSE and TMSE in our numerical assessments.

Indeed, there are infinitely many estimators  $\delta(\mathbf{Y})$  that improve upon  $\mathbf{Y}$ , including very complex and unrealistic ones [25]. In addition, it is quite easy to find an estimator that locally improves upon  $\mathbf{Y}$ , such as  $\delta(\mathbf{Y}) = \mathbf{0}$ . While the problem of deriving/assessing estimators has been intensively discussed in the statistical decision theory, it has rarely been appreciated in meta-analytical settings. The goal of this article is to facilitate the applications of shrinkage estimators in the context of meta-analyses.

In the subsequent sections, we will introduce three estimators that help reduce the WMSE and TMSE by shrinking  $\mathbf{Y}$  toward a restricted space of  $(\mu_1, \dots, \mu_G)$ . Section 3.1 will discuss the shrinkage towards the zero vector  $\mathbf{0} = \{(\mu_1, \dots, \mu_G): \mu_1 = \dots = \mu_G = 0\}$ , the most traditional shrinkage scheme. Section 3.2 will consider the shrinkage toward  $\mathbf{0}$  under constraints  $\{(\mu_1, \dots, \mu_G): \mu_1 \leq \dots \leq \mu_G\}$ . Section 3.3 will explore the shrinkage towards the *sparse* space  $\{(\mu_1, \dots, \mu_G): \sum_{i=1}^G I(\mu_i \neq 0) < G\}$ .

### 3.1. Shrinkage Estimation for Means

To estimate  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_G)$ , we propose the James–Stein (JS) estimator of the form

$$\delta^{\text{JS}} \equiv (\delta_1^{\text{JS}}, \dots, \delta_G^{\text{JS}}) \equiv \left( 1 - \frac{G-2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2} \right) \mathbf{Y}.$$

This estimator is a modification of the original JS estimator [29] that was derived under the unit variances ( $\sigma_i = 1$  for  $\forall i$ ). See Appendix A.1 for the details. The JS estimator reduces variance by shrinking the vector  $\mathbf{Y}$  toward  $\mathbf{0}$  while it produces bias. The degree of shrinkage is determined by the factor  $(G-2)/(\sum_{i=1}^G Y_i^2 / \sigma_i^2)$  that usually ranges from 0 (0% shrinkage) to 1 (100% shrinkage), and occasionally becomes greater than 1 (overshrinkage).

It can be shown in Appendix A.1 that  $\delta^{\text{JS}}$  has the following WMSE

$$E \left[ \sum_{i=1}^G \frac{(\delta_i^{\text{JS}}(\mathbf{Y}) - \mu_i)^2}{\sigma_i^2} \right] = G - (G-2) E \left[ \frac{1}{\chi_G^2(\lambda)} \right],$$

where  $\chi_G^2(\lambda)$  is a random variable having a noncentral  $\chi^2$ -distribution with the noncentral parameter  $\lambda = \sum_{i=1}^G \mu_i^2 / \sigma_i^2$  and the degrees of freedom  $G$ . Thus,  $\delta^{\text{JS}}$  has a smaller WMSE than  $\mathbf{Y}$  when  $G \geq 3$ . Indeed, the WMSE is minimized at  $\boldsymbol{\mu} = \mathbf{0}$  at which the WMSE is  $G - 1/(G-2)$  by the inverse moment of the central  $\chi^2$ -distribution with  $\lambda = 0$ . Thus, the JS estimator gains the greatest advantage if all the individual means are zero. This gain is appealing for meta-analyses for small individual effects (true means close to zero). Even if  $\boldsymbol{\mu} \neq$

$\mathbf{0}$ , the JS estimator has a smaller WMSE than  $\mathbf{Y}$ . The reduction of the WMSE diminishes as  $\lambda$  departs from zero.

One might ask where the special formula of the JS estimator comes from. The JS estimator  $\delta^{\text{JS}}$  can be derived as an empirical Bayes estimator under the prior  $\mu_i \sim N(0, \sigma_i^2 \tau^2)$ :

$$\delta^{\text{JS}} = \left(1 - \left(\frac{1}{1 + \tau^2}\right)\right) \mathbf{Y},$$

where the shrinkage factor  $(1/(1 + \tau^2)) \equiv (G - 2)/\sum_{i=1}^G Y_i^2/\sigma_i^2$  is the estimator of  $1/(1 + \tau^2)$ . See Appendix A.2 for the detailed derivations. Thus, if  $\mu_i \sim N(0, \sigma_i^2 \tau^2)$ , the JS estimator minimizes the Bayes risk, and hence, is the optimal estimator under the prior.

A minor modification to the JS estimator can reduce the WMSE further. The modification is made in order to avoid the effect of an overshrinkage,  $(G - 2)/(\sum_{i=1}^G Y_i^2/\sigma_i^2) > 1$ , by which all the signs of  $\mathbf{Y}$  are reverted. The overshrinkage phenomenon occurs with a small probability, and therefore, the modification is minor in the majority of cases. A modified estimator is the *positive-part JS estimator*

$$\delta^{\text{JS}+} \equiv (\delta_1^{\text{JS}+}, \dots, \delta_G^{\text{JS}+}) \equiv \left(1 - \frac{G - 2}{\sum_{i=1}^G Y_i^2/\sigma_i^2}\right)^+ \mathbf{Y},$$

where  $(\cdot)^+ \equiv \max(0, \cdot)$ . Consequently,  $\delta^{\text{JS}+}$  has a smaller WMSE than  $\delta^{\text{JS}}$  (p. 356, Theorem 5.4 of [25]).

In summary, this subsection proposes two estimators ( $\delta^{\text{JS}}$  and  $\delta^{\text{JS}+}$ ) that improve upon the standard unbiased estimator  $\mathbf{Y}$ .

### 3.2. Estimation under Ordered Means

Our next proposal is a shrinkage estimator under *ordered means*. We consider the case where the ordering constraints  $\mu_1 \leq \dots \leq \mu_G$  are known by the study design. Thus, the parameter  $\boldsymbol{\mu}$  is known to be restricted on the space  $\{(\mu_1, \dots, \mu_G) : \mu_1 \leq \dots \leq \mu_G\}$ . For instance, suppose that  $i$  ( $= 1, 2, \dots, G$ ) represents the time index ( $i = 1$  for the oldest study, and  $i = G$  for the newest study) at which a treatment effect  $\mu_i$  is estimated. Then, one may assume a trend  $\mu_1 \leq \dots \leq \mu_G$  due to the improvements of treatments over time.

For instance, the true means may be  $(\mu_1, \dots, \mu_5) = (-2, -1, 0, 1, 2)$ . This trend could be modeled by a meta-regression with  $\mu_i = a + bi$ , where values  $a$  and  $b$  are unknown. In practice, one does not know any structure of the means (e.g., linear regression) except for  $\mu_1 \leq \dots \leq \mu_5$ . If some knowledge, such as a linear model on covariates, is true, one could use meta-regression. However, we do not adopt any model, permitting various non-linear settings such as  $(-2, -2, 0, 0, 0)$  and  $(-2, -1, 4, 4, 5)$ .

The use of the standard unbiased estimator  $\mathbf{Y} = (Y_1, \dots, Y_G)$  is not desirable under the ordering constraints. Due to random variations, the estimator  $\mathbf{Y} = (Y_1, \dots, Y_G)$  can be outside the parameter space, namely,  $\mathbf{Y} \notin \{(\mu_1, \dots, \mu_G) : \mu_1 \leq \dots \leq \mu_G\}$ . Under this setting, an estimator accounting for the parameter restriction improves upon the unrestricted estimator  $\mathbf{Y}$ , even though the former is a biased estimator [30].

