



# On the use of historical estimates

Ori Davidov<sup>1</sup> · Tamás Rudas<sup>2</sup> 

Received: 12 August 2022 / Revised: 8 November 2022 / Published online: 6 January 2023  
© The Author(s) 2023

## Abstract

The use of historical, i.e., already existing, estimates in current studies is common in a wide variety of application areas. Nevertheless, despite their routine use, the uncertainty associated with historical estimates is rarely properly accounted for in the analysis. In this communication, we review common practices and then provide a mathematical formulation and a principled frequentist methodology for addressing the problem of drawing inferences in the presence of historical estimates. Three distinct variants are investigated in detail; the corresponding limiting distributions are found and compared. The design of future studies, given historical data, is also explored and relations with a variety of other well-studied statistical problems discussed.

**Keywords** Bliss-independence · Double-sampling · Loewner order

## 1 Introduction

There are many circumstances in which a statistical analysis either requires, or can greatly benefit, from the use of historical, that is existing, information. In this paper we focus on the situation where the historical information consists of parameter estimates. These may be essential for model fitting but impossible, and/or very expensive, to collect in the context of the current study. Although related, we will not explicitly discuss the large literature on data combination schemes or other two-stage plug-in methods. An example of the former is Ridder and Moffitt (2007) comprehensive review of methodologies for data combination common in econometrics whereas an example of the latter is Genest et al. (1995) seminal paper on two-step semi-parametric inference in copula models. These and similar problems have received considerable attention and have a long history in Statistics, see, e.g., Cochran (1954).

---

✉ Tamás Rudas  
trudas@elte.hu

<sup>1</sup> Department of Statistics, University of Haifa, Mount Carmel, 3498838 Haifa, Israel

<sup>2</sup> Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University, Pázmány Péter sétány 1/A, Budapest, Hungary

Historical estimates are used in a variety of applications in the social, physical, and biomedical sciences. For example, some models for the spread of infectious diseases, such as the SIR model (Becker 2017) popularized in connection with COVID-19, e.g., Cooper et al. (2020), require the input of age-specific transmission parameters which can be estimated from social contact networks (Edmunds et al. 1997; Wallinga et al. 2006) and then used to fit epidemic models (Mossong et al. 2008; Goeyvaerts et al. 2010, Yaari et al. 2018). Another interesting application is the optimization of cancer treatment where Kronik et al. (2010) develop a framework for predicting the outcome of prostate cancer immunotherapy by fitting personalized mathematical models. Their model consists of a set of differential equations whose behavior is governed by a collection of parameters, some of which are global parameters while others are subject specific. The values of the global parameters were obtained from at least ten different published studies, see their Table 2, whereas the subject level parameters were estimated by fitting a model to each participant assuming that the global parameters were estimated without error. See Kogan et al. (2012) and Kozłowska et al. (2018) for similar applications. It is worth noting that the applications above may be viewed as a model for situations in which knowledge collected in one setting, experimental or observational, is then used to estimate quantities arising in a different experiment and is quite common in the biomedical sciences. See Lee and Zelen (1998) and Davidov and Zelen (2004) for a similar structure arising in the planning of early detection programs.

Another very important application in which historical estimates are used is clinical trials. Consider, for example, the situation in which the effect of a combination of treatments is assessed (e.g., Tamma et al. 2012, Kanda et al. 2016). In such cases there exists a collection of therapies which have been independently proven to be somewhat successful at treating a medical condition. The objective of a new study may then be to assess whether a combination of these therapies provides an even better outcome. In the simplest case, one may view this problem as a three armed clinical trial comparing treatments  $A$ ,  $B$  and  $A + B$  in which historical estimates on the efficacy of treatments  $A$  and  $B$  already exist. An important example of such situations is the Food and Drug Administration (2006) guidelines for submitting applications for approval of fixed dose combinations, i.e., co-packaged drug products, of previously approved antiretrovirals for the treatment of HIV. In particular, Attachment A of the aforementioned document considers the scenario in which a non-innovator, i.e., a generic drug company, wants to obtain approval for a combination of already approved ingredients. In this case, only efficacy data for the combination needs to be submitted. We will revisit and thoroughly analyze two forms of this example later on. More broadly, the use of historical data in the contexts of clinical trials has been investigated by numerous researchers using multiple perspectives, cf., Pocock (1976), Peto et al. (1976), Neuenschwander et al. (2010), Viele et al. (2014), and Piantadosi (2017) among many others. As noted by a referee, a particularly relevant class of designs are platform trials which allow adding new treatments to the experiment and thus controls may become non-concurrent, cf. Lee and Wason (2020) and Roig et al. (2022).

The use of historical estimates is also widespread in the social sciences. For example, in the fitting of some econometric models researchers may use values estimated from previously collected survey data. We point to the paper of Newey et al. (2005)

which focuses on the asymptotic bias of the estimated parameters. The complexity of using historical estimates in the social sciences is further illustrated by the work of Tasseva (2019). In a microsimulation study investigating the effect of the recent expansion in higher education in Great Britain on household inequalities, previously obtained estimates from the Family Resources Survey for Great Britain (GOV.UK 2019) were used. While sampling variability could be taken into account using bootstrap methods, as noted by the author, measurement error, inevitably present in income information collected in surveys, see, e.g., Moore et al. (2000), could not be accounted for using this method. Similarly, Doudich et al. (2016) describe an imputation-related-method for incorporating estimates obtained in labor force surveys (which are easily and cheaply conducted) into household expenditure surveys (which are much more time consuming and expensive) in order to estimate poverty rates in Morocco. Likewise, demographic model fitting and projections rely on historical data. The standard method of population projections (see United Nations 2014) is based on the combination of cohort survival rates, i.e., historical data, with current data on cohort sizes. Raftery et al. (2014) proposed a Bayesian approach to take the uncertainty associated with historical data into account. It is worth noting that in this case the uncertainty accounted for by Bayesian modeling did not come from observational errors but rather from the fact that the true population figures may have changed over time.

Researchers often do not adequately account for the variability of the historical estimates when incorporating them into a current analysis. In fact, the practice of plugging-in the estimated values for certain parameters is widespread. However, this practice is often not disclosed as many practitioners view this strategy as a natural way of “doing things”. Consequently, the objectives and contributions of this communication are twofold: first, we draw attention to current practice, and secondly, and more importantly, we provide a principled methodology for incorporating historical estimates into a current analysis. Surprisingly, despite the ubiquity of historical data and estimates and the many papers that touch on various aspects thereof, a general methodology discussing the use of historical estimates, as given here, has been thus far lacking. In particular we consider two broad settings in which historical estimates are employed. Three such estimators are presented in Sect. 3 and one in Sect. 4.1. Their limiting distributions are found and a theoretical analysis comparing their precision is conducted. When comparable, a preference order among the different approaches is established. Our paper goes beyond the existing knowledge by providing an inventory of ways in which historical estimates can be used, and by quantifying the properties of the resulting estimators. We also demonstrate how these results may be used in the design of experiments.

The paper is organized as follows. Our notation and formulation are outlined in Sect. 2. Section 3 provides our main theoretical findings which include the limiting distributions of the estimates in the presence of historical estimates and a comparison thereof. In Sect. 4 two applications are described in conjunction with accompanying numerical experiments. The first application addresses the two-way analysis of variance (ANOVA) problem introduced in Sect. 2. The second, related application, deals with a drug interaction study within the framework of Bliss-independence (Bliss 1939), an old concept which has garnered much recent attention. We conclude with a discussion in Sect. 5. All proofs are collected in an Appendix.

## 2 Notation and formulation

Consider a designed experiment or observational study, denoted by  $\mathcal{S}$ , in which data  $\mathcal{D}$  consisting of  $n$  observations are collected. Usually, the observations are independent and identically distributed. Suppose further that the model describing the distribution of  $\mathcal{D}$  is indexed by  $\omega^T = (\theta^T, \eta^T)$  where  $\theta \in \mathbb{R}^p$  and  $\eta \in \mathbb{R}^q$  is the concatenation of  $\eta_1 \in \mathbb{R}^{q_1}, \dots, \eta_K \in \mathbb{R}^{q_K}$  with  $q = q_1 + \dots + q_K$ . Let  $\Phi(\omega)$  be some function of the model parameters which is of interest to the researchers. Clearly,  $\Phi(\omega)$  may be a function of  $\theta$  alone,  $\eta$  alone or of both  $\theta$  and  $\eta$ . The primary goal of the study  $\mathcal{S}$ , which we refer to as the current study, is inference on  $\Phi(\omega)$  in the presence of historical data which we view as a collection of  $K$ , independent estimates  $\hat{\eta}_1, \dots, \hat{\eta}_K$  obtained from historical studies  $\mathcal{S}_1, \dots, \mathcal{S}_K$  of sizes  $m_1, \dots, m_K$  and  $m = m_1 + \dots + m_K$  denotes the total sample size in the historical studies.

In some circumstances, it may not be possible to estimate  $\omega$  using the data  $\mathcal{D}$ . However, if  $\eta$  were known in advance then it would be possible to estimate  $\theta$ . As an example, such a situation would arise if the model  $f(\cdot; \theta, \eta)$  is not identifiable whereas the model  $f(\cdot; \theta, \eta_0)$  is identifiable for every fixed value of  $\eta_0$ . In other circumstances given the data  $\mathcal{D}$  both  $\theta$  and  $\eta$  are estimable (e.g., Peddada et al. 2007). Thus, in this communication we consider two distinct settings, the second of which has two variants. In the first setting, referred to as a Type I Problem, only the parameter  $\theta$  is estimable using the data  $\mathcal{D}$ , while  $(\eta_1, \dots, \eta_K)$  are fixed at their historical estimated values  $(\hat{\eta}_1, \dots, \hat{\eta}_K)$ . In the second setting, referred to as Type II Problem, both  $\theta$  and  $\eta$  are estimable using  $\mathcal{D}$  and a two-step procedure is utilized to estimate  $\theta$  while updating the estimates for  $(\eta_1, \dots, \eta_K)$ . It may also happen that the available data corresponds to a Type II problem and while a Type II analysis would be possible, the researcher may decide to conduct a Type I Analysis, i.e., estimate  $\theta$  as if the data came from a Type I Problem. One of our results shows that this is an inferior strategy, i.e., if the data  $\mathcal{D}$  identifies  $\omega$  it is always advisable to re-estimate  $\eta$  and the loss of precision is quantified in terms of a simple decomposition of the variance matrices of the resulting estimators. It is also important to emphasize that there are situations in which the investigator, by means of the design of the study  $\mathcal{S}$ , may control whether the problem is of Type I or a Type II.

To fix ideas consider the two-way ANOVA model in which the expected value of an outcome  $Y$  is given by

$$\mathbb{E}(Y|T_1, T_2) = \eta_0 + \eta_1 T_1 + \eta_2 T_2 + \theta T_1 T_2 \quad (1)$$

where for  $i = 1, 2$ ,  $T_i \in \{0, 1\}$  indicates whether treatment  $i$  is administered. Here  $\eta_0$  denotes the mean of  $Y$  when neither treatment is administered,  $\eta_i$  models the marginal increase in the expectation of  $Y$  when treatment  $i$  is administered and  $\theta$  models the interaction  $T_1 \times T_2$ . Suppose, now that the historical data consists of two studies  $\mathcal{S}_1$  and  $\mathcal{S}_2$  of sizes  $m_1$  and  $m_2$ , respectively, where in the study  $\mathcal{S}_i$  treatment  $i$  was compared with a control. Clearly the historical data provides no information on  $\theta$ . Thus inference on  $\theta$  would require a new study  $\mathcal{S}$  in which  $T_1 = T_2 = 1$  for some subset of the observations. For simplicity, *interchangeability* is assumed, i.e., all experimental units, in  $\mathcal{S}_1$  and  $\mathcal{S}_2$  as well as  $\mathcal{S}$ , are assumed to be drawn from the same

population, e.g., Peddada et al. (2007), and therefore any change in the mean response may be attributed solely to the treatment combination received. The assumption of *interchangeability* may be relaxed as discussed in Sect. 5.

One objective of this communication is to provide a methodology for effective design and analysis of a new study  $\mathcal{S}$  of size  $n$  which allows the estimation of  $\theta$  and utilizes the historical estimates of  $(\eta_0, \eta_1, \eta_2)$  obtained from  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . Depending on its objectives, the study  $\mathcal{S}$  may be of various forms. For example, one may choose to allocate all  $n$  observations to receive both treatments, i.e.,  $T_1 = T_2 = 1$ . In this case, the data  $\mathcal{D}$  is an IID sample of observations with mean  $\eta_0 + \eta_1 + \eta_2 + \theta$  and variance  $\sigma^2$ . Although the parameter  $\theta$  is not identifiable from  $\mathcal{D}$  alone it is estimable given the historical data, so this is clearly a Type I Problem. Alternatively, if  $\mathcal{S}$  allocates observation to all treatment combinations then  $\theta$  as well as  $(\eta_0, \eta_1, \eta_2)$  are estimable from  $\mathcal{S}$  and this falls within the framework of a Type II Problem. This example will be further analyzed in Sect. 4.1.

### 3 Results

Our main theoretical findings, i.e., Theorems 3.1, 3.2, and 3.4 describe the limiting distributions of estimators for  $\omega$  which are then compared in Theorems 3.3, 3.5 and 3.6. Remark 7 provides a brief summary of the results of this Section.