The restricted maximum likelihood (RML) estimator satisfying  $\delta_1 \leq \dots \leq \delta_G$  is calculated by the pool-adjacent-violators algorithm (PAVA)

$$\delta_i^{\text{RML}} = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{j=s}^t Y_j}{t - s + 1}.$$

This gives the RML estimator  $\delta^{\text{RML}} \equiv (\delta_1^{\text{RML}}, \dots, \delta_G^{\text{RML}})$ , which has a smaller WMSE than  $\mathbf{Y}$ . For an example of  $G = 3$ , one has the data of  $(Y_1, Y_2, Y_3)$ , and the PAVA results in

$$\delta_1^{\text{RML}} = \min\left(Y_1, \frac{Y_1 + Y_2}{2}, \frac{Y_1 + Y_2 + Y_3}{3}\right),$$

$$\delta_2^{\text{RML}} = \max\left\{\min\left(\frac{Y_1 + Y_2}{2}, \frac{Y_1 + Y_2 + Y_3}{3}\right), \min\left(Y_2, \frac{Y_2 + Y_3}{2}\right)\right\},$$

$$\delta_3^{\text{RML}} = \max\left(\frac{Y_1 + Y_2 + Y_3}{3}, \frac{Y_2 + Y_3}{2}, Y_3\right).$$

Hence,  $\delta_i^{\text{RML}}$  is equal to  $Y_i$  itself or an average including  $Y_i$ . Of course,  $\delta_i^{\text{RML}} = Y_i \forall i$  if  $Y_1 \leq \dots \leq Y_G$ . For theories and applications of the PAVA, we refer to [31–33]. The max min formula written above can be found in Chapter 8 of [30] or [34].

It is clear that  $\delta^{\text{RML}}$  is different from order statistics  $Y_{(1)} \leq \dots \leq Y_{(G)}$  that are a permutation of  $(Y_1, \dots, Y_G)$  and also improve the WMSE in some cases [34]. However, the permuted estimator loses the information of individual studies' identifications, and therefore, is not considered in this article.

Below, we further improve  $\delta^{\text{RML}}$  with the aid of the JS estimator. Let  $I(\cdot)$  be the indicator function;  $I(A) = 1$  or  $I(A) = 0$  if  $A$  is true or false, respectively. We adjust the estimator of Chang [35] who proposed the JS-type estimator under the order restriction as follows:

$$\delta^{\text{RJS}} \equiv \left(1 - \frac{G-2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2}\right) YI(Y_1 \leq \dots \leq Y_G) + \delta^{\text{RML}}(1 - I(Y_1 \leq \dots \leq Y_G)), \quad (2)$$

where “RJS” stands for “Restricted JS”. Note that  $\delta^{\text{RJS}}$  has a smaller WMSE than  $\delta^{\text{RML}}$ , and hence, the former improves upon the latter [35].

We further improve the RJS estimator by the positive-part RJS estimator given by

$$\delta^{\text{RJS}^+} \equiv \left(1 - \frac{G-2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2}\right)^+ YI(Y_1 \leq \dots \leq Y_G) + \delta^{\text{RML}}(1 - I(Y_1 \leq \dots \leq Y_G))$$

Consequently,  $\delta^{\text{RJS}^+}$  has a smaller WMSE than  $\delta^{\text{RJS}}$  (Theorem 5.4 of Lehmann and Casella [25]). Note that, if  $Y_1 \leq \dots \leq Y_G$  is not satisfied, then  $\delta^{\text{RML}} = \delta^{\text{RJS}} = \delta^{\text{RJS}^+}$ .

In summary, this subsection proposes three estimators ( $\delta^{\text{RML}}$ ,  $\delta^{\text{RJS}}$ , and  $\delta^{\text{RJS}^+}$ ) that improve upon the standard unbiased estimator  $Y$  under  $\mu_1 \leq \dots \leq \mu_G$ .

### 3.3. Estimation under Sparse Means

Our third proposal is a shrinkage estimator under *sparse normal means* where most of the  $\mu_i$ s are zero [27,28]. The vector  $(\mu_1, \dots, \mu_G)$  is called *sparse* if the number  $\sum_{i=1}^G I(\mu_i \neq 0)$  is much smaller than  $G$ , e.g.,  $(\mu_1, \dots, \mu_{10}) = (-5, 0, 0, 0, 0, 0, 0, 0, 0, 5)$ . In practice, one does not know which components are zeros, and how many components are zero. Nonetheless, one could assume that many of  $(\mu_1, \dots, \mu_G)$  are zero. However, the elements of  $(Y_1, \dots, Y_G)$  are almost always nonzero, which disagree with the true values  $(\mu_1, \dots, \mu_G)$ .

Under the sparse means, it is quite reasonable to estimate  $\mu_i$  as exactly zero if  $Y_i$  is close to zero. Accordingly, one can use a thresholding estimator  $Y_i I(|Y_i| > c_i)$  for a critical value  $c_i > 0$ . The idea was proposed by Bancroft [36] who formulated *pretest estimators* that incorporate a preliminary hypothesis test into estimation. Judge and Bock [37] extensively studied pretest estimators with applications to econometrics; see also more recent works [38–43]. Among all, we particularly note that Shih et al. [41] proposed the *general pretest* (GPT) estimator that includes empirical Bayes and Types I-II shrinkage pretest estimators for the univariate normal mean.

We modify the GPT estimator to be adopted to meta-analyses as follows:

$$\delta_i^{\text{GPT}} = Y_i I\left(\left|\frac{Y_i}{\sigma_i}\right| > z_{\alpha_1}\right) + q\left(\frac{Y_i}{\sigma_i}\right) Y_i I\left(z_{\alpha_2} < \left|\frac{Y_i}{\sigma_i}\right| \leq z_{\alpha_1}\right), \quad i = 1, 2, \dots, G.$$

for  $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ ,  $q: (-\infty, \infty) \mapsto (0, 1)$ , and  $z_p$  is the upper  $p$ -th quantile of  $N(0, 1)$  for  $0 < p < 1$ . To implement the GPT estimator, the values of  $\alpha_1$  and  $\alpha_2$ , and the probability function  $q(\cdot)$  must be chosen. They cannot be chosen to minimize the WMSE and TMSE criteria since pretest estimators do not permit tractable forms of MSEs [38,41]. Fortunately, for any value of  $\alpha_1$  and  $\alpha_2$ , and a function  $q$ , one can show that  $\delta^{\text{GPT}} \equiv (\delta_1^{\text{GPT}}, \dots, \delta_G^{\text{GPT}})$  has smaller WMSE and TMSE values than  $Y$  provided  $\mu \approx \mathbf{0}$ ; see Appendix A.3 for the proof.

For the above reasons, we apply statistically interpretable choices of  $\alpha_1$ ,  $\alpha_2$ , and  $q(\cdot)$ . The special case of  $\alpha_1 = \alpha_2 = \alpha = 0.025$  leads to the usual pretest estimator

$$\delta_i^{\text{PT}} = Y_i I\left(\left|\frac{Y_i}{\sigma_i}\right| > 1.96\right),$$

for which  $q(\cdot)$  is arbitrary. Thus, we retain  $Y_i$  if  $H_0: \mu_i = 0$  is rejected in favor of  $H_1: \mu_i \neq 0$  at the 5% level. Otherwise, we discard  $Y_i$ , and conclude  $\mu_i = 0$ .

For the GPT estimator, we set  $q(z) = 1/2$  (50% shrinkage) as suggested by Shih et al. [41]. To facilitate the interpretability of pretests, we chose  $\alpha_1 = 0.025$  (5% level) and  $\alpha_2 = 0.05$  (10% level). Thus, we propose the estimator

$$\delta_i^{\text{GPT}} = Y_i I\left(\left|\frac{Y_i}{\sigma_i}\right| > 1.96\right) + \frac{Y_i}{2} I\left(1.645 < \left|\frac{Y_i}{\sigma_i}\right| \leq 1.96\right), \quad i = 1, 2, \dots, G.$$

Thus, if  $H_0: \mu_i = 0$  is rejected at the 5% level, we set  $\delta_i^{\text{GPT}} = Y_i$ . If  $H_0: \mu_i = 0$  is accepted at the 5% level, but rejected at the 10% level, we set  $\delta_i^{\text{GPT}} = Y_i/2$ . If  $H_0: \mu_i = 0$  is accepted at the 10% level, we set  $\delta_i^{\text{GPT}} = 0$ . Thus,  $\delta_i^{\text{GPT}}$  gives a weaker shrinkage than  $\delta_i^{\text{PT}}$  does. Obviously, we obtain a relationship  $|\delta_i^{\text{PT}}| \leq |\delta_i^{\text{GPT}}| \leq |Y_i|$ .

The GPT estimator introduced above is not an empirical Bayesian estimator. If  $\alpha_1 = 0$ ,  $\alpha_2 = 1$ , and  $q(y) = 1 - \sigma_i^2 / \max\{\sigma_i^2, y^2\}$  were chosen, the resultant GPT estimator would be an empirical Bayes estimator [41]. However, we do not consider this estimator in our analysis.

In summary, this subsection proposes two estimators ( $\delta^{\text{PT}}$  and  $\delta^{\text{GPT}}$ ) that improve upon the standard unbiased estimator  $\mathbf{Y}$  provided  $\boldsymbol{\mu} \approx \mathbf{0}$ .