#### 3.1 Type I problems

Suppose first that we are in the setting of a Type I Problem. Recall that in such circumstances only  $\theta$  is estimated while  $(\eta_1, \dots, \eta_K)$  are fixed at their historical values. Thus, let  $\bar{\theta}_A$  solve

$$\Psi(\theta, \hat{\eta}) = \mathbf{0} \tag{2}$$

where  $\hat{\eta} = (\hat{\eta}_1^T, \dots, \hat{\eta}_K^T)^T$ . The estimating Eq. (2) may be a score equation motivated by likelihood theory, a generalized estimating equation derived by quasi-likelihood or any other statistical estimation framework. Observe that the solution  $\bar{\theta}_A$  of (2) is obtained by plugging-in the sample values of the  $K$  independent estimators  $\hat{\eta}_1, \dots, \hat{\eta}_K$ . For simplicity we may further assume that the data  $\mathcal{D}$  is a random sample  $Y_1, \dots, Y_n$  and (2) is of the form

$$\Psi(\theta, \hat{\eta}) = n^{-1} \sum_{i=1}^n \psi(\theta, \hat{\eta}, Y_i).$$

The function  $\psi$  is assumed to be: (i) continuously differentiable with respect to both  $\theta$  and  $\eta_1, \dots, \eta_K$ ; it is further assumed to satisfy (ii)  $\mathbb{E}_0(\psi) = \mathbf{0}$ ; (iii)  $\mathbb{E}_0(\psi \psi^T) < \infty$ ; (iv) the matrix  $\mathbb{E}_0(\partial \psi / \partial \eta)$  exists; and (v) the matrix  $\mathbb{E}_0(\partial \psi / \partial \theta)$  exists and is invertible. Here  $\mathbb{E}_0(\cdot)$  denotes the expectation taken at  $\omega_0 = (\theta_0, \eta_0) = (\theta_0, \eta_{1,0}, \dots, \eta_{K,0})$ , the true value of all parameters. Conditions (i) – (v) are all stan-

dard regularity conditions often imposed in the literature (cf., Heyde 2008, Van der Vaart 2000). We now have the following:

**Theorem 3.1** *Let  $\bar{\theta}_A$  be a solution to (2) and set  $\bar{\eta}_A = \widehat{\eta}$ . Assume that: (i)  $\bar{\theta}_A$  is consistent at  $\omega_0$ ; (ii) the estimating function  $\psi$  satisfies the regularity conditions listed above; and (iii) the historical estimates satisfy  $\sqrt{m_j}(\widehat{\eta}_j - \eta_{j,0}) \Rightarrow \mathcal{N}_{q_j}(\mathbf{0}, \Sigma_j)$  and are independent of each other and of the current study. Then if  $(m/m_j) \rightarrow \kappa_j < \infty$  for all  $j = 1, \dots, K$  as  $m_j \rightarrow \infty$  and  $n/m \rightarrow \rho \in (0, \infty)$  as  $n \rightarrow \infty$  we have*

$$\sqrt{n}(\bar{\theta}_A - \theta_0, \bar{\eta}_A - \eta_0)^T \Rightarrow \mathcal{N}_{p+q}(\mathbf{0}, A)$$

where

$$A = \begin{pmatrix} A_{\theta\theta} & A_{\theta\eta} \\ A_{\eta\theta} & A_{\eta\eta} \end{pmatrix}$$

with

$$\begin{aligned} A_{\theta\theta} &= (D_{\theta_0}^{-1})[\Sigma_\psi + \rho D_{\eta_0} \Sigma D_{\eta_0}^T](D_{\theta_0}^{-1})^T, \\ A_{\theta\eta} &= -\rho(D_{\theta_0})^{-1}D_{\eta_0} \Sigma, \\ A_{\eta\eta} &= \rho \Sigma. \end{aligned}$$

where  $D_{\theta_0} = \mathbb{E}_0(\partial\psi/\partial\theta)$ ,  $D_{\eta_0} = \mathbb{E}_0(\partial\psi/\partial\eta)$ ,  $\Sigma_\psi = \mathbb{E}_0(\psi\psi^T)$  and  $\Sigma = \text{BlockDiag}(\kappa_1 \Sigma_1, \dots, \kappa_K \Sigma_K)$ .

**Remark 1** Clearly,  $A_{\theta\theta}$  is the  $p \times p$  asymptotic variance matrix of  $\bar{\theta}_A$ ,  $A_{\eta\eta}$  is the  $q \times q$  asymptotic variance matrix of  $\bar{\eta}_A$  and  $A_{\theta\eta} = A_{\eta\theta}^T$  is their  $p \times q$  asymptotic covariance matrix.

**Remark 2** As pointed out by a referee the application of Theorem 3.1 is predicated on the fact that the historical studies report the estimates of the variance matrices  $\Sigma_1, \dots, \Sigma_K$ .

The proof of Theorem 3.1 is a straightforward, but somewhat involved, application of the delta method. In contrast with Randles (1982) and Pierce (1982) which describe the limiting distribution of statistics that are explicit functions of estimated parameters, the estimator  $\bar{\theta}_A$  is an implicit function of  $\bar{\eta}_A$ . For a related but less general result see Benichou and Gail (1989). Further note that  $(D_{\theta_0}^{-1})\Sigma_\psi(D_{\theta_0}^{-1})^T$  is the asymptotic variance of  $\bar{\theta}_A$  when the true values of  $\eta_1, \dots, \eta_K$  are known in advance. Thus the term

$$\rho D_{\eta_0} \Sigma D_{\eta_0}^T$$

may be viewed as the penalty for substituting estimates for the true values of the parameters. The penalty may also be rewritten as  $\rho \sum_{j=1}^K \kappa_j D_j \Sigma_j D_j^T$  where  $D_j = \mathbb{E}_0(\partial\psi/\partial\eta_j)$  which expresses its dependence on the relative sample sizes,

the asymptotic variances of the historical estimators and the sensitivity of the estimation procedure with respect to the historical estimates, embodied in the matrices  $D_1, \dots, D_K$ .

**Remark 3** Note that if  $\rho$  is very small which occurs when  $m \gg n$ , then the penalty is inconsequential, i.e., the asymptotic variance of  $\tilde{\theta}_A$  is close to its variance when  $\eta_1, \dots, \eta_K$  are fully known.

### 3.2 Type II problems

Next, consider the case where both  $\theta$  and  $\eta$  are estimable using the data  $\mathcal{D}$  observed in the current study  $\mathcal{S}$ . Further assume that two estimating functions  $\Psi$  and  $\Gamma$  are available to us;  $\Psi$  is an estimating function for  $\theta$  given a known fixed value of  $\eta$ , as in Type I Problems, whereas  $\Gamma$  is an estimating function for  $\eta$  given a known fixed value of  $\theta$ . For example, within the likelihood framework  $\Psi$  is the score with respect to  $\theta$  while  $\Gamma$  is the score with respect to  $\eta$ .

We propose estimating  $\omega$  using a two step procedure. In the first step the data  $\mathcal{D}$  is used to obtain the pair  $(\tilde{\theta}, \tilde{\eta})^T$  which simultaneously solve

$$\Psi(\theta, \eta) = \mathbf{0} \quad \text{and} \quad \Gamma(\theta, \eta) = \mathbf{0}. \tag{3}$$

Under standard regularity conditions, cf., the conditions listed just before the statement of Theorem 3.1, the estimators  $(\tilde{\theta}, \tilde{\eta})^T$  satisfy

$$\sqrt{n}(\tilde{\theta} - \theta_0, \tilde{\eta} - \eta_0)^T \Rightarrow \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{Y}) \tag{4}$$

where  $\mathbf{Y}$  is assumed to be a non-singular variance matrix which can be consistently estimated from the data by, say  $\tilde{\mathbf{Y}}$ , the standard sandwich estimator (Van der Vaart 2000). For convenience, we may partition  $\mathbf{Y}$  as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{\theta\theta} & \mathbf{Y}_{\theta\eta} \\ \mathbf{Y}_{\eta\theta} & \mathbf{Y}_{\eta\eta} \end{pmatrix} \tag{5}$$

where  $\mathbf{Y}_{\theta\theta}$  and  $\mathbf{Y}_{\eta\eta}$  denote the marginal asymptotic variances of  $\tilde{\theta}$  and  $\tilde{\eta}$ , respectively, and  $\mathbf{Y}_{\theta\eta}$  is their asymptotic covariance. Naturally, a similar partition holds for  $\tilde{\mathbf{Y}}$ . Furthermore, as in Sect. 3.1, at our disposal are  $K$  independent historical estimates of  $\eta_1, \dots, \eta_K$  obtained using studies of sizes  $m_1, \dots, m_K$  which satisfy  $\sqrt{m_j}(\hat{\eta}_j - \eta_{j,0}) \Rightarrow \mathcal{N}_{q_j}(\mathbf{0}, \mathbf{\Sigma}_j)$ , where, again, it is assumed that  $\mathbf{\Sigma}_j$  are non-singular and can be consistently estimated for all  $j = 1, \dots, K$ . Thus

$$\sqrt{m}(\hat{\eta} - \eta_0) \Rightarrow \mathcal{N}_q(\mathbf{0}, \mathbf{\Sigma}) \tag{6}$$

where  $\mathbf{\Sigma}$  is given in the statement of Theorem 3.1. Let  $\hat{\mathbf{\Sigma}}$  be a consistent estimator of  $\mathbf{\Sigma}$ .

The historic and current estimates of  $\eta$  can be aggregated, or combined, in many ways. Lemma 2, appearing in the Appendix, suggests using the estimator

$$\bar{\eta} = (n\tilde{\Upsilon}_{\eta\eta}^{-1} + m\widehat{\Sigma}^{-1})^{-1}(n\tilde{\Upsilon}_{\eta\eta}^{-1}\tilde{\eta} + m\widehat{\Sigma}^{-1}\widehat{\eta}) \tag{7}$$

which is the MLE under normality assuming that the matrices  $\Upsilon_{\eta\eta}$  and  $\Sigma$  are known. Note that

$$\bar{\eta} = W_1\tilde{\eta} + W_2\widehat{\eta} + o_p(1)$$

where the weights  $W_1$  and  $W_2$  are the symmetric matrices

$$\begin{aligned} W_1 &= (\gamma\Upsilon_{\eta\eta}^{-1} + (1 - \gamma)\Sigma^{-1})^{-1}\gamma\Upsilon_{\eta\eta}^{-1} \quad \text{and} \\ W_2 &= (\gamma\Upsilon_{\eta\eta}^{-1} + (1 - \gamma)\Sigma^{-1})^{-1}(1 - \gamma)\Sigma^{-1} \end{aligned} \tag{8}$$

which satisfy  $I = W_1 + W_2$  with  $\gamma = \lim(n/(n + m))$ . Thus (7) differs from the best linear unbiased estimator by at most an  $o_p(1)$  term.

In the second step we find  $\bar{\theta}_B$  by solving

$$\Psi(\theta, \bar{\eta}) = 0. \tag{9}$$

where  $\bar{\eta}$  is given by (7). We now have:

**Theorem 3.2** *Let  $\bar{\theta}_B$  be a solution to (9) where  $\bar{\eta}_B = \bar{\eta}$  is given in (7). Assume that the regularity conditions of Theorem 3.1 hold. Then*

$$\sqrt{n}(\bar{\theta}_B - \theta_0, \bar{\eta}_B - \eta_0) \Rightarrow \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{B}) \tag{10}$$

where

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\theta\theta} & \mathbf{B}_{\theta\eta} \\ \mathbf{B}_{\eta\theta} & \mathbf{B}_{\eta\eta} \end{pmatrix}$$

with

$$\begin{aligned} \mathbf{B}_{\theta\theta} &= (\mathbf{D}_{\theta_0}^{-1})[\Sigma\psi + \mathbf{D}_{\eta_0}(\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}\mathbf{D}_{\eta_0}^T](\mathbf{D}_{\theta_0}^{-1})^T, \\ \mathbf{B}_{\theta\eta} &= -\mathbf{D}_{\theta_0}^{-1}\mathbf{D}_{\eta_0}(\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}, \\ \mathbf{B}_{\eta\eta} &= (\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}. \end{aligned}$$

Although the mechanics are slightly more involved, the proof of Theorem 3.2 builds on the proof of Theorem 3.1. Moreover, the structures of the asymptotic variance matrices  $\mathbf{A}$  and  $\mathbf{B}$  are analogous with the exception that the variance matrix  $\rho\Sigma$  appearing in  $\mathbf{A}$  is replaced with  $(\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}$  in  $\mathbf{B}$ .

**Remark 4** Observe that  $(\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1} \rightarrow \mathbf{0}$  as  $\rho \rightarrow 0$  so the conclusions of Remark 3 hold here as well.



It is clear that whenever the model for  $\mathcal{D}$  identifies  $\omega$ , both  $(\bar{\theta}_B, \bar{\eta}_B)$  and  $(\bar{\theta}_A, \bar{\eta}_A)$  can be computed. Next, using the concept of the Loewner order we show the former is superior to the latter. Recall that the matrix  $V_1$  is said to be smaller in the Loewner order compared with the matrix  $V_2$  if  $V_2 - V_1$  is non-negative definite (Pukelsheim 2006). This relationship is denoted by  $V_1 \preceq V_2$ . Suppose now that  $V_1$  and  $V_2$  are the variances of two (asymptotically) unbiased estimators. Then  $V_1 \preceq V_2$  implies that the estimator associated with  $V_1$  is more efficient than the estimator associated with  $V_2$ . This means, for example, that the confidence ellipsoid associated with  $V_1$  lies within the confidence ellipsoid associated with  $V_2$ .

**Theorem 3.3** *Whenever the data  $\mathcal{D}$  identifies  $\omega$  we have*

$$B \preceq A. \tag{11}$$

Moreover, for any function  $\Phi$  we have  $V_B^\Phi \preceq V_A^\Phi$  where  $V_A^\Phi$  and  $V_B^\Phi$  are the asymptotic variances of  $\Phi(\bar{\theta}_A, \bar{\eta}_A)$  and  $\Phi(\bar{\theta}_B, \bar{\eta}_B)$  respectively.

Theorem 3.3 indicates that, if possible, it is always asymptotically beneficial to estimate both  $\theta$  and  $\eta$  using the data  $\mathcal{D}$  collected in the study  $\mathcal{S}$ . Moreover, Theorem 3.3 holds also when only a sub-vector of  $\eta$  is identified by the data  $\mathcal{D}$ .