#### 4. Simulation

We conducted Monte Carlo simulations to examine the performance of the proposed estimators. The purpose of the simulations is to clarify how the proposed estimators improve upon the standard unbiased estimator  $\mathbf{Y}$  in terms of the WMSE and TMSE, defined as

$$\text{WMSE} \equiv E \left[ \sum_{i=1}^G \frac{(\delta_i(\mathbf{Y}) - \mu_i)^2}{\sigma_i^2} \right],$$

$$\text{TMSE} \equiv E \left[ \sum_{i=1}^G (\delta_i(\mathbf{Y}) - \mu_i)^2 \right].$$

We also examined the bias defined as,

$$\text{Bias} \equiv E[\delta_i(\mathbf{Y}) - \mu_i], \quad i = 1, 2, \dots, G.$$

##### 4.1. Simulation Design

We generated normally distributed data  $Y_i \sim N(\mu_i, \sigma_i^2)$ ,  $i = 1, 2, \dots, G$ , under the following scenarios:

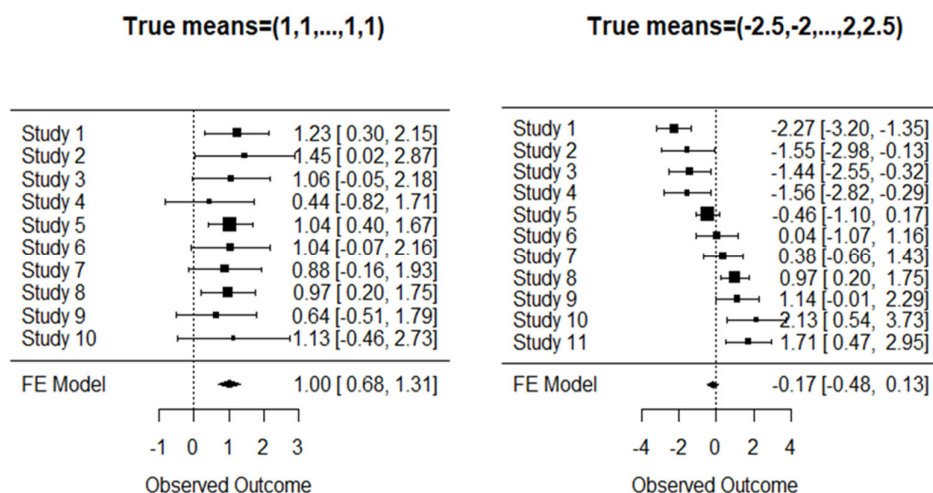
**Scenario (a):** Common means:  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_G) = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ ,

**Scenario (b):** Zero means:  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_G) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ ,

**Scenario (c):** Ordered means:  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_G) = (-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5)$ ,

**Scenario (d):** Sparse means:  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_G) = (-5, 0, 0, 0, 0, 0, 0, 0, 0, 5)$ .

For the known variances, we generated  $\sigma_i^2 \sim \chi_{df=1}^2/4$ , restricted in the interval  $[0.009, 0.6]$ , as previously considered [15,44], leading to  $E[\sigma_i^2] = 0.173$ . Figure 1 depicts simulated data  $(Y_1, \dots, Y_G)$  and  $(\sigma_1, \dots, \sigma_G)$  under Scenarios (a) and (c). Using the data, we computed the proposed estimators  $\delta^{\text{S}}$ ,  $\delta^{\text{S+}}$ ,  $\delta^{\text{RML}}$ ,  $\delta^{\text{RJS}}$ ,  $\delta^{\text{RJS+}}$ ,  $\delta^{\text{PT}}$ , and  $\delta^{\text{GPT}}$  for estimating  $\boldsymbol{\mu}$ . Our interest is to see how and when these estimators improve upon the standard unbiased estimator  $\mathbf{Y}$ . The WMSE, TMSE and Bias of all the estimators were computed by the Monte Carlo average of 10,000 replications of such data.



**Figure 1.** Simulated values  $(Y_1, \dots, Y_G)$  and the 95%CI  $(Y_1, \dots, Y_G) \pm 1.96 \times (\sigma_1, \dots, \sigma_G)$ , where  $Y_i \sim N(\mu_i, \sigma_i^2)$ , and  $\sigma_i^2 \sim \chi_{df=1}^2/4$  being restricted in the interval  $[0.009, 0.6]$ , for  $i = 1, 2, \dots, G$ . The “FE Model” is the common mean estimator and its 95%CI under the fixed-effect (FE) model.

#### 4.2. Simulation Results

Tables 1–4 summarize the simulation results for comparing eight estimators:  $Y$ ,  $\delta^{JS}$ ,  $\delta^{JS+}$ ,  $\delta^{RML}$ ,  $\delta^{RJS}$ ,  $\delta^{RJS+}$ ,  $\delta^{PT}$ , and  $\delta^{GPT}$ .

The JS estimators ( $\delta^{JS}$  and  $\delta^{JS+}$ ) improved upon  $Y$  in all the scenarios (see the WMSE and TMSE in Tables 1–4). Scenario (b) produced the greatest reduction in the values of the WMSE and TMSE, where all the true means are equal to zero (Table 2). This is exactly what the theoretical results imply (Section 3.1). The two estimators  $\delta^{JS}$  and  $\delta^{JS+}$  showed similar performance except for Scenario (b), where  $\delta^{JS+}$  exhibited smaller values for the WMSE and TMSE than  $\delta^{JS}$  (Table 2). In summary, we observed a clear advantage of the two JS estimators ( $\delta^{JS}$  and  $\delta^{JS+}$ ) over the standard estimator ( $Y$ ) in all the scenarios examined.

The order-restricted estimators ( $\delta^{RML}$ ,  $\delta^{RJS}$  and  $\delta^{RJS+}$ ) were even more advantageous than  $\delta^{JS}$  and  $\delta^{JS+}$  according to the smallest values of the WMSE and TMSE in Tables 1 and 3. The substantially reduced values of the WMSE and TMSE were brought by the strict constraints by isotonic regression. We found that  $\delta^{RJS+}$  had the smallest value of the WMSE and TMSE under Scenarios (a) and (c) (Tables 1 and 3). However, we observed large and systematic biases of the order-restricted estimators, which are caused by isotonic regression.

The pretest estimator  $\delta^{PT}$  showed the best performance under Scenario (d) (Table 4), followed by the general pretest estimator  $\delta^{GPT}$ . However, the comparison between  $\delta^{PT}$  and  $\delta^{GPT}$  depends on the scenarios:  $\delta^{PT}$  was better in Scenarios (b) and (d) while  $\delta^{GPT}$  was better in Scenarios (a) and (c). Both  $\delta^{PT}$  and  $\delta^{GPT}$  did not perform well in Scenarios (a) and (c), where the sparsity hypothesis  $H_0: \mu_i = 0$  fails. Naturally, the sparsity of individual means is essential for the pretest estimators to perform well.

In summary, our simulations have shown that the proposed estimators have their own advantageous scenarios:  $\delta^{JS+}$  for Scenario (b);  $\delta^{RJS+}$  for Scenarios (a) and (c);  $\delta^{PT}$  or  $\delta^{GPT}$  for Scenario (d). These proposed estimators substantially reduced the values of the WMSE and TMSE over the standard estimator  $Y$ . Furthermore, the proposed estimators  $\delta^{JS}$ ,  $\delta^{JS+}$ ,  $\delta^{RML}$ ,  $\delta^{RJS}$ , and  $\delta^{RJS+}$  improved upon  $Y$  across all the scenarios. We therefore conclude that there are good reasons to apply the proposed estimators to estimate  $\mu$  in order to improve the accuracy of estimation.

**Table 1.** Simulation results under **Scenario (a):**  $\mu = (\mu_1, \dots, \mu_{10}) = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ .

	$Y$	$\delta^{JS}$	$\delta^{JS+}$	$\delta^{RML}$	$\delta^{RJS}$	$\delta^{RJS+}$	$\delta^{PT}$	$\delta^{GPT}$
TMSE for $\mu$	1.743	1.608	1.608	<b>0.503</b>	<b>0.503</b>	<b>0.503</b>	3.579	3.137
WMSE for $\mu$	10.045	9.635	9.635	<b>5.761</b>	<b>5.761</b>	<b>5.761</b>	17.491	15.586
Bias for $\mu_1$	0.01	−0.04	−0.04	−0.21	−0.21	−0.21	−0.15	−0.12
Bias for $\mu_2$	<b>0.00</b>	−0.05	−0.05	−0.12	−0.12	−0.12	−0.15	−0.12
Bias for $\mu_3$	<b>0.00</b>	−0.05	−0.05	−0.07	−0.07	−0.07	−0.16	−0.13
Bias for $\mu_4$	<b>0.00</b>	−0.05	−0.05	−0.04	−0.04	−0.04	−0.16	−0.13
Bias for $\mu_5$	<b>0.00</b>	−0.05	−0.05	−0.01	−0.01	−0.01	−0.16	−0.13
Bias for $\mu_6$	<b>0.00</b>	−0.05	−0.05	0.01	0.01	0.01	−0.16	−0.13
Bias for $\mu_7$	<b>0.00</b>	−0.05	−0.05	0.04	0.04	0.04	−0.16	−0.13
Bias for $\mu_8$	<b>0.00</b>	−0.05	−0.05	0.07	0.07	0.07	−0.16	−0.13
Bias for $\mu_9$	<b>0.00</b>	−0.05	−0.05	0.12	0.12	0.12	−0.15	−0.13
Bias for $\mu_{10}$	<b>0.00</b>	−0.05	−0.05	0.21	0.21	0.21	−0.16	−0.13

Notes:  $TMSE \equiv E[\sum_{i=1}^G (\delta_i(Y) - \mu_i)^2]$ ,  $WMSE \equiv E[\sum_{i=1}^G \frac{(\delta_i(Y) - \mu_i)^2}{\sigma_i^2}]$ ,  $Bias \equiv E[\delta_i(Y) - \mu_i]$ .

**Table 2.** Simulation results under **Scenario (b):**  $\mu = (\mu_1, \dots, \mu_{10}) = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ .

	$Y$	$\delta^{JS}$	$\delta^{JS+}$	$\delta^{RML}$	$\delta^{RJS}$	$\delta^{RJS+}$	$\delta^{PT}$	$\delta^{GPT}$
TMSE for $\mu$	1.743	0.353	<b>0.223</b>	0.503	0.503	0.503	0.489	0.558
WMSE for $\mu$	10.045	2.051	<b>1.287</b>	5.761	5.761	5.761	2.837	3.236
Bias for $\mu_1$	0.01	<b>0.00</b>	<b>0.00</b>	−0.21	−0.21	−0.21	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_2$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.12	−0.12	−0.12	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_3$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.07	−0.07	−0.07	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_4$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.04	−0.04	−0.04	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_5$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.01	−0.01	−0.01	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_6$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.01	0.01	0.01	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_7$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.04	0.04	0.04	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_8$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.07	0.07	0.07	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_9$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.12	0.12	0.12	<b>0.00</b>	<b>0.00</b>
Bias for $\mu_{10}$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.21	0.21	0.21	<b>0.00</b>	<b>0.00</b>

Notes:  $TMSE \equiv E[\sum_{i=1}^G (\delta_i(Y) - \mu_i)^2]$ ,  $WMSE \equiv E[\sum_{i=1}^G \frac{(\delta_i(Y) - \mu_i)^2}{\sigma_i^2}]$ ,  $Bias \equiv E[\delta_i(Y) - \mu_i]$ .

**Table 3.** Simulation results under **Scenario (c):**  $\mu = (\mu_1, \dots, \mu_{11}) = (-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5)$ .

	$Y$	$\delta^{JS}$	$\delta^{JS+}$	$\delta^{RML}$	$\delta^{RJS}$	$\delta^{RJS+}$	$\delta^{PT}$	$\delta^{GPT}$
TMSE for $\mu$	1.914	1.847	1.847	1.314	<b>1.309</b>	<b>1.309</b>	2.826	2.547
WMSE for $\mu$	11.041	10.831	10.831	9.259	<b>9.240</b>	<b>9.240</b>	14.600	13.494
Bias for $\mu_1$	0.01	0.07	0.07	−0.03	−0.02	−0.02	0.02	0.02
Bias for $\mu_2$	<b>0.00</b>	0.05	0.05	<b>0.00</b>	0.01	0.01	0.03	0.03
Bias for $\mu_3$	<b>0.00</b>	0.03	0.03	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.08	0.06
Bias for $\mu_4$	<b>0.00</b>	0.02	0.02	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.16	0.13
Bias for $\mu_5$	<b>0.00</b>	0.01	0.01	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.20	0.17
Bias for $\mu_6$	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.00	0.00
Bias for $\mu_7$	<b>0.00</b>	−0.01	−0.01	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.20	−0.17
Bias for $\mu_8$	<b>0.00</b>	−0.02	−0.02	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.16	−0.13
Bias for $\mu_9$	<b>0.00</b>	−0.04	−0.04	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	−0.08	−0.06
Bias for $\mu_{10}$	<b>0.00</b>	−0.05	−0.05	<b>0.00</b>	−0.01	−0.01	−0.03	−0.02
Bias for $\mu_{11}$	<b>0.00</b>	−0.06	−0.06	0.03	0.03	0.03	−0.01	−0.01

Notes:  $TMSE \equiv E[\sum_{i=1}^G (\delta_i(Y) - \mu_i)^2]$ ,  $WMSE \equiv E[\sum_{i=1}^G \frac{(\delta_i(Y) - \mu_i)^2}{\sigma_i^2}]$ ,  $Bias \equiv E[\delta_i(Y) - \mu_i]$ .

**Table 4.** Simulation results under **Scenario (d):**  $\mu = (\mu_1, \dots, \mu_{10}) = (-5, 0, 0, 0, 0, 0, 0, 0, 0, 5)$ .

	$Y$	$\delta^{JS}$	$\delta^{JS+}$	$\delta^{RML}$	$\delta^{RJS}$	$\delta^{RJS+}$	$\delta^{PT}$	$\delta^{GPT}$
TMSE for $\mu$	1.743	1.719	1.719	0.823	0.823	0.823	<b>0.739</b>	0.794
WMSE for $\mu$	10.1045	9.897	9.897	7.272	7.272	7.272	<b>4.281</b>	4.599
Bias for $\mu_1$	0.01	0.10	0.10	0.01	0.01	0.01	0.01	0.01
Bias for $\mu_2$	0.00	0.00	0.00	−0.21	−0.21	−0.21	0.00	0.00
Bias for $\mu_3$	0.00	0.00	0.00	−0.11	−0.11	−0.11	0.00	0.00
Bias for $\mu_4$	0.00	0.00	0.00	−0.06	−0.06	−0.06	0.00	0.00
Bias for $\mu_5$	0.00	0.00	0.00	−0.02	−0.02	−0.02	0.00	0.00
Bias for $\mu_6$	0.00	0.00	0.00	−0.02	−0.02	−0.02	0.00	0.00
Bias for $\mu_7$	0.00	0.00	0.00	0.02	0.02	0.02	0.00	0.00
Bias for $\mu_8$	0.00	0.00	0.00	0.05	0.05	0.05	0.00	0.00
Bias for $\mu_9$	0.00	0.00	0.00	0.10	0.10	0.10	0.00	0.00
Bias for $\mu_{10}$	0.00	0.00	0.00	0.21	0.21	0.21	0.00	0.00
Bias for $\mu_{11}$	0.00	−0.09	−0.09	0.00	0.00	0.00	0.00	0.00

Notes:  $TMSE \equiv E[\sum_{i=1}^G (\delta_i(Y) - \mu_i)^2]$ ,  $WMSE \equiv E[\sum_{i=1}^G \frac{(\delta_i(Y) - \mu_i)^2}{\sigma_i^2}]$ ,  $Bias \equiv E[\delta_i(Y) - \mu_i]$ .

## 5. Data Analyses

We analyze two datasets to illustrate the proposed method. One dataset comes from gastric cancer patients and the other from COVID-19 pneumonia patients.

### 5.1. Gastric Cancer Data

We illustrate the proposed methods using a dataset from the Global Advanced/Adjuvant Stomach Tumor Research International Collaboration (GASTRIC) [45]. The data provide survival times for gastric cancer patients from 14 studies. One of the goals of this collaboration is to estimate the effects of chemotherapy on disease-free survival (DFS) for gastric cancer patients [46]. We used the data available in the R package *surrosurv* [47].