**Remark 5** As noted by a referee, an alternative approach to Type II Problems would be to combine the first and second estimation steps. This can be done by simultaneously solving the estimating equations

$$\Psi(\theta, \eta) = \mathbf{0} \quad \text{and} \quad \Gamma(\theta, \eta) + m\hat{\Sigma}^{-1}(\hat{\eta} - \eta) = \mathbf{0}. \tag{12}$$

The system (12) is obtained from (3) by augmenting the estimating function  $\Gamma$  with the term  $m\hat{\Sigma}^{-1}(\hat{\eta} - \eta)$ . The latter is a pseudo-score equation which follows directly from (6). By appropriately modifying the proof of Theorem 3.2 it can be shown that  $(\bar{\theta}_B, \bar{\eta}_B)$  have the same limiting distribution as the solution of (12). It thus follows that the estimator (7) is asymptotically efficient up to the first order.

Another variant of Type II Problems occurs when the data  $\mathcal{D}$  is not available, but nevertheless the estimates  $(\tilde{\theta}, \tilde{\eta})$  from the current study as well as their estimated variance, i.e.,  $\tilde{\Upsilon}$  is given. The objective is then to combine the current estimators (4) with the historical estimators (6). To this end we propose estimating  $\theta$  by

$$\bar{\theta}_C = \tilde{\theta} - \tilde{\Upsilon}_{\theta\eta} \tilde{\Upsilon}_{\eta\eta}^{-1}(\tilde{\eta} - \bar{\eta}_C) \tag{13}$$

where  $\bar{\eta}_C = \bar{\eta}$  is given by (7). The estimators (7) as well as (13) are motivated by Lemma 2 and Remark 8 appearing in the Appendix.

**Theorem 3.4** *Let  $(\bar{\theta}_C, \bar{\eta}_C)^T$  be defined by (7) and (13). Suppose further that (4) and (6) hold and both  $\Upsilon$  and  $\Sigma$  can be consistently estimated. Then as  $n \rightarrow \infty$  we have*

$$\sqrt{n}(\bar{\theta}_C - \theta_0, \bar{\eta}_C - \eta_0)^T \Rightarrow \mathcal{N}_{p+q}(\mathbf{0}, C)$$

where  $C = MVM^T$  with

$$V = \begin{pmatrix} \mathbf{Y} & \mathbf{0} \\ \mathbf{0} & \rho \Sigma \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} \mathbf{I} & -\mathbf{R}\mathbf{W}_2 & \mathbf{R}\mathbf{W}_2 \\ \mathbf{0} & \mathbf{W}_1 & \mathbf{W}_2 \end{pmatrix}. \tag{14}$$

The matrices  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are defined in (8) and  $\mathbf{R} = \mathbf{Y}_{\theta\eta}\mathbf{Y}_{\eta\eta}^{-1}$ . Moreover, we have:

$$\begin{aligned} C_{\theta\theta} &= \mathbf{Y}_{\theta\theta} - \mathbf{Y}_{\theta\eta}\mathbf{Y}_{\eta\eta}^{-1}\mathbf{W}_2\mathbf{Y}_{\theta\eta}^T, \\ C_{\theta\eta} &= \mathbf{Y}_{\theta\eta}\mathbf{W}_1 \\ C_{\eta\eta} &= (\mathbf{Y}_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}. \end{aligned}$$

Theorem 3.4 describes the large sample behavior of the estimators (7) and (13). Further insight is facilitated by considering the simplest possible situation, i.e., when  $(\theta, \eta) \in \mathbb{R}^2$ , in which case  $\sqrt{m}(\hat{\eta} - \eta_0) \Rightarrow \mathcal{N}(0, \sigma^2)$  for the historical data, whereas for the current study  $\sqrt{n}(\hat{\theta} - \theta_0, \hat{\eta} - \eta_0)^T \Rightarrow \mathcal{N}(0, \mathbf{Y})$  where

$$\mathbf{Y} = \begin{pmatrix} v_{\theta\theta}^2 & v_{\theta\eta} \\ v_{\theta\eta} & v_{\eta\eta}^2 \end{pmatrix}.$$

It is not hard to see that (13) reduces to  $\bar{\theta} = \tilde{\theta} - (\tilde{v}_{\theta\eta}/\tilde{v}_{\eta\eta}^2)(\tilde{\eta} - \bar{\eta})$  where  $\bar{\eta} = w_1^*\tilde{\eta} + w_2^*\hat{\eta}$  with

$$w_1^* = \frac{n/\tilde{v}_{\eta\eta}^2}{n/\tilde{v}_{\eta\eta}^2 + m/\hat{\sigma}^2} \quad \text{and} \quad w_2^* = \frac{m/\hat{\sigma}^2}{n/\tilde{v}_{\eta\eta}^2 + m/\hat{\sigma}^2}.$$

Furthermore  $C_{\theta\theta}$  simplifies to

$$v_{\theta\theta}^2 - \frac{v_{\theta\eta}^2}{v_{\eta\eta}^2}w_2 = v_{\theta\theta}^2(1 - w_2r^2), \tag{15}$$

where  $r = v_{\theta\eta}/(v_{\theta\theta}v_{\eta\eta})$  is the asymptotic correlation between  $\tilde{\theta}$  and  $\tilde{\eta}$  and

$$w_2 = \frac{(1 - \gamma)/\sigma^2}{\gamma/v_{\eta\eta}^2 + (1 - \gamma)/\sigma^2}$$

is the limiting value of  $w_2^*$  as  $n/(n + m) \rightarrow \gamma$ . It follows that the asymptotic relative efficiency of  $\bar{\theta}$  to  $\tilde{\theta}$  is  $1 - w_2r^2$ , which is at most unity (when  $v_{\theta\eta} = 0$ ) and no less than  $1 - r^2$  (when  $\gamma$  is close to 0). Clearly, the historical estimates are useful only if the covariance  $v_{\theta\eta}$  is non-zero and highly useful whenever  $w_2$  is close to unity. A similar but more involved analysis applies when the parameters are multidimensional.

We emphasize that the structure of the estimators  $\bar{\eta}_C$  and  $\bar{\theta}_C$  as well as the form of  $C$  are related to, but much more general, than results obtained in the literature on both double sampling and monotone missing normal data (Anderson 1957; Morrison 1971; Kanda and Fujikoshi 1998). Double sampling is a widely used technique in survey

sampling, where the estimator is also known as the generalized regression estimator (Thompson 1997), as well as in other applications, cf. Davidov and Haitovsky (2000), Chen and Chen (2000) and the references therein. We also note that Eq. (15) is a generalization of the formulas obtained for the usual double sampling estimator (e.g., Tamhane 1978) where  $w_2 = m/(n + m)$ . The following Theorem substantially generalizes on results obtained in the literature on both the double sampling and monotone missing data.

**Theorem 3.5** *We have*

$$C \preceq \Upsilon \tag{16}$$

Moreover, for any function  $\Phi$  we have  $V_C^\Phi \preceq V_Y^\Phi$  where  $V_C^\Phi$  and  $V_Y^\Phi$  are the asymptotic variances of  $\Phi(\bar{\theta}_C, \bar{\eta}_C)$  and  $\Phi(\tilde{\theta}, \tilde{\eta})$  respectively.

In words, the estimator  $(\bar{\theta}_C, \bar{\eta}_C)$ , incorporating the historical estimates and derived by combining  $(\tilde{\theta}, \tilde{\eta})$  and  $\hat{\eta}$ , is more precise than  $(\tilde{\theta}, \tilde{\eta})$ , the estimator based only on the current study.

**Remark 6** It is also important to emphasize that in finite, typically small samples, the estimator  $C_{\theta\theta}$  may be in fact inferior to  $\Upsilon_{\theta\theta}$ . This typically occurs when the “regression matrix”  $R$ , see the statement of Theorem 3.4, is poorly estimated. This feature has been also recognized in the double sampling literature (Tamhane 1978).

A little algebra shows that

$$\begin{aligned} C_{\theta\theta} &= \Upsilon_{\theta\theta} - \Upsilon_{\theta\eta} \Upsilon_{\eta\eta}^{-1} (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1} (\rho \Sigma)^{-1} \Upsilon_{\theta\eta}^T, \\ C_{\theta\eta} &= \Upsilon_{\theta\eta} \Upsilon_{\eta\eta}^{-1} (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1} \\ C_{\eta\eta} &= (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1}. \end{aligned}$$

so we can remove the dependence of  $C$  on the matrices  $W_1$  and  $W_2$ .

Clearly, whenever the data  $\mathcal{D}$  is available both  $(\bar{\theta}_B, \bar{\eta}_B)$  and  $(\bar{\theta}_C, \bar{\eta}_C)$  can be calculated where  $\bar{\eta}_B = \bar{\eta}_C$  are given in (7). Recall that  $\bar{\theta}_B$  solves  $\Psi(\theta, \bar{\eta}) = \mathbf{0}$  where  $\Psi(\theta, \eta) = n^{-1} \sum_{i=1}^n \psi(\theta, \eta, Y_i)$ . Similarly, we can view  $\bar{\theta}_C$  as a solution to an estimating equation  $\Lambda(\theta, \bar{\eta}) = \mathbf{0}$  where  $\Lambda(\theta, \eta) = n^{-1} \sum_{i=1}^n \lambda(\theta, \eta, Y_i)$ . The form of  $\Lambda$  can be easily deduced from Lemma 2 and that of  $\lambda$  by plugging in the influence functions for  $\tilde{\theta}$  and  $\tilde{\eta}$  into  $\Lambda$ . In fact, the precise form of the influence function of  $\bar{\theta}_C$  is readily derived, for more details see Remark 9 appearing in the Appendix. It is worth noting that  $\Psi$  operates on the full data  $\mathcal{D}$  whereas  $\Lambda$  operates on functions thereof namely the estimators  $(\tilde{\theta}, \tilde{\eta})$  and  $\hat{\eta}$ . Thus  $(\bar{\theta}_C, \bar{\eta}_C)$  can be viewed as functions of a coarsening of the data  $\mathcal{D}$  and therefore is expected to be less efficient than  $(\bar{\theta}_B, \bar{\eta}_B)$ . This indeed is the case under mild regularity conditions. A formal statement requires the introduction of some additional notation. Let  $\mathbf{h} = \mathbf{h}(\theta, \eta, Y)$  denote any estimating function and denote  $D_{\theta_0}(\mathbf{h}) = \mathbb{E}_0(\partial \mathbf{h} / \partial \theta)$  and  $D_{\eta_0}(\mathbf{h}) = \mathbb{E}_0(\partial \mathbf{h} / \partial \eta)$ . Note that earlier we referred to  $D_{\theta_0}(\psi)$  and  $D_{\eta_0}(\psi)$  simply as  $D_{\theta_0}$  and  $D_{\eta_0}$ . Now:

**Theorem 3.6** *Suppose that both  $\bar{\omega}_B$  and  $\bar{\omega}_C$  can be obtained. If*

$$D_{\theta_0}(\psi)^{-1}D_{\eta_0}(\psi) \leq D_{\theta_0}(\lambda)^{-1}D_{\eta_0}(\lambda) \tag{17}$$

*component-wise and*

$$(D_{\theta_0}(\psi)^{-1})\mathbb{E}_0(\psi\psi^T)(D_{\theta_0}(\psi)^{-1})^T \preceq (D_{\theta_0}(\lambda)^{-1})\mathbb{E}_0(\lambda\lambda^T)(D_{\theta_0}(\lambda)^{-1})^T \tag{18}$$

*in the Loewner order, then*

$$B \preceq C. \tag{19}$$

*Moreover, for any function  $\Phi$  we have  $V_B^\Phi \preceq V_C^\Phi$  where  $V_B^\Phi$  and  $V_C^\Phi$  are the asymptotic variances of  $\Phi(\bar{\theta}_B, \bar{\eta}_B)$  and  $\Phi(\bar{\theta}_C, \bar{\eta}_C)$  respectively.*

Condition (18) holds when the estimating equation  $\Psi(\theta, \eta_0) = \mathbf{0}$  results in more efficient estimators for  $\theta$  than those resulting from  $\Lambda(\theta, \eta_0) = \mathbf{0}$  when  $\eta = \eta_0$  is set to its true value. This condition holds for any sensible choice of  $\Psi$ . In particular it holds for the score equations associated with maximum likelihood estimation. Condition (17) roughly means that  $\Psi$  is less sensitive to small perturbations in both  $\theta$  and  $\eta$  compared with  $\Lambda$ . Conditions (17) are (18) are not necessary. For example, the conclusion of Theorem 3.6 may hold if  $\psi$  is more sensitive to small perturbations but at the same time much more efficient. We believe that the aforementioned conditions hold broadly and the estimators  $(\bar{\theta}_C, \bar{\eta}_C)$ , described in Theorem 3.4, are generally less efficient than  $(\bar{\theta}_B, \bar{\eta}_B)$ , described in Theorem 3.2. For an additional discussion see Remark 9 in the Appendix. There are, however, situations in which  $B = C$  and situations where  $\bar{\omega}_B = \bar{\omega}_C$  for any data  $\mathcal{D}$ . As we shall see in the next section this is the case in normal linear models in which the estimators  $(\hat{\theta}, \hat{\eta})$  and  $\hat{\eta}$  are actually sufficient statistics. Finally, it is worth noting that if  $\Upsilon_{\theta\eta} = \mathbf{0}$  then the estimator  $(\bar{\theta}_C, \bar{\eta}_C)$  does not improve  $\bar{\theta}$  whereas there is always an improvement when the full data  $\mathcal{D}$  is available.

**Remark 7** To summarize, the proposed estimators are designed to extract as much information as possible from the data. Recall that in Type I Problems the parameter  $\eta$  is not estimable given the current data  $\mathcal{D}$ . The pair  $(\bar{\theta}_A, \bar{\eta}_A)$  solves the system of equations  $\Psi(\bar{\theta}_A, \bar{\eta}_A) = \mathbf{0}$  with  $\bar{\eta}_A = \hat{\eta}$  where  $\hat{\eta}$  is the historical estimator. In Type II problems the pair  $(\bar{\theta}_B, \bar{\eta}_B)$  solves the system of equations  $\Psi(\bar{\theta}_B, \bar{\eta}_B) = \mathbf{0}$  where  $\bar{\eta}_B$ , given by (7), is a weighted combination of  $\hat{\eta}$  and  $\tilde{\eta}$ . Moreover, as noted earlier it can be shown that the resulting estimators are asymptotically efficient. Our approach to Type I and the first variant of Type II problems are of similar structure: plug into  $\Psi$  the best available estimator of  $\eta$ . In the second variant of Type II Problems the pair  $(\bar{\theta}_C, \bar{\eta}_C)$  solves a different system of estimating equations which we denote by  $\Lambda(\bar{\theta}_C, \bar{\eta}_C) = \mathbf{0}$ . These equations are the likelihood equations given in Lemma 2 assuming that the variance matrices  $\Sigma$  and  $\Upsilon$  are fully known. Since  $(\bar{\theta}_C, \bar{\eta}_C)$  are based on further coarsening of the data they are generally less efficient.

### 4 Illustrations, applications and numerical results

In this section two applications are discussed in detail. In Sect. 4.1 the two-way ANOVA problem introduced in Sect. 2 is investigated. In particular, various design options for the current study  $\mathcal{S}$  are evaluated. It is worth noting that although the abovementioned ANOVA problem is among the simplest possible, its analysis is far from trivial. Next, in Sect. 4.2 we discuss the use of historical estimates in the design of drug interaction studies in the context of Bliss independence. A simple algorithm for the design of such studies is proposed.

#### 4.1 Two way ANOVA

Recall the ANOVA model of Sect. 2 where the studies  $\mathcal{S}_1$  and  $\mathcal{S}_2$  were designed to estimate  $\boldsymbol{\eta}_1 = (\eta_0, \eta_1)^T$  and  $\boldsymbol{\eta}_2 = (\eta_0, \eta_2)^T$ , respectively. Note that the parameter  $\eta_0$  is estimated in both studies so  $\boldsymbol{\eta}_1$  and  $\boldsymbol{\eta}_2$  are not distinct. Therefore employing any of the aforementioned findings requires the aggregation of the historical estimates as if they came from a single experiment. The historical studies result in the estimates  $(\widehat{\eta}_0(\mathcal{S}_1), \widehat{\eta}_1(\mathcal{S}_1))$  and  $(\widehat{\eta}_0(\mathcal{S}_2), \widehat{\eta}_2(\mathcal{S}_2))$  as well as their standard errors. Given these we can easily back calculate the unobserved means and sample sizes in the studies  $\mathcal{S}_1$  and  $\mathcal{S}_2$  and estimate  $(\eta_0, \eta_1, \eta_2)$  by:

$$\widehat{\eta}_1 = \bar{Y}_1(\mathcal{S}_1) - \widehat{\eta}_0, \quad \widehat{\eta}_2 = \bar{Y}_2(\mathcal{S}_2) - \widehat{\eta}_0, \quad \text{and} \quad \widehat{\eta}_0 = \frac{m_{1,0}\bar{Y}_0(\mathcal{S}_1) + m_{2,0}\bar{Y}_0(\mathcal{S}_2)}{m_{1,0} + m_{2,0}}, \tag{20}$$

where the quantity  $\bar{Y}_j(\mathcal{S}_i)$  is the average response on treatment  $j \in \{0, 1, 2\}$  in study  $i \in \{1, 2\}$  and  $m_{i,j}$  is the size of of treatment group  $j$  in study  $i$ . Under the usual conditions

$$\sqrt{m}(\widehat{\eta}_0 - \eta_0, \widehat{\eta}_1 - \eta_1, \widehat{\eta}_2 - \eta_2)^T \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

for some matrix  $\boldsymbol{\Sigma}$ . Furthermore, if (1) is a homoscedastic model with variance  $\sigma^2$  and  $m_{1,0} = m_{1,1} = m_{2,0} = m_{2,2}$ , i.e., the studies  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are balanced and of the same size, then it is easy to see that

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 2 & -2 & -2 \\ -2 & 6 & 2 \\ -2 & 2 & 6 \end{pmatrix}.$$

We will now investigate various designs for a new study  $\mathcal{S}$ . When the primary focus of  $\mathcal{S}$  is inference on  $\theta$  then it may be advantageous in some circumstances to allocate all  $n$  observations to the treatment arm receiving both treatments one and two, i.e.,  $T_1 = T_2 = 1$  for all observations. This is clearly a Type I problem since  $\boldsymbol{\omega}$  is not identifiable from  $\mathcal{D}$  but given  $\boldsymbol{\eta}$  the parameter  $\theta$  is estimable. Note that an unbiased estimator for  $\theta$  is

$$\bar{\theta}_A = \bar{Y}_{12}(\mathcal{S}) - (\hat{\eta}_0 + \hat{\eta}_1 + \hat{\eta}_2) \tag{21}$$

and it is not hard to see that (21) solves (2) when  $\psi(\theta, \eta_0, \eta_1, \eta_2, Y_i) = Y_i - \eta_0 - \eta_1 - \eta_2 - \theta$ . Thus,  $\Sigma_\psi = \sigma^2$ ,  $D_{\theta_0} = 1$  and  $D_{\eta_0} = -(1, 1, 1)$  and it follows that  $A_{\theta\theta}$ , the asymptotic variance of (21) as described in Theorem 3.1, reduces to

$$\sigma^2 \times (1 + 10\rho) \quad \text{where} \quad \rho = \lim \frac{n}{m}.$$

The second term appearing in the parentheses in the above display is an inflation factor, i.e., the price to pay for substituting estimates for the unknown value of  $(\eta_0, \eta_1, \eta_2)$ . Note that when  $n/m \rightarrow 0$  as both  $m \rightarrow \infty$  and  $n \rightarrow \infty$  the asymptotic variance of  $\bar{\theta}_A$  approaches  $\sigma^2$ . In practice this requires a large current study and even larger historical data. Incidentally, since  $\bar{\theta}_A$  is a linear function of  $\bar{Y}_{12}(\mathcal{S})$  and  $(\hat{\eta}_0, \hat{\eta}_1, \hat{\eta}_2)$  it is not hard to see that its exact variance is  $\sigma^2(1/n + 10/m)$  which coincides with the asymptotic form.

Alternatively, suppose that the study  $\mathcal{S}$  allocates  $n/4$  observations to all treatment combinations. In this case the data  $\mathcal{D}$  identifies  $\omega = (\theta, \eta_0, \eta_1, \eta_2)^T$ , so this is a Type II problem. The usual estimators for this design are  $\tilde{\eta}_0 = \bar{Y}_0(\mathcal{S})$ ,  $\tilde{\eta}_1 = \bar{Y}_1(\mathcal{S}) - \bar{Y}_0(\mathcal{S})$ ,  $\tilde{\eta}_2 = \bar{Y}_2(\mathcal{S}) - \bar{Y}_0(\mathcal{S})$  and

$$\tilde{\theta} = \bar{Y}_{12}(\mathcal{S}) - (\bar{Y}_1(\mathcal{S}) + \bar{Y}_2(\mathcal{S})) + \bar{Y}_0(\mathcal{S})$$

and thus the limiting variance of  $(\tilde{\theta}, \tilde{\eta})^T$  is

$$\Upsilon = \sigma^2 \begin{pmatrix} 16 & 4 & -8 & -8 \\ 4 & 4 & -4 & -4 \\ -8 & -4 & 8 & 4 \\ -8 & -4 & 4 & 8 \end{pmatrix}.$$

Next we aggregate the historical and current estimates for  $\eta$ . As in Sect. 3 we estimate  $\eta$  by  $\bar{\eta} = W_1\hat{\eta} + W_2\tilde{\eta}$  where

$$W_1 = (n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}n\Upsilon_{\eta\eta}^{-1} \quad \text{and} \quad W_2 = (n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}m\Sigma^{-1}.$$

Note that the weight matrices are functions of the variances  $\Upsilon_{\eta\eta}$  and  $\Sigma$  as well as the ratio  $n/(n + m)$ . Since  $\mathcal{D}$  is fully available to us then we can estimate  $\theta$  by

$$\bar{\theta}_B = \bar{Y}_{12}(\mathcal{S}) - (\bar{\eta}_0 + \bar{\eta}_1 + \bar{\eta}_2). \tag{22}$$

Note that the estimators (21) and (22) are of the same functional form. Further note that the statistic  $\bar{Y}_{12}(\mathcal{S})$  in (22) is a function of the  $n_{12}$  observations  $Y_1, \dots, Y_{n_{12}}$  receiving the treatment combination  $T_1 = T_2 = 1$ . A straightforward calculation shows that  $B_{\theta\theta}$  is given by

$$\sigma^2 \times (\xi_{11}^{-1} + \mathbf{1}^T(\Upsilon_{\eta\eta}^{-1} + (\rho\Sigma)^{-1})^{-1}\mathbf{1}) \quad \text{where} \quad \rho = \lim \frac{n}{m},$$

**Table 1** Asymptotic variances for  $\theta$  for Type I and Type II Problems (with a balanced design) as function of the sizes of the historical and current studies

| $m$  | 100                |                    | 200                |                    | 500                |                    | 1000               |                    | 2000               |                    | 5000               |                    |
|------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| $n$  | $A_{\theta\theta}$ | $B_{\theta\theta}$ | $A_{\theta\theta}$ | $B_{\theta\theta}$ | $A_{\theta\theta}$ | $B_{\theta\theta}$ | $A_{\theta\theta}$ | $B_{\theta\theta}$ | $A_{\theta\theta}$ | $B_{\theta\theta}$ | $A_{\theta\theta}$ | $B_{\theta\theta}$ |
| 100  | 11.0               | 9.30               | 6.0                | 7.50               | 3.0                | 5.70               | 2.0                | 4.90               | 1.5                | 4.50               | 1.2                | 4.20               |
| 200  | 21.0               | 11.33              | 11.0               | 9.33               | 5.0                | 6.95               | 3.0                | 5.70               | 2.0                | 4.92               | 1.4                | 4.39               |
| 500  | 51.0               | 13.52              | 26.0               | 11.94              | 11.0               | 9.33               | 6.0                | 7.47               | 3.5                | 6.04               | 2.0                | 4.39               |
| 1000 | 101.0              | 14.61              | 51.0               | 13.52              | 21.0               | 11.3               | 11.0               | 9.33               | 6.0                | 7.47               | 3.0                | 5.70               |
| 2000 | 201.0              | 15.26              | 101.0              | 14.61              | 41.0               | 13.07              | 21.0               | 11.33              | 11.0               | 9.33               | 5.0                | 6.95               |
| 5000 | 501.0              | 15.69              | 251.0              | 15.40              | 101.0              | 14.61              | 51.0               | 13.52              | 26.0               | 11.94              | 11.0               | 9.33               |

where  $\xi_{11}$  is the fraction of the observations which are assigned to receive both treatments. In situations where the full data is not available to us but  $(\tilde{\theta}, \tilde{\eta}_0, \tilde{\eta}_1, \tilde{\eta}_2)$  are known we may estimate  $\theta$  by  $\bar{\theta}_C = \tilde{\theta} - \mathbf{Y}_{\theta\eta} \mathbf{Y}_{\eta\eta}^{-1} (\tilde{\eta} - \bar{\eta})$ . It can be verified that in this application, in which a normal linear model is involved and all estimators are functions of sufficient statistics, the estimators  $\bar{\theta}_B$  and  $\bar{\theta}_C$  coincide. Therefore  $\bar{\theta}_C$  is not discussed any further.

Table 1 provides a comparison of the asymptotic variances of (21) and (22) for a range of values of  $m$  and  $n$ . Table 1 displays asymptotic variances; the variances themselves are found by dividing any entry in the table by the size of the current study in the relevant row. Observe that both  $A_{\theta\theta}$  and  $B_{\theta\theta}$  decrease as a function of  $m$  for any fixed value of  $n$  and increase in  $n$  for any fixed  $m$ . For example when  $n = m = 100$  then  $A_{\theta\theta} = 11$  and  $B_{\theta\theta} = 9.3$  whereas when  $m = 100$  and  $n = 5000$  then  $A_{\theta\theta} = 501$  and  $B_{\theta\theta} = 15.69$  and when  $m = 5000$  and  $n = 100$  then  $A_{\theta\theta} = 1.2$  and  $B_{\theta\theta} = 4.2$ . Thus going down the first column of Table 1 the asymptotic variance  $A_{\theta\theta}$  increases by a factor of approximately 45 whereas that of  $B_{\theta\theta}$  by the much more modest 1.4. Similarly going across the first row the asymptotic variances of  $A_{\theta\theta}$  and  $B_{\theta\theta}$  are reduced by a factor of 9.2 and 2.2 respectively. Each pair  $(n, m)$  provides a direct comparison between the two proposed designs (design A, say, in which all experimental units in the current study receive both treatments and design B, say, which is a balanced design). Clearly, design A seems preferable in situations where  $m$  is much larger than  $n$ , otherwise design B is to be preferred.