Let  $Y_i$  be the estimator for a Cox regression coefficient for a treatment indicator (1=treatment vs. 0=control) and let  $\sigma_i$  be the SE for  $i = 1, \dots, 14$ . Table 5 summarizes  $Y_i$ ,  $\sigma_i$ , and sample sizes for the 14 studies. It is reasonable to assume the asymptotic normal approximation  $Y_i \sim N(\mu_i, \sigma_i^2)$  since the sample size of each study is large enough. Table 5 also shows the  $p$ -values for testing  $H_0: \mu_i = 0$  through Z-values  $Y_i/\sigma_i$ .

**Table 5.** The 14 studies on gastric cancer patients for estimating the effects of chemotherapy on DFS.

	Sample Size	Treatment Effect= $Y_i$	SE= $\sigma_i$	$p$ -Value
Study 1	269	−0.18312	0.15372	0.234
Study 2	190	−0.72266	0.28686	0.012 **
Study 3	252	−0.48507	0.33192	0.144
Study 4	536	−0.23961	0.21558	0.266
Study 5	219	−0.13226	0.14691	0.368
Study 6	306	−0.27228	0.14416	0.059 *
Study 7	88	−0.5867	0.24885	0.018 **
Study 8	281	−0.13969	0.14542	0.337
Study 9	271	−0.1004	0.16404	0.541
Study 10	178	−0.31143	0.17038	0.068 *
Study 11	126	−0.04949	0.19818	0.803
Study 12	180	−0.11685	0.16476	0.478
Study 13	206	−0.13044	0.19268	0.498
Study 14	186	0.04391	0.17632	0.803

\* 10% significance; \*\* 5% significance.

Table 5 shows that only two studies reached 5% significance. Thus, it is reasonable to assume  $\mu_i \approx 0 \forall i$ , and therefore, the JS estimators are suitable. It is also reasonable to assume  $\mu_i = 0$  for most of  $i$ s, and therefore, the PT and GPT estimators are suitable. We therefore calculated  $\delta^{JS}$ ,  $\delta^{JS+}$ ,  $\delta^{PT}$ , and  $\delta^{GPT}$  to estimate  $\mu$  as shown in Table 6.

**Table 6.** Estimates of the treatment effects on DFS from the 14 studies on gastric cancer patients.

	$Y$	$\delta^{JS}$	$\delta^{JS+}$	$\delta^{PT}$	$\delta^{GPT}$
$\mu_1$	−0.18	−0.10	−0.10	0	0
$\mu_2$	−0.72	−0.40	−0.40	−0.72	−0.72
$\mu_3$	−0.49	−0.27	−0.27	0	0
$\mu_4$	−0.24	−0.13	−0.13	0	0
$\mu_5$	−0.13	−0.07	−0.07	0	0
$\mu_6$	−0.27	−0.15	−0.15	0	−0.14
$\mu_7$	−0.59	−0.32	−0.32	−0.59	−0.59
$\mu_8$	−0.14	−0.08	−0.08	0	0
$\mu_9$	−0.10	−0.06	−0.06	0	0
$\mu_{10}$	−0.31	−0.17	−0.17	0	−0.16
$\mu_{11}$	−0.05	−0.03	−0.03	0	0
$\mu_{12}$	−0.12	−0.06	−0.06	0	0
$\mu_{13}$	−0.13	−0.07	−0.07	0	0
$\mu_{14}$	0.04	0.02	0.02	0	0

All values are rounded up to 2 digits.

Table 6 shows that  $\delta^{JS}$  and  $\delta^{JS+}$  shrink the values of  $Y$  by 45%, namely,  $\delta^{JS} = \delta^{JS+} = (1 - 0.45) \times Y$ . We also observe that  $\delta^{PT}$  and  $\delta^{GPT}$  cause 0%, 50% or 100% shrinkage of  $Y$ . The values of  $\delta^{PT}$  suggested the null effects ( $\mu_i = 0$ ) for 12 studies while the values of  $\delta^{GPT}$  suggested the null effects for 10 studies. By shrinkage, the estimators  $\delta^{JS}$ ,  $\delta^{JS+}$ ,  $\delta^{PT}$ , and  $\delta^{GPT}$  more accurately estimate  $\mu = (\mu_1, \dots, \mu_{14})$  than  $Y$  do. The reason is that the data structure resembles Scenario (b), where  $\delta^{JS}$  and  $\delta^{JS+}$  performed the best (Table 2) or Scenario (d), where  $\delta^{PT}$  and  $\delta^{GPT}$  performed the best (Table 4).

## 5.2. COVID-19 Data

We illustrate the proposed methods using data from Pranata et al., [7] who examined the effect of hypertension on mortality for patients with COVID-19 pneumonia. One of their conclusions was that hypertension increases mortality. We obtained mortality data from Pranata et al. [7], consisting of 11 published studies (Table 7).

**Table 7.** Summary of the 11 studies (shown by the author and publication year) on COVID-19 patients for examining the effects of hypertension on mortality.

Study	Sample Size	Male (%)	Log (RR)	SE	p-Value
Akbari 2020	440	56.4	0.6881	0.6732	0.307
Bai 2000	127	63.0	0.5933	0.2754	0.031 **
Cao 2020	102	52.0	1.1756	0.2821	0.000 **
Chen 2020	123	49.0	0.5365	0.2493	0.031 **
Chen T 2020	274	62.0	0.6780	0.1713	0.000 **
Fu 2020	200	49.5	0.5878	0.3302	0.075 *
Grasselli 2020	1591	82.0	0.4637	0.0956	0.000 **
Li 2020	102	58.0	0.5247	0.3272	0.109
Luo 2020	403	47.9	1.2326	0.1489	0.000 **
Yuan 2020	27	45.0	2.8904	1.4263	0.043 **
Zhou 2020	191	62.0	1.1378	0.2097	0.000 **

\* 10% significance; \*\* 5% significance.

Let  $Y_i$ ,  $i = 1, \dots, 11$ , be the log of the risk ratio (RR) calculated from two-by-two contingency tables examining the association (mortality vs. hypertension). Let  $\sigma_i$  be the SE for  $i = 1, \dots, 11$ . Table 7 summarizes the 11 studies. Table 7 also shows the  $p$ -values for testing the no association ( $H_0: \mu_i = 0$ ) by Z-values  $Y_i/\sigma_i$  assuming  $Y_i \sim N(\mu_i, \sigma_i^2)$ .

Before applying the proposed methods, we changed the order of the 11 studies by the percentages of male patients (Table 8). The largest percentage (82%) gave the smallest RR ( $Y_1 = 0.46$ ) while the smallest percentage (45%) gave the largest RR ( $Y_{11} = 2.89$ ). However, the values of  $Y_i$ s are not monotonically increasing ( $Y_1 < Y_2 < Y_3 < Y_4 > Y_5 < Y_6 < Y_7 > Y_8 > Y_9 < Y_{10} < Y_{11}$  in Table 8). While Pranata et al. [7] regressed  $Y_i$ s on the percentages by meta-regression, we prefer not to assume a linear model. Alternatively, we considered the order-restricted estimators ( $\delta^{\text{RML}}$ ,  $\delta^{\text{RJS}}$ , and  $\delta^{\text{RJS}+}$ ) by merely assuming that the true means are ordered by the percentages of male patients.

**Table 8.** Estimated effects of hypertension on mortality from the 11 studies on COVID-19 patients.

Study	Male (%)	Mean	$Y$	$\delta^{\text{JS}}$	$\delta^{\text{JS}+}$	$\delta^{\text{RML}}$	$\delta^{\text{RJS}}$	$\delta^{\text{RJS}+}$
Grasselli 2020	82.0	$\mu_1$	0.46	0.44	0.44	0.46	0.46	0.46
Bai 2000	63.0	$\mu_2$	0.59	0.56	0.56	0.59	0.59	0.59
Chen T 2020	62.0 *	$\mu_3$	0.68	0.64	0.64	0.68	0.68	0.68
Zhou 2020	62.0 *	$\mu_4$	1.14	1.08	1.08	0.78	0.78	0.78
Li 2020	58.0	$\mu_5$	0.52	0.50	0.50	0.78	0.78	0.78
Akbari 2020	56.4	$\mu_6$	0.69	0.65	0.65	0.78	0.78	0.78
Cao 2020	52.0	$\mu_7$	1.18	1.12	1.12	0.78	0.78	0.78
Fu 2020	49.5	$\mu_8$	0.59	0.56	0.56	0.78	0.78	0.78
Chen 2020	49.0	$\mu_9$	0.54	0.51	0.51	0.78	0.78	0.78
Luo 2020	47.9	$\mu_{10}$	1.23	1.17	1.17	1.23	1.23	1.23
Yuan 2020	45.0	$\mu_{11}$	2.89	2.74	2.74	2.89	2.89	2.89

All values are rounded up to 2 digits. \* ties are sorted by the age (62 years-old in Chen T 2020; 56 years-old in Zhou 2020).