We now look a bit deeper into the question of optimal design. Suppose for simplicity that the historical sample satisfies  $m_{1,0} = m_{1,1} = m_{2,0} = m_{2,2}$ . Note that Table 1 considers only designs with  $\xi = (0, 0, 0, 1)$  and  $\xi = (1/4, 1/4, 1/4, 1/4)$ . Therefore we next consider designs for the study  $\mathcal{S}$  where  $\xi = (\xi_{00}, \xi_{10}, \xi_{01}, \xi_{11})$  is any value in the unit simplex. Clearly here  $\xi_{ij}$  denotes the proportion of observations who received treatment combination  $i \times j$  where  $i$  and  $j$  are in  $\{0, 1\}$ . It is not hard to see that the optimal design in the interior of the simplex, i.e. for a Type II Problem, is attained by Theorem 3.2 when  $\xi_{11}^{-1} + \mathbf{1}^T (\mathbf{Y}_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1} \mathbf{1}$  is minimized where

**Table 2** Optimal design for Type II problems as a function of the sampling ratio  $\rho$

| Sampling ratio ( $\rho$ ) | Minimal variance $B_{\theta\theta}$ | Optimal design ( $\xi$ )     |
|---------------------------|-------------------------------------|------------------------------|
| 1/8                       | 2.27                                | (0.020, 0.001, 0.001, 0.978) |
| 1/4                       | 3.51                                | (0.020, 0.001, 0.001, 0.978) |
| 1/2                       | 5.53                                | (0.020, 0.160, 0.160, 0.660) |
| 1                         | 8.00                                | (0.020, 0.243, 0.243, 0.494) |
| 2                         | 10.66                               | (0.125, 0.250, 0.250, 0.375) |
| 4                         | 12.80                               | (0.187, 0.250, 0.250, 0.313) |
| 8                         | 14.22                               | (0.219, 0.250, 0.250, 0.281) |

$$\mathbf{Y} = \sigma^2 \begin{pmatrix} \frac{1}{\xi_{00}} + \frac{1}{\xi_{10}} + \frac{1}{\xi_{01}} + \frac{1}{\xi_{11}} & \frac{1}{\xi_{00}} & -\left(\frac{1}{\xi_{00}} + \frac{1}{\xi_{10}}\right) & -\left(\frac{1}{\xi_{00}} + \frac{1}{\xi_{01}}\right) \\ \frac{1}{\xi_{00}} & \frac{1}{\xi_{00}} & -\frac{1}{\xi_{00}} & -\frac{1}{\xi_{00}} \\ -\left(\frac{1}{\xi_{00}} + \frac{1}{\xi_{10}}\right) & -\frac{1}{\xi_{00}} & \left(\frac{1}{\xi_{00}} + \frac{1}{\xi_{10}}\right) & \frac{1}{\xi_{00}} \\ -\left(\frac{1}{\xi_{00}} + \frac{1}{\xi_{01}}\right) & -\frac{1}{\xi_{00}} & \frac{1}{\xi_{00}} & \frac{1}{\xi_{00}} + \frac{1}{\xi_{10}} \end{pmatrix}.$$

Symmetry considerations imply that under optimality  $\xi_{01} = \xi_{10}$  and since  $\xi_{00} = 1 - 2\xi_{10} - \xi_{11}$  the minimization involves only a two dimensional search. Table 2 provides the optimal design, i.e., the vector  $\xi$  for estimating  $\theta$  for various values of the ratio  $\rho = n/m$  found by a grid search with step size 0.001 and the restriction that  $\xi_{00} \geq 0.02$ . This restriction is necessary; otherwise the matrix  $\mathbf{Y}$  can not be inverted.

Table 2 provides the optimal design for estimating  $\theta$  as a function of the sampling ratio. Note that for large  $\rho$ , i.e., when  $n$  is larger than  $m$ , we find that  $\xi_{01} = \xi_{10} = 1/4$  and that the difference between  $\xi_{11}$  and  $\xi_{00}$  decreases in  $\rho$ . We believe that the balanced design is optimal when  $\rho \rightarrow \infty$ . It is also clear that for large  $\rho$  the designs appearing in Table 2 are generally superior to those in Table 1. For example when  $\rho = 1$  we find that the asymptotic variances in Table 1 are 11.0 and 9.3 whereas the corresponding optimal asymptotic variance given in Table 2 is 8.0. However, for values of  $\rho$  smaller than a 1/4, i.e., when  $n$  is relatively small to  $m$ , then the optimal design sets  $\xi_{00} = 0.02$ , and  $\xi_{01} = \xi_{10} = 0.001$  which are the smallest possible values allowed by our search algorithm. This suggest that further, at most minor, improvements are possible by setting  $\xi_{00} = 0$  and/or  $\xi_{01} = \xi_{10} = 0$ . Clearly when  $\xi_{01} = \xi_{10} = \xi_{00} = 0$  we have a Type I Problem.

Therefore we next consider the situation that  $\xi_{00} = 0$  and  $\xi_{01} = \xi_{10} > 0$ , in which case the current study comprises of three arms and thus three group means:  $\bar{Y}_1(\mathcal{S})$ ,  $\bar{Y}_2(\mathcal{S})$  and  $\bar{Y}_{12}(\mathcal{S})$ . We emphasize that this estimation problem is neither a Type I nor a Type II problem. Further observe that with these data alone we can not estimate  $\omega$ . Nevertheless, the pair  $(\bar{Y}_1(\mathcal{S}), \bar{Y}_2(\mathcal{S}))^T$  whose mean is  $(\eta_0 + \eta_1, \eta_0 + \eta_2)$  can be aggregated with  $\hat{\eta}$  the historical estimate of  $\eta$ . By an appropriate modification of Lemma 2 it can be shown that  $\eta$  can be estimated by

$$\eta^\dagger = (nA^T V^{-1} A + m \Sigma^{-1})^{-1} (nA^T V^{-1} S + m \Sigma^{-1} \hat{\eta}) \tag{23}$$



where  $S = (\bar{Y}_1(S), \bar{Y}_2(S))^T$ ,  $V = \sigma^2 \text{diag}(\xi_{01}^{-1}, \xi_{10}^{-1})$  is its asymptotic variance and

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

is the matrix which satisfies  $\mathbb{E}(S) = A\eta$ . Note that (23) is of the same form as (7) but with  $A^T V^{-1} A$  instead of  $\Upsilon_{\eta\eta}$ . Now, let  $\bar{\theta}_D$  denote the solution to  $\Psi(\theta, \eta^\dagger) = 0$  which is nothing but

$$\bar{\theta}_D = \bar{Y}_{12}(S) - (\eta_0^\dagger + \eta_1^\dagger + \eta_2^\dagger) \tag{24}$$

A straightforward calculation shows that the asymptotic variance of  $\bar{\theta}_D$  is given by

$$\sigma^2 \times (\xi_{11}^{-1} + \mathbf{1}^T (A^T V^{-1} A + (\rho \Sigma)^{-1})^{-1} \mathbf{1}) \quad \text{where } \rho = \lim \frac{n}{m}.$$

The formula above is useful in finding the optimal design for small values of  $\rho$  when  $\xi_{00} = 0$ . For example when  $\rho = 1/8$  then the design  $\xi = (0, 0.0005, 0.0005, 0.9990)$  results in a variance of 2.25 (actually 2.250751) which is slightly smaller than 2.27 the reported variance in the first row of Table 2. Finally we note that when  $\rho = 1/8$  then  $A_{\theta\theta}$  equals (precisely) 2.25 which means that in this application a design for Type I Problem would be the most effective.

As noted by a referee, in addition to the above mentioned three arm trial one could choose various two arm designs for  $S$ . For example, one can choose a design for which  $\xi_{00} > 0$ ,  $\xi_{11} > 0$  and  $\xi_{01} = \xi_{10} = 0$ . Or alternatively designs for which  $\xi_{01} > 0$ ,  $\xi_{11} > 0$  and  $\xi_{00} = \xi_{10} = 0$ . It can be shown, however, that for small values of  $\rho$ , where such designs are of interest, these two armed designs are not superior to a Type I design.

### 4.2 Using historical estimates in drug interaction studies

This subsection deals with the optimal design of drug interaction studies. Consider two drugs  $D_1$  and  $D_2$  with no-effect probabilities  $\eta_1$  and  $\eta_2$ , respectively and let  $\theta$  denote the no-effect probability when both drugs are administered together. The drugs are called Bliss independent, see, Bliss (1939), Liu et al. (2018), if

$$\theta = \eta_1 \eta_2. \tag{25}$$

If (25) does not hold and  $\theta < \eta_1 \eta_2$  there is synergy among the drugs, otherwise there is antagonism. The concept of Bliss independence has seen a recent resurgence of interest as the need to assess the benefit of combination therapies and drug–drug interactions has increased. Some current references are Pallmann and Schaarschmidt (2016), Palmer and Sorger (2017), Russ and Kishony (2018) and Niu et al. (2019). Drug interaction studies are often carried out as single-dose experiments, e.g., Ansari et al. (2008), where the interaction is assessed by considering a single dose of each of the two drugs. A more elaborate design, which we will not consider here, assesses multiple drugs and doses using response surface methodology as in Lee (2010).

Naturally, the quantity of interest in drug interaction studies is

$$\Phi(\theta, \eta) = \log(\theta) - \log(\eta_1) - \log(\eta_2). \quad (26)$$

The formulation in (26) links the problem discussed here to the ANOVA setup considered earlier. In many applications of single dose interaction tests, whether using historical data or not, an explicit or implicit asymptotic argument is used, and the theoretical results for the asymptotic case presented above are relevant. For example, Demidenko and Miller (2019) describes a *Daphnia* acute test with two stressors, single doses of CuSO<sub>4</sub> and of NiCl, where the numbers of surviving organisms in water were counted after 48 hours. The observations reported were the surviving fractions of organisms only, without reporting their original numbers thus, essentially, assuming their original numbers were very high, i.e., applying an asymptotic argument. But as pointed out by Pallmann and Schaarschmidt (2016), in single-dose experiments, correct statistical analysis should rely on the observed frequencies, and not on the observed rates of success or failure. Therefore the sample sizes used in each arm of the experiment are of crucial importance and in this subsection we provide finite sample results.

For simplicity suppose that there exist historical estimates of  $\eta_1$  and  $\eta_2$  based on independent binomial experiments with sizes  $m_1$  and  $m_2$ . Suppose further that the current study allows for the recruitment of  $n$  experimental units,  $n_1$  of which will receive  $D_1$ ,  $n_2$  will receive  $D_2$  and  $n_{12}$  will receive both drugs. Obviously

$$n = n_1 + n_2 + n_{12} \quad (27)$$

and  $\theta$  can not be estimated unless  $n_{12} > 0$ . However it is possible that  $n_1 = n_2 = 0$ . The goal is to allocate the experimental units optimally, which is equivalent to the problem of optimally allocating  $n + m_1 + m_2$  observations in an experiment in which the single dose arms are no smaller than  $m_1$  and  $m_2$ , respectively. The optimal design problem can be approximated as the minimization of the large sample variance of (26)

$$\frac{1}{n_{12}} \frac{1 - \theta}{\theta} + \frac{1}{n_1 + m_1} \frac{1 - \eta_1}{\eta_1} + \frac{1}{n_2 + m_2} \frac{1 - \eta_2}{\eta_2}, \quad (28)$$

subject to the constraint (27).

In contrast with the design problem encountered in Sect. 4.1 the design criterion depends on the unknown parameters, i.e., the probabilities  $\theta$  and  $\eta$ . We propose allocating observations as if  $\eta_1 = \hat{\eta}_1$ ,  $\eta_2 = \hat{\eta}_2$  and  $\theta = \hat{\eta}_1 \hat{\eta}_2$  is equal to its estimated value under the hypothesis of Bliss independence.

One simple approach to the minimization of (28) is the following greedy iterative procedure, which sequentially allocates observations into the condition where the variance is reduced most:

**ALGORITHM**

```

 $n_{12} \leftarrow 1, n_1 \leftarrow 0, n_2 \leftarrow 0$ 
if  $n = n_1 + n_2 + n_{12}$  then
  stop
else
   $R_{12}(n_{12}) \leftarrow \frac{1}{n_{12}} \frac{1 - \hat{\eta}_1 \hat{\eta}_2}{\hat{\eta}_1 \hat{\eta}_2}$ 
   $R_1(n_1) \leftarrow \frac{1}{n_1 + m_1} \frac{1 - \hat{\eta}_1}{\hat{\eta}_1}$ 
   $R_2(n_2) \leftarrow \frac{1}{n_2 + m_2} \frac{1 - \hat{\eta}_2}{\hat{\eta}_2}$ 
   $C \leftarrow \min\{R_{12}(n_{12} + 1) - R_{12}(n_{12}), R_1(n_1 + 1) - R_1(n_1), R_2(n_2 + 1) - R_2(n_2)\}$ 

  if  $C = R_{12}(n_{12} + 1) - R_{12}(n_{12})$  then
     $n_{12} \leftarrow n_{12} + 1$ 
  end if
  if  $C = R_1(n_1 + 1) - R_1(n_1)$  then
     $n_1 \leftarrow n_1 + 1$ 
  end if
  if  $C = R_2(n_2 + 1) - R_2(n_2)$  then
     $n_2 \leftarrow n_2 + 1$ 
  end if
end if

```

For example, if  $m_1 = 30, m_2 = 50, \hat{\eta}_1 = 0.7$  and  $\hat{\eta}_2 = 0.8$ , and one had 56 observations, then 55 observations would be put in the arm where both treatments are administered, and only one would be used to improve the estimate of  $\hat{\eta}_1$ . Table 3 contains a tabulation of the optimal allocation of  $(n_{12}, n_1, n_2)$ . For selected combinations of the values of  $m_1, m_2, \eta_1, \eta_2$  the table gives the minimal value of  $n$ , denoted as  $n_{\min}$ , for which replications of the historic observations are needed, and then the optimal allocation for  $n_{\min}$ . As one would expect, when  $\theta$  is closer to 0.5 than  $\eta_1$  or  $\eta_2$ , a larger sample size  $n_{12}$  is allocated in the optimal design to estimating  $\theta$ , than  $m_1$  or  $m_2$ . In the opposite case,  $n_{12}$  is smaller than  $m_1$  or  $m_2$ .

As pointed out by a reviewer, the optimal design may not be unique. For example, in the first entry of the table, when  $\eta_1 = \eta_2 = 0.3, m + 1 - m_2 = 10$ , and one has 23 observations to allocate, 22 observations should be used for estimating joint effect  $\theta$  and one observation should be used to improve the estimate of either of the individual effects  $\eta_1$  or  $\eta_2$ .

**5 Summary and discussion**

This paper focuses on the situation when historically available information in the form of parameter estimates are either incorporated in the analysis of a current study or used to plan a future experiment. We did not explicitly discuss the large literature on data combination schemes or other two-stage plug-in methods. As mentioned, when

**Table 3** The minimal sample size  $n_{min}$ , at which optimal allocation requires to repeat historical observations, for selected values of the no-success probabilities  $\eta_1, \eta_2$  and historical sample sizes  $m_1, m_2$  in a test of drug interaction

| $m_1$                           | $m_2$ | $n_{min}$ | $n_{12}$ | $n_1$ | $n_2$ | $m_1$                           | $m_2$ | $n_{min}$ | $n_{12}$ | $n_1$ | $n_2$ |
|---------------------------------|-------|-----------|----------|-------|-------|---------------------------------|-------|-----------|----------|-------|-------|
| $(\eta_1, \eta_2) = (0.3, 0.3)$ |       |           |          |       |       | $(\eta_1, \eta_2) = (0.5, 0.7)$ |       |           |          |       |       |
| 10                              | 10    | 23        | 22       | 1     | 0     | 10                              | 10    | 15        | 14       | 1     | 0     |
| 20                              | 10    | 23        | 22       | 0     | 1     | 10                              | 20    | 15        | 14       | 1     | 0     |
| 30                              | 10    | 23        | 22       | 0     | 1     | 10                              | 30    | 15        | 14       | 1     | 0     |
|                                 |       |           |          |       |       | 20                              | 10    | 23        | 22       | 0     | 1     |
| $(\eta_1, \eta_2) = (0.3, 0.5)$ |       |           |          |       |       | 30                              | 10    | 23        | 22       | 0     | 1     |
| 10                              | 10    | 17        | 16       | 1     | 0     | $(\eta_1, \eta_2) = (0.5, 0.9)$ |       |           |          |       |       |
| 20                              | 10    | 26        | 25       | 0     | 1     |                                 |       |           |          |       |       |
| 30                              | 10    | 27        | 26       | 0     | 1     | 10                              | 10    | 13        | 12       | 1     | 0     |
| 10                              | 20    | 17        | 16       | 1     | 0     | 10                              | 20    | 13        | 12       | 1     | 0     |
| 10                              | 30    | 17        | 16       | 1     | 0     | 10                              | 30    | 13        | 12       | 1     | 0     |
|                                 |       |           |          |       |       | 20                              | 10    | 24        | 23       | 1     | 0     |
| $(\eta_1, \eta_2) = (0.3, 0.7)$ |       |           |          |       |       | 30                              | 10    | 35        | 34       | 1     | 0     |
| 10                              | 10    | 14        | 13       | 1     | 0     | $(\eta_1, \eta_2) = (0.7, 0.7)$ |       |           |          |       |       |
| 10                              | 20    | 14        | 13       | 1     | 0     |                                 |       |           |          |       |       |
| 10                              | 30    | 14        | 13       | 1     | 0     | 10                              | 10    | 17        | 16       | 1     | 0     |
| 20                              | 10    | 28        | 27       | 1     | 0     | 20                              | 10    | 17        | 16       | 0     | 1     |
| 30                              | 10    | 32        | 31       | 0     | 1     | 30                              | 10    | 17        | 16       | 0     | 1     |
| $(\eta_1, \eta_2) = (0.3, 0.9)$ |       |           |          |       |       | $(\eta_1, \eta_2) = (0.7, 0.9)$ |       |           |          |       |       |
| 10                              | 10    | 12        | 11       | 1     | 0     | 10                              | 10    | 18        | 17       | 1     | 0     |
| 10                              | 20    | 12        | 11       | 1     | 0     | 20                              | 10    | 25        | 24       | 1     | 0     |
| 10                              | 30    | 12        | 11       | 1     | 0     | 30                              | 10    | 25        | 24       | 0     | 1     |
| 20                              | 10    | 23        | 22       | 1     | 0     | 10                              | 20    | 13        | 12       | 1     | 0     |
| 30                              | 10    | 34        | 33       | 1     | 0     | 10                              | 30    | 13        | 12       | 1     | 0     |

historical estimates are incorporated in an analysis their variability is rarely accounted for in the analysis. A partial list of examples, drawn from the scientific literature was furnished earlier and many more exist. However, it seems very difficult to find published research where the details are given to the extent which would make the replication of the analysis possible. This limits one’s ability to apply the results of this paper to published research. However, the results presented here will inform future researchers of the scope and use of historical estimates and provide a toolkit for doing so. We hope that our investigation may have an effect on the quality of future analyses and publication standards.

We also agree with a referee who has noted that an estimate is always a coarsening of the full data, and it is clear that having only access to a historical estimate instead of the entire data leads to less efficient inferences.

Different disciplines exhibit different modes of using historical estimates. Social scientists often incorporate estimates from surveys in the process of model fitting, whereas biologists and engineers may use parameters estimated in experiments which are very different than their own. One way, of course, of incorporating historical esti-

mates is using prior distributions within the Bayesian framework. For recent examples see Hoff (2019) and Bryan and Hoff (2020). Our approach, however, is frequentist, as are most of the applications in the literature. In particular, we show how to incorporate historical estimates in a principled way in scenarios which we classify at Type I Problems, where the historical parameters are not re-estimated, and Type II Problems, where they are. Two variants of Type II Problems are described. See Theorems 3.1, 3.2 and 3.4. We also show that if, given the current data  $\mathcal{D}$ , it is possible to re-estimate the historical parameters then it is beneficial to do so at least for large sample sizes (Theorem 3.3). Other preference relations, in fact a hierarchy, among the estimators and any function thereof, were also established, cf. Theorems 3.5, 3.6. The loss of precision in the above mentioned settings is quantified in terms of a decomposition of the variance matrices. It was also demonstrated that the availability of historical estimates should be taken into account when an optimal experiment is designed. In particular, relevant methods for a two-way ANOVA and for testing drug interaction were discussed. Thus the results of this paper go beyond the existing knowledge on the use of historical estimates.

In our analysis we have assumed that the current data  $\mathcal{D}$  is a random sample and that the estimating Eq. (2) is of an additive form. These assumptions have been used merely to simplify the exposition and are easily modified to dependent data and various other estimating functions. It is clear that Type I and II Problems describe a broad range of possibilities, nevertheless they are insufficient for describing the rich collection of problems in which historical estimates may play a role. For example, our formulation assumes that the historical parameters  $\eta_1, \dots, \eta_K$  are distinct. However, in many situations this is not so. In fact, some of the historical studies may be full or partial replicates of each other. In cases when the current study is a partial replicate of a historical study, simple plug-in methods or re-estimation methods may be used. One has to be careful, though, about the choice of the estimates. We are aware of situations where a simple plug-in estimator performs better than a less than optimal re-estimating method. Throughout, we have assumed interchangeability. Clearly there are many experimental settings, especially in the sciences, where this assumption is realistic. In other situations, say clinical trials, heterogeneity rather than interchangeability is the rule. In such cases some modification of the methods proposed, using random effect models, may be possible. See Rukhin (2007) and the references therein.

Finally, it is also worth mentioning that the problem of accounting for historical estimates is naturally related, for obvious reasons, to sequential analysis, where data is collected over time, to meta-analysis, where the effort is to combine information from different sources and double sampling, and especially non-nested double sampling (Hidioglou 2001), which attempts to provide better inferences by augmenting and predicting unobserved quantities from existing data sets. The literature on combining surveys (Kim and Rao 2012) is also relevant. Further understanding can be possibly attained by incorporating ideas from these fields.

**Acknowledgements** Both authors thank COST Action IC1408 for Computationally Intensive Methods for the Robust Analysis of Non-standard Data that supported this research with grants for short visits. In addition the work of Ori Davidov was partially supported by the Israel Science Foundation Grants No. 456/17 and 2200/22 gratefully acknowledged. The authors are indebted to Anna Klimova for drawing their

attention to the importance of Bliss independence and to the anonymous reviewers for their insightful and detailed comments.

**Funding** Open access funding provided by Eötvös Loránd University.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix: Proofs

### Proof of Theorem 3.1:

*Proof* Since  $\bar{\theta}_A$  solves (2) we have

$$\Psi(\bar{\theta}_A, \hat{\eta}) = \mathbf{0}. \quad (29)$$

By assumption  $\psi$  is continuous and differentiable with respect to  $\theta$  and  $\eta_1, \dots, \eta_K$ . Thus, so is  $\Psi$ . Hence, by the mean value theorem

$$\Psi(\bar{\theta}_A, \hat{\eta}) = \Psi(\theta_0, \hat{\eta}) + \frac{\partial}{\partial \theta} \Psi(\theta_0, \hat{\eta})(\bar{\theta}_A - \theta_0) + o(\|\bar{\theta}_A - \theta_0\|).$$

Applying the mean value theorem to  $\Psi(\theta_0, \hat{\eta})$  in the display above yields

$$\Psi(\theta_0, \hat{\eta}) = \Psi(\theta_0, \eta_0) + \frac{\partial}{\partial \eta} \Psi(\theta_0, \eta_0)(\hat{\eta} - \eta_0) + o(\|\hat{\eta} - \eta_0\|),$$

so (29) can be rewritten as

$$\Psi(\theta_0, \eta_0) + \frac{\partial}{\partial \eta} \Psi(\theta_0, \eta_0)(\hat{\eta} - \eta_0) + \frac{\partial}{\partial \theta} \Psi(\theta_0, \hat{\eta})(\bar{\theta}_A - \theta_0) + \mathbf{R} = \mathbf{0}$$

where, assuming consistency  $\mathbf{R} = o(\|\bar{\theta}_A - \theta_0\|) + o(\|\hat{\eta} - \eta_0\|) = o_p(1)$ . Now, by the continuous mapping theorem and the law of large numbers we have:

$$\begin{aligned} \frac{\partial}{\partial \theta} \Psi(\theta_0, \hat{\eta}) &= \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \psi(\theta_0, \hat{\eta}, Y_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \psi(\theta_0, \eta_0, Y_i) + o_p(1) = \mathbb{E}_0(\partial \psi / \partial \theta) + o_p(1) \end{aligned}$$

which is a  $p \times p$  matrix. Similarly,

$$\frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{\Psi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{\psi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0, \mathbf{Y}_i) = \mathbb{E}_0(\partial \boldsymbol{\psi} / \partial \boldsymbol{\eta}) + o_p(1),$$

which is a  $p \times q$  matrix. For convenience we set  $\mathbf{D}_{\boldsymbol{\theta}_0} = \mathbb{E}_0(\partial \boldsymbol{\psi} / \partial \boldsymbol{\theta})$  and  $\mathbf{D}_{\boldsymbol{\eta}_0} = \mathbb{E}_0(\partial \boldsymbol{\psi} / \partial \boldsymbol{\eta})$ . Hence we can reexpress (29) more concisely as

$$\boldsymbol{\Psi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) + \mathbf{D}_{\boldsymbol{\eta}_0}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + \mathbf{D}_{\boldsymbol{\theta}_0}(\bar{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) + o_p(1) = 0$$

from which it follows, by the invertibility of  $\mathbf{D}_{\boldsymbol{\theta}_0}$ , that

$$\sqrt{n}(\bar{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) = -\mathbf{D}_{\boldsymbol{\theta}_0}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\psi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0, \mathbf{Y}_i) + \frac{\sqrt{n}}{\sqrt{m}} \mathbf{D}_{\boldsymbol{\eta}_0} \sqrt{m}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \right\} + o_p(1). \tag{30}$$

Since the first term in the curly brackets above is a function of the data  $\mathcal{D}$  collected in  $\mathcal{S}$  and the second term depends on the historical data, i.e., the studies  $\mathcal{S}_1, \dots, \mathcal{S}_K$  the two terms are independent. Now, by the central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \boldsymbol{\psi}(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0, \mathbf{Y}_i) \Rightarrow \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\psi}})$$

where  $\boldsymbol{\Sigma}_{\boldsymbol{\psi}} = \mathbb{E}_0(\boldsymbol{\psi} \boldsymbol{\psi}^T)$ . By assumption  $\sqrt{m_j}(\widehat{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_{j,0}) \Rightarrow \mathcal{N}_{q_j}(\mathbf{0}, \boldsymbol{\Sigma}_j)$  for each  $j$ . Thus  $\sqrt{m}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \Rightarrow \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\Sigma} = \text{BlockDiag}(\kappa_1 \boldsymbol{\Sigma}_1, \dots, \kappa_K \boldsymbol{\Sigma}_K)$  with  $\kappa_j = \lim(m/m_j)$  for  $j = 1, \dots, K$ . Thus,

$$\frac{\sqrt{n}}{\sqrt{m}} \mathbf{D}_{\boldsymbol{\eta}_0} \sqrt{m}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \Rightarrow \mathcal{N}_p(\mathbf{0}, \rho \mathbf{D}_{\boldsymbol{\eta}_0} \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{\eta}_0}^T). \tag{31}$$

Collecting terms shows that  $\sqrt{n}(\bar{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) \Rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}\boldsymbol{\theta}})$  where  $\mathbf{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}$  is as stated. Now, recall that  $\bar{\boldsymbol{\eta}}_A = \widehat{\boldsymbol{\eta}}$ . Thus, marginally  $\sqrt{n}(\bar{\boldsymbol{\eta}}_A - \boldsymbol{\eta}_0) \Rightarrow \mathcal{N}_q(\mathbf{0}, \rho \boldsymbol{\Sigma})$ , so  $\mathbf{A}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \rho \boldsymbol{\Sigma}$ . Clearly the joint asymptotic distribution of  $\sqrt{n}(\bar{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0, \bar{\boldsymbol{\eta}}_A - \boldsymbol{\eta}_0)^T$ , is also multivariate normal. Thus,

$$\begin{aligned} \mathbf{A}_{\boldsymbol{\theta}\boldsymbol{\eta}} &= \lim_n \text{Cov}(\sqrt{n}(\bar{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0), \sqrt{n}(\bar{\boldsymbol{\eta}}_A - \boldsymbol{\eta}_0)) = \lim_n [-\mathbf{D}_{\boldsymbol{\theta}_0}^{-1} \mathbf{D}_{\boldsymbol{\eta}_0} \text{Cov}(\sqrt{n}(\bar{\boldsymbol{\eta}}_A - \boldsymbol{\eta}_0), \\ &\quad \sqrt{n}(\bar{\boldsymbol{\eta}}_A - \boldsymbol{\eta}_0))] = -\rho \mathbf{D}_{\boldsymbol{\theta}_0}^{-1} \mathbf{D}_{\boldsymbol{\eta}_0} \boldsymbol{\Sigma}, \end{aligned}$$

as required, completing the proof. □

**Proof of Theorem 3.2:**

*Proof* Since  $\bar{\theta}_B$  solves (9) where  $\bar{\eta}$  is given in (7) we have

$$\Psi(\bar{\theta}_B, \bar{\eta}) = \mathbf{0}. \tag{32}$$

Following the derivations in the proof of Theorem 3.1, but with  $\bar{\eta}$  instead of  $\hat{\eta}$ , we find that (32) can be rewritten as

$$\Psi(\theta_0, \eta_0) + D_{\eta_0}(\bar{\eta} - \eta_0) + D_{\theta_0}(\bar{\theta}_B - \theta_0) + o_p(1) = \mathbf{0}$$

from which it follows, by the invertibility of  $D_{\theta_0}$ , that

$$\sqrt{n}(\bar{\theta}_B - \theta_0) = -D_{\theta_0}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n \psi(\theta_0, \eta_0, Y_j) + D_{\eta_0} \sqrt{n}(\bar{\eta} - \eta_0) \right\} + o_p(1). \tag{33}$$