Table 8 shows the proposed estimators  $\delta^{\text{JS}}$ ,  $\delta^{\text{JS}+}$ ,  $\delta^{\text{RML}}$ ,  $\delta^{\text{RJS}}$ , and  $\delta^{\text{RJS}+}$ . The JS shrinkage estimators were  $\delta^{\text{JS}} = \delta^{\text{JS}+} = (1 - 0.05) \times Y$ . They reduced only 5% of  $Y$ , and hence gave very weak shrinkage toward zero. This is because  $\mu_i$ s are unlikely to be zero. The order-restricted estimator  $\delta^{\text{RML}}$  successfully enforced order restrictions,  $\delta_1^{\text{RJS}} \leq \dots \leq \delta_{11}^{\text{RJS}}$ , which was not achieved by  $Y$ . Since the order restrictions were not met for  $Y$ , we had  $\delta^{\text{RML}} = \delta^{\text{RJS}} = \delta^{\text{RJS}+}$  by definition. Thus, the last three estimators can clearly incorporate the knowledge that hypertension-related mortality is monotonically influenced by gender without imposing a linear model.

In summary, it would be advantageous and informative to add the order restriction to individual studies' estimators on the association (mortality vs. hypertension). The resultant order-restricted estimators are expected to provide more accurate estimates than individual studies' estimators. The reason is that the data structure resembles Scenario (c), where  $\delta^{\text{RML}}$ ,  $\delta^{\text{RJS}}$ , and  $\delta^{\text{RJS}+}$  performed the best (Table 3).

## s6. Conclusions and Future Extensions

We have shown that the proposed shrinkage estimators can be more precise than the standard unbiased estimators for estimating individual means in meta-analyses. Theoretical evidence (Section 3) and simulation studies (Section 4) identified the remarkable advantages of the proposed estimators under certain scenarios of the true individual means. However, no single estimator is the best for all the scenarios. Hence, it is important to choose a suitable estimator for each scenario. We have provided two data examples to demonstrate the scenarios under which the proposed estimators are beneficial (Section 5). However, the choice of estimators depends on the goal of research, and hence, it is difficult to set a purely statistical strategy to choose the estimator. What we have concluded in this

article is that there is good potential for the proposed estimators by their careful applications to real meta-analyses.

Recall that the shrinkage effects on the general pretest estimators,  $(\delta_1^{\text{GPT}}, \dots, \delta_G^{\text{GPT}})$ , are component-wise, and hence, do not involve the JS-type shrinkage as in Sections 3.1 and 3.2. In the presence of pretests, it is unclear how the JS-type shrinkage can be introduced. This issue is an interesting topic, yet it is beyond the scope of the article.

This article focused on the shrinkage toward zero because a meta-analysis is typically a collection of individual studies of weak effects. However, other shrinkage schemes can also be considered, such as the shrinkage toward an arbitrary mean vector, a truncated space, a polyhedral convex, or a positive halfplane [12,25,48–54]. Different shrinkage schemes could be considered depending on the goals of meta-analyses. Therefore, there is good potential for shrinkage estimation methods to work with meta-analyses.

This article focuses on univariate response. One may consider extensions of this article to multiple (e.g., bivariate) responses. For instance, educational research may involve meta-analyses of bivariate test scores on verbal and mathematics tests [55], and mathematics and statistics tests [44]. In our data example of gastric cancer patients (Section 5.1), bivariate correlated endpoints are of great interest [55]. Correlations between responses should be employed to improve the efficiency and reduce the bias of univariate analyses [56–65] and to predict a primary outcome by the secondary outcome [66–69]. Multivariate shrinkage estimators of multivariate normal means, such as [70,71], can be considered for this extension.

One may also consider extensions of this article to non-normal (e.g., gamma distributed) responses. For non-normal distributions, shrinkage estimators exist, such as [72] for gamma distribution. Thus, introduction of shrinkage estimators to multivariate and/or non-normal models, and even their regression models, is a promising future direction.

Finally, we note that all meta-analyses are valid under the questionable assumption of no publication bias. Frequentist and Bayesian models for checking the publication biases are often considered [73,74]. It would be interesting to see if the proposed shrinkage methods can mitigate the effect of publication biases.

**Author Contributions:** Conceptualization, N.T. and T.E.; methodology, N.T. and T.E.; data curation, N.T. and T.E.; writing, N.T., T.E., Y.K. and Y.-T.C.; supervision, T.E., Y.K., and Y.-T.C.; funding acquisition, Y.-T.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** Chang YT is financially supported by JSPS KAKENHI Grant Number JP26330047 and JP18K11196.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank referees for their valuable suggestions that improved the article. The authors kindly thank Leila Zhang from MDPI for her invitation to submit our work to *Axioms* free of article processing charge.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Derivations of the JS Estimator

Define  $\mathbf{X} \equiv (Y_1/\sigma_1, \dots, Y_G/\sigma_G)$  and  $\boldsymbol{\theta} \equiv (\theta_1, \dots, \theta_G) \equiv (\mu_1/\sigma_1, \dots, \mu_G/\sigma_G)$ . Then, we obtain  $\mathbf{X} \sim N_G(\boldsymbol{\theta}, \mathbf{I}_G)$ , where  $\mathbf{I}_G$  is the identity matrix of dimension  $G$ . The random vector  $\mathbf{X}$  having the unit variance is the traditional setup of the simultaneous inference (pp. 346–355 of [25]). Now,  $\mathbf{X}$  is a maximum likelihood estimator of  $\boldsymbol{\theta}$ . Consider an arbitrary estimator  $\boldsymbol{\theta}(\mathbf{X}) \equiv (\theta_1(\mathbf{X}), \dots, \theta_G(\mathbf{X}))$  of  $\boldsymbol{\theta}$  under the square error loss  $\|\boldsymbol{\theta}(\mathbf{X}) - \boldsymbol{\theta}\|^2 \equiv \sum_{i=1}^G |\theta_i(\mathbf{X}) - \theta_i|^2$ . Let the JS estimator be

$$\boldsymbol{\theta}^{\text{JS}}(\mathbf{X}) \equiv \left(1 - \frac{G-2}{\sum_{i=1}^G X_i^2}\right) \mathbf{X}.$$

It can be shown that

$$E_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^{\text{JS}}(\mathbf{X}) - \boldsymbol{\theta}\|^2 = G - (G-2)^2 E_{\boldsymbol{\theta}} \left[ \frac{1}{\chi_G^2(\lambda)} \right],$$

where  $\chi_G^2(\lambda)$  is the noncentral  $\chi^2$ -distribution with the noncentral parameter  $\lambda = \|\boldsymbol{\theta}\|^2$ ; see p. 355 of [25] for the proofs. From this equation,

$$E_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^{\text{JS}}(\mathbf{X}) - \boldsymbol{\theta}\|^2 < G = E_{\boldsymbol{\theta}} \|\mathbf{X} - \boldsymbol{\theta}\|^2,$$

for  $G \geq 3$ . Thus,  $\boldsymbol{\theta}^{\text{JS}}(\mathbf{X})$  improves upon  $\mathbf{X}$  for  $G \geq 3$ . Rewriting the above results, we arrive at

$$E_{\boldsymbol{\theta}} \|\boldsymbol{\theta}^{\text{JS}}(\mathbf{X}) - \boldsymbol{\theta}\|^2 = E_{\boldsymbol{\theta}} \left[ \sum_{i=1}^G \frac{(\delta_i^{\text{JS}}(\mathbf{Y}) - \mu_i)^2}{\sigma_i^2} \right] < E \left[ \sum_{i=1}^G \frac{(Y_i - \mu_i)^2}{\sigma_i^2} \right] = G, \quad \forall (\mu_1, \dots, \mu_G).$$

#### Appendix A.2. Derivations of the Empirical Bayes Estimator

Under the prior  $\mu_i \sim N(0, \sigma_i^2 \tau^2)$ , the Bayes estimator is derived as

$$\boldsymbol{\delta}^{\text{Bayes}} = E(\boldsymbol{\mu}|\mathbf{Y}) = \left(1 - \frac{1}{1 + \tau^2}\right) \mathbf{Y}.$$

It can be shown that the marginal distribution of  $Y_i$  is  $Y_i \sim N(0, \sigma_i^2 + \sigma_i^2 \tau^2)$ . Thus,  $\sum_{i=1}^G Y_i^2 / (\sigma_i^2 + \sigma_i^2 \tau^2) = 1/(1 + \tau^2) \sum_{i=1}^G Y_i^2 / \sigma_i^2 \sim \chi_G^2$ , a central  $\chi^2$ -distribution with  $G$  degrees of freedom. Taking the inverse,

$$\frac{1 + \tau^2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2} \sim \frac{1}{\chi_G^2}.$$

Taking the expectation for left and right sides,

$$E \left[ \frac{1 + \tau^2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2} \right] = \frac{1}{G-2}.$$

Thus, by the method of moment estimation,

$$\left( \frac{1}{1 + \tau^2} \right) = \frac{G-2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2}.$$

This estimator is unbiased since

$$E \left( \frac{1}{1 + \tau^2} \right) \equiv E \left[ \frac{G-2}{\sum_{i=1}^G Y_i^2 / \sigma_i^2} \right] = \frac{1}{1 + \tau^2} E \left[ \frac{G-2}{\sum_{i=1}^G Y_i^2 / (\sigma_i^2 + \sigma_i^2 \tau^2)} \right] = \frac{1}{1 + \tau^2}.$$

#### Appendix A.3. Proof for the Pretest Estimators

The proofs for the WMSE and TMSE are similar. Thus, we focus on the latter. Then, it suffices to show

$$E[(Y_i - \mu_i)^2] > E[(\delta_i^{\text{GPT}} - \mu_i)^2] \quad \forall i \text{ for } \mu_i \approx 0.$$

It follows that

$$E[(Y_i - \mu_i)^2] = E[(Y_i - \mu_i)^2 I(|Y_i| > \sigma_i z_{\alpha_1})] + E[(Y_i - \mu_i)^2 I(\sigma_i z_{\alpha_2} < |Y_i| \leq \sigma_i z_{\alpha_1})] + E[(Y_i - \mu_i)^2 I(|Y_i| \leq \sigma_i z_{\alpha_2})]$$

There exists a small value  $\varepsilon_i > 0$  such that for  $\forall \mu_i \in (-\varepsilon_i, \varepsilon_i)$

$$E[(Y_i - \mu_i)^2 I(|Y_i| \leq \sigma_i z_{\alpha_2})] > E[(-\mu_i)^2 I(|Y_i| \leq \sigma_i z_{\alpha_2})],$$

$$E[(Y_i - \mu_i)^2 I(\sigma_i z_{\alpha_2} < |Y_i| \leq \sigma_i z_{\alpha_1})] > E[(q(Y_i/\sigma_i)Y_i - \mu_i)^2 I(\sigma_i z_{\alpha_2} < |Y_i| \leq \sigma_i z_{\alpha_1})]$$

Thus, for  $\forall \mu_i \in (-\varepsilon_i, \varepsilon_i)$ ,

$$\begin{aligned} E[(Y_i - \mu_i)^2] &> E[(Y_i - \mu_i)^2 I(|Y_i| > \sigma_i z_{\alpha_1})] + E\left[\left(q\left(\frac{Y_i}{\sigma_i}\right)Y_i - \mu_i\right)^2 I(\sigma_i z_{\alpha_2} < |Y_i| \leq \sigma_i z_{\alpha_1})\right] + E[(-\mu_i)^2 I(\sigma_i z_{\alpha_2} \geq |Y_i|)] \\ &= E\left\{Y_i I\left(\left|\frac{Y_i}{\sigma_i}\right| > z_{\alpha_1}\right) + q\left(\frac{Y_i}{\sigma_i}\right)Y_i I\left(z_{\alpha_2} < \left|\frac{Y_i}{\sigma_i}\right| \leq z_{\alpha_1}\right) - \mu_i\right\}^2 = E[(\delta_i^{\text{GPT}} - \mu_i)^2]. \end{aligned}$$

## References

- Borenstein, M.; Hedges, L.V.; Higgins, J.P.; Rothstein, H.R. *Introduction to Meta-Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
- Kaiser, T.; Menkhoff, L. Financial education in schools: A meta-analysis of experimental studies. *Econ. Educ. Rev.* **2020**, *78*, 101930.
- Leung, Y.; Oates, J.; Chan, S.P. Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. *J. Speech Lang. Hear. Res.* **2018**, *61*, 266–297.
- DerSimonian, R.; Laird, N. Meta-analysis in clinical trials revisited. *Contemp. Clin. Trials* **2015**, *45*, 139–145.
- Fleiss, J.L. Review papers: The statistical basis of meta-analysis. *Stat. Methods Med. Res.* **1993**, *2*, 121–145.
- Batra, K.; Singh, T.P.; Sharma, M.; Batra, R.; Schvaneveldt, N. Investigating the psychological impact of COVID-19 among healthcare workers: A meta-analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9096.
- Pranata, R.; Lim, M.A.; Huang, I.; Raharjo, S.B.; Lukito, A.A. Hypertension is associated with increased mortality and severity of disease in COVID-19 pneumonia: A systematic review, meta-analysis and meta-regression. *J. Renin-Angiotensin-Aldosterone Syst.* **2020**, *21*, 1470320320926899.
- Wang, Y.; Kala, M.P.; Jafar, T.H. Factors associated with psychological distress during the coronavirus disease 2019 (COVID-19) pandemic on the predominantly general population: A systematic review and meta-analysis. *PLoS ONE* **2020**, *15*, e0244630.
- Rice, K.; Higgins, J.P.; Lumley, T. A re-evaluation of fixed effect (s) meta-analysis. *J. R. Stat. Soc. Ser. A* **2018**, *181*, 205–227.
- Lehmann, E.L. *Elements of Large-Sample Theory*; Springer Science & Business Media: Berlin, Germany, 2010.
- Shinozaki, N.; Chang, Y.T. Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances. *Commun. Stat.-Theor. Methods* **1993**, *8*, 2147–2169.
- Shinozaki, N.; Chang, Y.T. Minimaxity of empirical Bayes estimators shrinking toward the grand mean when variances are unequal. *Commun. Stat.-Theor. Methods* **1996**, *25*, 183–199.
- Singh, H.P.; Vishwakarma, G.K. A family of estimators of population mean using auxiliary information in stratified sampling. *Commun. Stat.-Theor. Methods* **2008**, *37*, 1038–1050.
- DerSimonian, R.; Laird, N. Meta-analysis in clinical trials. *Control Clin. Trials* **1986**, *7*, 177–188.
- Kontopantelis, E.; Reeves, D. Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Stat. Methods Med. Res.* **2012**, *21*, 409–426.
- Raudenbush, S.W.; Bryk, A.S. Empirical bayes meta-analysis. *J. Educ. Stat.* **1985**, *10*, 75–98.
- Schmid, C. Using bayesian inference to perform meta-analysis. *Eval. Health Prof.* **2001**, *24*, 165–189.
- Röver, C.; Friede, T. Dynamically borrowing strength from another study through shrinkage estimation. *Stat. Methods Med. Res.* **2020**, *29*, 293–308.
- Röver, C.; Friede, T. Bounds for the weight of external data in shrinkage estimation. *Biom. J.* **2021**, *63*, 1131–1143.
- Malekzadeh, A.; Kharrati-Kopaei, M. Inferences on the common mean of several normal populations under hetero-scedasticity. *Comput. Stat.* **2018**, *33*, 1367–1384.
- Shinozaki, N. A note on estimating the common mean of k normal distributions and the stein problem. *Commun. Stat.-Theory Methods* **1978**, *7*, 1421–1432.
- Chen, Z.; Zhang, G.; Li, J. Goodness-of-fit test for meta-analysis. *Sci. Rep.* **2015**, *5*, 16983.
- Everitt, B. *Modern Medical Statistics: A Practical Guide*; Wiley: Hoboken, NJ, USA, 2003.
- Lin, L. Hybrid test for publication bias in meta-analysis. *Stat. Methods Med. Res.* **2020**, *29*, 2881–2899.
- Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
- Shao, J. *Mathematical Statistics*; Springer: New York, NY, USA, 2003.
- Johnstone, I.M. On minimax estimation of a sparse normal mean vector. *Ann. Stat.* **1994**, *22*, 271–289.
- van der Pas, S.; Salomond, J.-B.; Schmidt-Hieber, J. Conditions for posterior contraction in the sparse normal means problem. *Electron. J. Stat.* **2016**, *10*, 976–1000.
- James, W.; Stein, C. Estimation with quadratic loss. In *Breakthroughs in Statistics*; Springer: New York, NY, USA, 1992; Volume 1, pp. 443–460.
- van Eeden, C. *Restricted Parameter Space Estimation Problems*; Springer: New York, NY, USA, 2006.
- Li, W.; Li, R.; Feng, Z.; Ning, J. Semiparametric isotonic regression analysis for risk assessment under nested case-control and case-cohort designs. *Stat. Methods Med. Res.* **2020**, *29*, 2328–2343.
- Robertson, T.; Wright, F.T.; Dykstra, R. *Order Restricted Statistical Inference*; Wiley: Chichester, UK, 1988.
- Silvapulle, M.J.; Sen, P.K. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*; Wiley: Hoboken, NJ, USA, 2005.