Now using (7) and (8) we find that

$$\sqrt{n}(\bar{\eta} - \eta_0) = \sqrt{n}(W_1(\tilde{\eta} - \eta_0) + W_2(\hat{\eta} - \eta_0)) + o_p(1)$$

which we may substitute into (33) yielding

$$\begin{aligned} \sqrt{n}(\bar{\theta}_B - \theta_0) = -D_{\theta_0}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n \psi(\theta_0, \eta_0, Y_j) + D_{\eta_0} W_1 \sqrt{n}(\tilde{\eta} - \eta_0) \right. \\ \left. + D_{\eta_0} W_2 \sqrt{\frac{n}{m}} \sqrt{m}(\hat{\eta} - \eta_0) \right\} + o_p(1). \end{aligned} \tag{34}$$

The three terms in the curly brackets in (34) satisfy:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\theta_0, \eta_0, Y_i) &\Rightarrow \mathcal{N}_p(\mathbf{0}, \Sigma_\psi), \\ D_{\eta_0} W_1 \sqrt{n}(\tilde{\eta} - \eta_0) &\Rightarrow \mathcal{N}_q(\mathbf{0}, D_{\eta_0} W_1 \Upsilon_{\eta\eta} W_1^T D_{\eta_0}^T), \\ D_{\eta_0} W_2 \sqrt{\frac{n}{m}} \sqrt{m}(\hat{\eta} - \eta_0) &\Rightarrow \mathcal{N}_q(\mathbf{0}, \rho D_{\eta_0} W_2 \Sigma W_2^T D_{\eta_0}^T). \end{aligned}$$

Now, since  $\tilde{\eta}$  is a solution to (3) the quantity  $\sqrt{n}(\tilde{\eta} - \eta_0)$  may be expressed as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\theta_0, \eta_0, Y_i) + o_p(1)$$



for some function  $\psi$  which is known as the influence function (cf. Van der Vaart 2000). For more details see Remark 9. It follows by the central limit theorem, that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\theta_0, \eta_0, Y_i) \quad \text{and} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(\theta_0, \eta_0, Y_i)$$

are asymptotically jointly multivariate normal, thus so are the first two terms in the curly brackets of (34). Moreover the third term, which depends on the historical data is independent of the first two terms and normally distributed. Now the covariance among the first two terms is

$$\begin{aligned} \text{Cov} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\theta_0, \eta_0, Y_i), \sqrt{n}(\tilde{\eta} - \eta_0) \right) &= \text{Cov} \left( \sum_{i=1}^n \psi(\theta_0, \eta_0, Y_i), (\tilde{\eta} - \eta_0) \right) \\ &= \text{Cov}(n\psi(\theta_0, \eta_0, Y_1), (\tilde{\eta} - \eta_0)) \\ &= \mathbb{E}_0(\psi(\theta_0, \eta_0, Y_1)n(\tilde{\eta} - \eta_0)) = \mathbb{E}_0(\mathbb{E}_0(\psi(\theta_0, \eta_0, Y_1)n(\tilde{\eta} - \eta_0)|Y_1)) \\ &= \mathbb{E}_0(\psi(\theta_0, \eta_0, Y_1)\mathbb{E}_0(n(\tilde{\eta} - \eta_0)|Y_1)). \end{aligned}$$

However, by assumption  $\tilde{\eta}$  is asymptotically unbiased and  $\sqrt{n}$ -consistent, i.e.,  $\mathbb{E}_0(\tilde{\eta}) = \eta_0 + b/n + o(1/n)$  so  $\mathbb{E}_0(n(\tilde{\eta} - \eta_0)|Y_1) = n\mathbb{E}_0((\tilde{\eta} - \eta_0)) + o(1) = O(1)$ . Plugging the latter into the above display shows that covariance above converges to 0 as  $n \rightarrow \infty$ . It now follows that all three terms appearing in (34) are asymptotically independent.

Set  $\tilde{\eta}_B = \tilde{\eta}$  and observe that using (34) we have  $\sqrt{n}(\tilde{\eta}_B - \eta_0) \Rightarrow \mathcal{N}_q(\mathbf{0}, \mathbf{B}_{\eta\eta})$  where

$$\mathbf{B}_{\eta\eta} = \mathbf{W}_1 \Upsilon_{\eta\eta} \mathbf{W}_1^T + \rho \mathbf{W}_2 \Sigma \mathbf{W}_2^T.$$

Since  $\gamma/(1 - \gamma) = \rho$  we may reexpress the weight matrices as  $\mathbf{W}_1 = (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1} \Upsilon_{\eta\eta}^{-1}$  and  $\mathbf{W}_2 = (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1} (\rho \Sigma)^{-1}$ . Now, using the fact that products of symmetric matrices commute and a bit of algebra it can be shown that

$$\mathbf{B}_{\eta\eta} = (\Upsilon_{\eta\eta}^{-1} + (\rho \Sigma)^{-1})^{-1}.$$

Collecting terms shows that  $\sqrt{n}(\tilde{\theta}_B - \theta_0) \Rightarrow \mathcal{N}_p(\mathbf{0}, \mathbf{B}_{\theta\theta})$  where  $\mathbf{B}_{\theta\theta}$  is as stated. The stochastic representation (30) shows that the joint asymptotic distribution of  $\sqrt{n}(\tilde{\theta} - \theta_0, \tilde{\eta} - \eta_0)$  is also multivariate normal with

$$\begin{aligned} \text{Cov}(\sqrt{n}(\tilde{\theta}_B - \theta_0), \sqrt{n}(\tilde{\eta}_B - \eta_0)) &= -\mathbf{D}_{\theta_0}^{-1} \mathbf{D}_{\eta_0} \text{Cov}(\sqrt{n}(\tilde{\eta}_B - \eta_0), \\ &\sqrt{n}(\tilde{\eta}_B - \eta_0)) \rightarrow -\mathbf{D}_{\theta_0}^{-1} \mathbf{D}_{\eta_0} \mathbf{B}_{\eta\eta} \end{aligned}$$

as required, completing the proof. □

**Proof of Theorem 3.3:**

The following preliminary Lemma will be used.

**Lemma 1** *Let  $X_1$  and  $X_2$  be random vectors with variances  $V_1 = \mathbb{V}(X_1)$  and  $V_2 = \mathbb{V}(X_2)$  with  $V_1 \preceq V_2$ . Then for any matrix  $A$  we have  $\mathbb{V}(AX_1) \preceq \mathbb{V}(AX_2)$ . As a consequence we also have  $V_1^{-1} \succeq V_2^{-1}$ .*

**Proof of Lemma 1:**

*Proof* Observe that

$$\begin{aligned} u^T \mathbb{V}(AX_1)u &= u^T A V_1 A^T u = (A^T u)^T V_1 (A^T u) = v^T V_1 v \preceq v^T V_2 v \\ &= (A^T u)^T V_2 (A^T u) \\ &= u^T A V_2 A^T u = u^T \mathbb{V}(AX_2)u. \end{aligned}$$

for any vector  $u$ . The inequality  $v^T V_1 v \preceq v^T V_2 v$  holds since  $V_2 - V_1$  is non-negative definite by assumption. Thus  $\mathbb{V}(AX_1) \preceq \mathbb{V}(AX_2)$  as claimed.

Now choose  $A = V_1^{-1/2} V_2^{-1/2}$  and note that

$$\begin{aligned} \mathbb{V}(AX_1) &= (V_1^{-1/2} V_2^{-1/2}) V_1 (V_1^{-1/2} V_2^{-1/2})^T = V_2^{-1} \\ \mathbb{V}(AX_2) &= (V_1^{-1/2} V_2^{-1/2}) V_2 (V_1^{-1/2} V_2^{-1/2})^T = V_1^{-1}. \end{aligned}$$

The equalities above hold since products of symmetric matrices commute. The inequality  $V_1^{-1} \succeq V_2^{-1}$  follows immediately. □

We now continue with the proof of Theorem 3.3:

*Proof* Observe that  $A$  is the variance matrix of the random vector

$$T_1 = \begin{pmatrix} -D_{\theta_0}^{-1} & -D_{\theta_0}^{-1} D_{\eta_0} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

where  $S_1 \sim \mathcal{N}_p(\mathbf{0}, \Sigma_\psi)$  and  $S_2 \sim \mathcal{N}_q(\mathbf{0}, \rho \Sigma)$  are independent. Similarly,  $B$  is the variance matrix of the random vector

$$T_2 = \begin{pmatrix} -D_{\theta_0}^{-1} & -D_{\theta_0}^{-1} D_{\eta_0} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} S_1 \\ S_3 \end{pmatrix}$$

where  $S_1 \sim \mathcal{N}_p(\mathbf{0}, \Sigma_\psi)$  and  $S_3 \sim \mathcal{N}_q(\mathbf{0}, (\Upsilon_{\eta\eta} + (\rho \Sigma)^{-1})^{-1})$  are independent. Now,

$$\mathbb{V} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} - \mathbb{V} \begin{pmatrix} S_1 \\ S_3 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \rho \Sigma - (\Upsilon_{\eta\eta} + (\rho \Sigma)^{-1})^{-1} \end{pmatrix}$$

It is easy to verify that  $\Upsilon_{\eta\eta} + (\rho\Sigma)^{-1} \succeq (\rho\Sigma)^{-1}$  so by the second part of Lemma 1 we have  $\rho\Sigma \succeq (\Upsilon_{\eta\eta} + (\rho\Sigma)^{-1})^{-1}$  and therefore

$$\mathbb{V} \begin{pmatrix} S_1 \\ S_3 \end{pmatrix} \preceq \mathbb{V} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}.$$

Applying Lemma 1 we find that

$$B = \mathbb{V}(T_2) \preceq \mathbb{V}(T_1) = A$$

as required.

Next, an application of the  $\delta$ -method and Theorems 3.1 and 3.2 shows that  $\sqrt{n}(\Phi(\hat{\theta}_A, \bar{\eta}_A) - \Phi(\theta_0, \eta_0)) \Rightarrow \mathcal{N}_r(\mathbf{0}, PAP^T)$  and  $\sqrt{n}(\Phi(\bar{\theta}_B, \bar{\eta}_B) - \Phi(\theta_0, \eta_0)) \Rightarrow \mathcal{N}_r(\mathbf{0}, PBP^T)$  where  $r$  is the dimension of  $\Phi$  and  $P = \mathbb{E}_0(\partial\Phi/\partial\omega)$ . Observe that  $PAP^T$  is the variance of the random vector  $PT_1$  whereas  $PBP^T$  is the variance  $PT_2$ . Since  $B \preceq A$  it follows from Lemma 1 that  $PBP^T \preceq PAP^T$  concluding the proof.  $\square$

The following lemma motivates the use of the estimators (7) and (13)

**Lemma 2** Let  $W \sim \mathcal{N}_q(\eta, m^{-1}\Sigma)$  and  $(U, V)^T \sim \mathcal{N}_{p+q}((\theta, \eta)^T, n^{-1}\Upsilon)$  be independent random vectors where

$$\Upsilon = \begin{pmatrix} \Upsilon_{\theta\theta} & \Upsilon_{\theta\eta} \\ \Upsilon_{\eta\theta} & \Upsilon_{\eta\eta} \end{pmatrix}.$$

Then the MLEs of  $\theta$  and  $\eta$  are

$$\bar{\theta} = U - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}(V - \bar{\eta}) \quad \text{and} \quad \bar{\eta} = (n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}(n\Upsilon_{\eta\eta}^{-1}V + m\Sigma^{-1}W) \tag{35}$$

**Proof of Lemma 2:**

*Proof* The likelihood is given by

$$\begin{aligned} L(\theta, \eta) &= f(U, V; \theta, \eta)f(W; \eta) \\ &= f(U|V; \theta, \eta)f(V; \eta)f(W; \eta). \end{aligned}$$

Now  $U|V \sim \mathcal{N}_p(\lambda, A)$  with

$$\begin{aligned} \lambda &= \mathbb{E}(U|V) = \theta + \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}(V - \eta), \\ A &= \mathbb{V}(U|V) = n^{-1}(\Upsilon_{\theta\theta} - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}\Upsilon_{\eta\theta}), \end{aligned}$$

so  $\lambda$  is linear in both  $\theta$  and  $\eta$ . Thus we may reparameterize  $f(U|V; \theta, \eta)$  as  $f(U|V; \lambda)$  where

$$f(U|V; \lambda) \propto \exp\{-\frac{1}{2}(U - \lambda)^T A^{-1}(U - \lambda)\}.$$

Also marginally  $V$  follows a  $\mathcal{N}_q(\eta, n^{-1}\Upsilon_{\eta\eta})$  distribution so

$$f(V; \eta)f(W; \eta) \propto \exp\{-\frac{1}{2}(V - \eta)^T n\Upsilon_{\eta\eta}^{-1}(V - \eta)\} \exp\{-\frac{1}{2}(W - \eta)^T m\Sigma^{-1}(W - \eta)\}$$

It now follows that the MLEs for  $(\lambda, \eta)$  are

$$\begin{aligned} \bar{\lambda} &= U, \\ \bar{\eta} &= (n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}(n\Upsilon_{\eta\eta}^{-1}V + m\Sigma^{-1}U). \end{aligned}$$

Thus by the invariance property of MLEs we find that the MLE of  $\theta$  is

$$\bar{\theta} = U - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}(V - \bar{\eta})$$

which completes the proof. □

**Remark 8** To obtain the estimators (7) and (13) apply Lemma 2 and substitute  $\tilde{\theta}$  for  $U$ ,  $\tilde{\eta}$  for  $V$  and  $\hat{\eta}$  for  $W$ . Further substitute  $\tilde{\Upsilon}$  and  $\hat{\Sigma}$  for  $\Upsilon$  and  $\Sigma$ , respectively.