34. Tsukuma, H. Simultaneous estimation of restricted location parameters based on permutation and sign-change. *Stat. Pap.* **2012**, *53*, 915–934.
35. Chang, Y.T. Stein-type estimators for parameters restricted by linear inequalities. *Keio Sci. Technol. Rep.* **1981**, *34*, 83–95.
36. Bancroft, T.A. On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Stat.* **1944**, *15*, 190–204.
37. Judge, G.G.; Bock, M.E. *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*; Elsevier: Amsterdam, The Netherlands, 1978.
38. Khan, S.; Saleh, A.K.M.E. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. *Stat. Pap.* **2001**, *42*, 451–473.
39. Magnus, J.R. The traditional pretest estimator. *Theory Probab. Its Appl.* **2000**, *44*, 293–308.
40. Magnus, J.R.; Wan, A.T.; Zhang, X. Weighted average least squares estimation with nonspherical disturbances and an application to the Hong Kong housing market. *Comput. Stat. Data Anal.* **2011**, *55*, 1331–1341.
41. Shih, J.-H.; Konno, Y.; Chang, Y.-T.; Emura, T. A class of general pretest estimators for the univariate normal mean. *Commun. Stat. -Theory Methods* **2021**, doi:10.1080/03610926.2021.1955384.
42. Shih, J.-H.; Lin, T.-Y.; Jimichi, M.; Emura, T. Robust ridge M-estimators with pretest and Stein-rule shrinkage for an intercept term. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 107–150.
43. Kibria, B.G.; Saleh, A.M.E. Optimum critical value for pre-test estimator. *Commun. Stat. -Simul. Comput.* **2006**, *35*, 309–319.
44. Shih, J.-H.; Konno, Y.; Chang, Y.-T.; Emura, T. Estimation of a common mean vector in bivariate meta-analysis under the FGM copula. *Statistics* **2019**, *53*, 673–695.
45. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group. Role of chemotherapy for advanced/recurrent gastric cancer: An individual-patient-data meta-analysis. *Eur. J. Cancer* **2013**, *49*, 1565–1577, doi:10.1016/j.ejca.2012.12.016.
46. Oba, K.; Paoletti, X.; Alberts, S.; Bang, Y.J.; Benedetti, J.; Bleiberg, H.; Catalano, P.; Lordick, F.; Michiels, S.; Morita, S.; et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. *J. Natl. Cancer Inst.* **2013**, *105*, 1600–1607.
47. Rotolo, F.; Paoletti, X.; Michiels, S. Surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. *Comput. Methods Programs Biomed.* **2018**, *155*, 189–198.
48. Kubokawa, T. Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Commun. Stat. -Theory Methods* **1999**, *28*, 613–650.
49. Fourdrinier, D.; Strawderman, W.E.; Wells, M.T. *Shrinkage Estimation*; Springer: New York, NY, USA, 2018.
50. Fourdrinier, D.; Ouassou, I.; Strawderman, W.E. Estimation of a parameter vector when some components are re-stricted. *J. Multivar. Anal.* **2003**, *86*, 14–27.
51. Kuriki, S.; Takemura, A. Shrinkage estimation towards a closed convex set with a smooth boundary. *J. Multivar. Anal.* **2000**, *75*, 79–111.
52. Tripathi, Y.M.; Kumar, S. Estimating a positive normal mean. *Stat. Pap.* **2007**, *48*, 609–629.
53. Tsukuma, H. Shrinkage estimation in elliptically contoured distribution with restricted parameter space. *Stat. Decis.* **2009**, *27*, 25–35.
54. Tsukuma, H.; Kubokawa, T. Stein's phenomenon in estimation of means restricted to a polyhedral convex cone. *J. Multivar. Anal.* **2008**, *99*, 141–164.
55. Gleser, L.J.; Olkin, L. Stochastically dependent effect sizes. In *the Handbook of Research Synthesis*; Russel Sage Foundation: New York, NY, USA, 1994.
56. Emura, T.; Matsui, S.; Rondeau, V. *Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models*; Springer: Singapore, 2019.
57. Emura, T.; Nakatochi, M.; Murotani, K.; Rondeau, V. A joint frailty-copula model between tumour progression and death for meta-analysis. *Stat. Methods Med. Res.* **2017**, *26*, 2649–2666.
58. Emura, T.; Sofeu, C.L.; Rondeau, V. Conditional copula models for correlated survival endpoints: Individual patient data meta-analysis of randomized controlled trials. *Stat. Methods Med. Res.* **2021**, doi:10.1177/09622802211046390.
59. Mavridis, D.; Salanti, G. A practical introduction to multivariate meta-analysis. *Stat. Methods Med. Res.* **2013**, *22*, 133–158.
60. Peng, M.; Xiang, L. Correlation-based joint feature screening for semi-competing risks outcomes with application to breast cancer data. *Stat. Methods Med. Res.* **2021**, 09622802211037071.
61. Riley, R.D. Multivariate meta-analysis: The effect of ignoring within-study correlation. *J. R. Stat. Soc. Ser. A* **2009**, *172*, 789–811.
62. Copas, J. B.; Jackson, D.; White, I. R.; Riley, R. D. The role of secondary outcomes in multivariate meta-analysis. *J. R. Stat. Soc. Ser. C* **2018**, *67*, 1177–1205.
63. Sofeu, C.L.; Emura, T.; Rondeau, V. A joint frailty-copula model for meta-analytic validation of failure time surrogate endpoints in clinical trials. *Biom. J.* **2021**, *63*, 423–446.
64. Wu, B.H.; Michimae, H.; Emura, T. Meta-analysis of individual patient data with semi-competing risks under the Weibull joint frailty-copula model. *Comput. Stat.* **2020**, *35*, 1525–1552.
65. Yamaguchi, Y.; Maruo, K. Bivariate beta-binomial model using Gaussian copula for bivariate meta-analysis of two binary outcomes with low incidence. *Jpn. J. Stat. Data Sci.* **2019**, *2*, 347–373.
66. Emura, T.; Nakatochi, M.; Matsui, S.; Michimae, H.; Rondeau, V. Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model. *Stat. Methods Med. Res.* **2018**, *27*, 2842–2858.

67. Kawakami, R.; Michimae, H.; Lin, Y.-H. Assessing the numerical integration of dynamic prediction formulas using the exact expressions under the joint frailty-copula model. *Jpn. J. Stat. Data Sci.* **2021**, 1–29, doi:10.1007/s42081-021-00133-z.
68. Noughabi, M.S.; Kayid, M. Bivariate quantile residual life: A characterization theorem and statistical properties. *Stat. Pap.* **2019**, *60*, 2001–2012.
69. Shinohara, S.; Lin, Y.-H.; Michimae, H.; Emura, T. Dynamic lifetime prediction using a Weibull-based bivariate failure time model: A meta-analysis of individual-patient data. *Commun. Stat. -Simul. Comput.* **2020**, 1–20, doi:10.1080/03610918.2020.1855449.
70. Bilodeau, M.; Kariya, T. Minimax estimators in the normal MANOVA model. *J. Multivar. Anal.* **1989**, *28*, 260–270.
71. Konno, Y. On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivar. Anal.* **1991**, *36*, 44–55.
72. Vishwakarma, G.K.; Gupta, S. Shrinkage estimator for scale parameter of gamma distribution. *Commun. Stat. -Simul. Comput.* **2020**, 1–8, doi:10.1080/03610918.2019.1705976.
73. Copas, J.B.; Shi, J.Q. A sensitivity analysis for publication bias in systematic reviews. *Stat. Methods Med Res.* **2001**, *10*, 251–265.
74. Larose, D.T.; Dey, D.K. Modeling publication bias using weighted distributions in a Bayesian framework. *Comput. Stat. Data Anal.* **1998**, *26*, 279–302.