**Proof of Theorem 3.4:**

**Proof** First note that the difference  $\tilde{\eta} - \bar{\eta}_C$  in (13) is a linear combination of  $\tilde{\eta}$  and  $\hat{\eta}$  given by

$$(n\tilde{\Upsilon}_{\eta\eta}^{-1} + m\hat{\Sigma}^{-1})^{-1}m\hat{\Sigma}^{-1}(\tilde{\eta} - \hat{\eta}).$$

Therefore,

$$\begin{pmatrix} \bar{\theta}_C \\ \bar{\eta}_C \end{pmatrix} = \begin{pmatrix} I - \tilde{\Upsilon}_{\theta\eta}\tilde{\Upsilon}_{\eta\eta}^{-1}(n\tilde{\Upsilon}_{\eta\eta}^{-1} + m\hat{\Sigma}^{-1})^{-1}m\hat{\Sigma}^{-1} & \tilde{\Upsilon}_{\theta\eta}\tilde{\Upsilon}_{\eta\eta}^{-1}(n\tilde{\Upsilon}_{\eta\eta}^{-1} + m\hat{\Sigma}^{-1})^{-1}m\hat{\Sigma}^{-1} \\ \mathbf{0} & (n\tilde{\Upsilon}_{\eta\eta}^{-1} + m\hat{\Sigma}^{-1})^{-1}n\tilde{\Upsilon}_{\eta\eta}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\theta} \\ \tilde{\eta} \\ \hat{\eta} \end{pmatrix}.$$

Since  $\Upsilon$  and  $\Sigma$  can be consistently estimated it follows that

$$\begin{pmatrix} \bar{\theta}_C \\ \bar{\eta}_C \end{pmatrix} = \begin{pmatrix} I - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}(n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}m\Sigma^{-1} & \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}(n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}m\Sigma^{-1} \\ \mathbf{0} & (n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}n\Upsilon_{\eta\eta}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\theta} \\ \tilde{\eta} \\ \hat{\eta} \end{pmatrix} + o_p(1).$$

Clearly, the fact that  $n/(n + m) \rightarrow \gamma$  implies that  $(n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}n\Upsilon_{\eta\eta}^{-1} \rightarrow \mathbf{W}_1$  and  $(n\Upsilon_{\eta\eta}^{-1} + m\Sigma^{-1})^{-1}m\Sigma^{-1} \rightarrow \mathbf{W}_2$  so we may rewrite the display above as

$$\begin{pmatrix} \bar{\theta}_C \\ \bar{\eta}_C \end{pmatrix} = \mathbf{M} \begin{pmatrix} \tilde{\theta} \\ \tilde{\eta} \\ \hat{\eta} \end{pmatrix} + o_p(1) \tag{36}$$

where  $\mathbf{M}$  is given in (14). Further observe that

$$\mathbf{M} \begin{pmatrix} \theta_0 \\ \eta_0 \\ \eta_0 \end{pmatrix} = \begin{pmatrix} \theta_0 \\ \eta_0 \end{pmatrix} \tag{37}$$

and that

$$\sqrt{n}(\tilde{\theta} - \theta_0, \tilde{\eta} - \eta_0, \hat{\eta} - \eta_0) \Rightarrow \mathcal{N}_{p+2q}(\mathbf{0}, \mathbf{V}) \tag{38}$$

where  $\mathbf{V}$  is given by (14). Now (36), (37) and (38) together imply that

$$\begin{pmatrix} \bar{\theta}_C - \theta_0 \\ \bar{\eta}_C - \eta_0 \end{pmatrix} \Rightarrow \mathcal{N}_{p+q}(\mathbf{0}, \mathbf{M}\mathbf{V}\mathbf{M}^T) \tag{39}$$

as stated. In particular  $\mathbf{C}_{\theta\theta}$  is the appropriate submatrix of  $\mathbf{M}\mathbf{V}\mathbf{M}^T$ . Multiplying out we find that

$$\mathbf{C}_{\theta\theta} = \Upsilon_{\theta\theta} - \Upsilon_{\theta\eta}\mathbf{W}_2\mathbf{R}^T - \mathbf{R}\mathbf{W}_2\Upsilon_{\eta\theta} + \mathbf{R}\mathbf{W}_2\Upsilon_{\eta\eta}\mathbf{W}_2\mathbf{R}^T + \rho\mathbf{R}\mathbf{W}_2\Sigma\mathbf{W}_2\mathbf{R}^T. \tag{40}$$

The matrices  $\Upsilon_{\eta\eta}$ ,  $\Sigma$  and  $\mathbf{W}_2$  are symmetric and thus their products commute. It follows that  $\mathbf{R}\mathbf{W}_2\Upsilon_{\eta\eta}\mathbf{W}_2\mathbf{R}^T$  equals  $\mathbf{R}\Upsilon_{\eta\eta}\mathbf{W}_2^2\mathbf{R}^T$  and  $\rho\mathbf{R}\mathbf{W}_2\Sigma\mathbf{W}_2\mathbf{R}^T$  equals  $\rho\mathbf{R}\Sigma\mathbf{W}_2^2\mathbf{R}^T$ . It is also easy to verify that  $\Upsilon_{\theta\eta}\mathbf{W}_2\mathbf{R}^T = \mathbf{R}\mathbf{W}_2\Upsilon_{\eta\theta}$  so

$$\Upsilon_{\theta\eta}\mathbf{W}_2\mathbf{R}^T + \mathbf{R}\mathbf{W}_2\Upsilon_{\eta\theta} = 2\Upsilon_{\theta\eta}\mathbf{W}_2\mathbf{R}^T = 2\mathbf{R}\Upsilon_{\eta\eta}\mathbf{W}_2\mathbf{R}^T.$$

Combining and simplifying we obtain

$$\mathbf{C}_{\theta\theta} = \Upsilon_{\theta\theta} - \mathbf{R}\Upsilon_{\eta\eta}\mathbf{S}\mathbf{R}^T.$$

where

$$\mathbf{S} = 2\mathbf{W}_2 - \mathbf{W}_2^2 - \rho\mathbf{W}_2^2\Sigma\Upsilon_{\eta\eta}^{-1}. \tag{41}$$

Now, using symmetry, standard algebraic manipulation and the fact that  $\rho = \gamma/(1 - \gamma)$  we have

$$\begin{aligned} \mathbf{S} &= (\gamma\Upsilon_{\eta\eta}^{-1} + (1 - \gamma)\Sigma)^{-2}\{2(1 - \gamma)\Sigma^{-1}(\gamma\Upsilon_{\eta\eta}^{-1} + (1 - \gamma)\Sigma) - (1 - \gamma)^2\Sigma^{-2} \\ &\quad - \frac{\gamma}{1 - \gamma}\Sigma\Upsilon_{\eta\eta}^{-1}(1 - \gamma)^2\Sigma^{-2}\} \end{aligned}$$

$$\begin{aligned}
 &= (\gamma \mathbf{Y}_{\eta\eta}^{-1} + (1 - \gamma) \mathbf{\Sigma})^{-2} (1 - \gamma) (\gamma \mathbf{Y}_{\eta\eta}^{-1} + (1 - \gamma) \mathbf{\Sigma}) \\
 &= (\gamma \mathbf{Y}_{\eta\eta}^{-1} + (1 - \gamma) \mathbf{\Sigma})^{-1} (1 - \gamma) \mathbf{\Sigma}^{-1} = \mathbf{W}_2
 \end{aligned}$$

Thus  $\mathbf{C}_{\theta\theta} = \mathbf{Y}_{\theta\theta} - \mathbf{R}\mathbf{Y}_{\eta\eta}\mathbf{W}_2\mathbf{R}^T = \mathbf{Y}_{\theta\theta} - \mathbf{Y}_{\theta\eta}\mathbf{Y}_{\eta\eta}^{-1}\mathbf{W}_2\mathbf{Y}_{\theta\eta}^T$  as required. It is also clear that  $\mathbf{C}_{\eta\eta} = \mathbf{B}_{\eta\eta}$  and that

$$\begin{aligned}
 \mathbf{C}_{\theta\eta} &= \lim_n \text{Cov}(\sqrt{n}(\bar{\theta}_C - \theta_0), \sqrt{n}(\bar{\eta}_C - \eta_0)) = \lim_n n\text{Cov}(\bar{\theta}_C, \bar{\eta}_C) \\
 &= \lim_n n\text{Cov}(\tilde{\theta} - \tilde{\mathbf{R}}(\tilde{\eta} - \bar{\eta}), \bar{\eta}) \\
 &= \lim_n n(\text{Cov}(\tilde{\theta}, \bar{\eta}) - \tilde{\mathbf{R}}\text{Cov}(\tilde{\eta} - \bar{\eta}, \bar{\eta}))
 \end{aligned}$$

where  $\tilde{\mathbf{R}} = \tilde{\mathbf{Y}}_{\theta\eta}\tilde{\mathbf{Y}}_{\eta\eta}^{-1}$ . Now  $n\text{Cov}(\tilde{\theta}, \bar{\eta}) = n\text{Cov}(\tilde{\theta}, \mathbf{W}_1\tilde{\eta} + \mathbf{W}_2\hat{\eta} + o_p(1)) \rightarrow \mathbf{Y}_{\theta\eta}\mathbf{W}_1$ . Further note that  $n\text{Cov}(\tilde{\eta} - \bar{\eta}, \bar{\eta}) = n\text{Cov}(\tilde{\eta}, \mathbf{W}_1\tilde{\eta} + \mathbf{W}_2\hat{\eta} + o_p(1)) - n\text{Cov}(\bar{\eta}, \bar{\eta}) \rightarrow \mathbf{Y}_{\eta\eta}\mathbf{W}_1 - (\mathbf{Y}_{\eta\eta}^{-1} + (\rho\mathbf{\Sigma})^{-1})^{-1} = \mathbf{0}$  since

$$\mathbf{Y}_{\eta\eta}\mathbf{W}_1 = \mathbf{Y}_{\eta\eta}(\gamma\mathbf{Y}_{\eta\eta}^{-1} + (1 - \gamma)\mathbf{\Sigma}^{-1})^{-1}\gamma\mathbf{Y}_{\eta\eta}^{-1} = (\mathbf{Y}_{\eta\eta}^{-1} + (\rho\mathbf{\Sigma})^{-1})^{-1}$$

where we have used the fact that  $\rho = \gamma/(1 - \gamma)$ . Thus  $\mathbf{C}_{\theta\eta} = \mathbf{Y}_{\theta\eta}\mathbf{W}_1$  concluding the proof. □

**Proof of Theorem 3.5:**

*Proof* Suppose that  $(\mathbf{U}, \mathbf{V})^T \sim \mathcal{N}((\theta, \eta)^T, \mathbf{Y})$  and  $\mathbf{W} \sim \mathcal{N}(\eta, \rho\mathbf{\Sigma})$  are independent. Let  $I_\omega(\mathbf{U}, \mathbf{V})$  and  $I_\omega(\mathbf{U}, \mathbf{V}, \mathbf{W})$  denote the Fisher Information about  $\omega = (\theta, \eta)^T$  in  $(\mathbf{U}, \mathbf{V})$  and  $(\mathbf{U}, \mathbf{V}, \mathbf{W})$  respectively. It is clear that  $I_\omega(\mathbf{U}, \mathbf{V}) = \mathbf{Y}^{-1}$ . Moreover, repeating the calculations in proofs of Lemma 2 and Theorem 3.4 we deduce that  $I_\omega(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \mathbf{C}^{-1}$ . The additivity of Fisher’s Information implies that

$$I_\omega(\mathbf{U}, \mathbf{V}, \mathbf{W}) \geq I_\omega(\mathbf{U}, \mathbf{V}). \tag{42}$$

Equation (42) and Lemma 1 imply that

$$\mathbf{C} \leq \mathbf{Y}$$

as stated. The fact that  $\mathbf{V}_C^\Phi \leq \mathbf{V}_Y^\Phi$  now follows as in Theorem 3.3. □

**Proof of Theorem 3.6:**

*Proof* By Equation (33) in the proof of Theorem 3.2 we have

$$\sqrt{n}(\bar{\theta}_B - \theta_0) = -(\mathbf{D}_{\theta_0}(\psi))^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\theta_0, \eta_0, Y_i) + \mathbf{D}_{\eta_0}(\psi)\sqrt{n}(\bar{\eta} - \eta_0) \right\} + o_p(1),$$

and similarly,

$$\sqrt{n}(\bar{\theta}_C - \theta_0) = -(D_{\theta_0}(\lambda))^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \lambda(\theta_0, \eta_0, Y_i) + D_{\eta_0}(\lambda) \sqrt{n}(\bar{\eta} - \eta_0) \right\} + o_p(1).$$

The analysis in the proof of Theorem 3.2 shows that in both equations above the terms in the curly brackets are asymptotically independent. Conditions (18) and (17) immediately imply the conclusion of the Theorem.  $\square$

**Remark 9** Recall that  $(\tilde{\theta}, \tilde{\eta})$  simultaneously solve  $\Psi(\theta, \eta) = \mathbf{0}$  and  $\Gamma(\theta, \eta) = \mathbf{0}$  where  $\Gamma(\theta, \eta) = n^{-1} \sum_{i=1}^n \gamma(\theta, \eta, Y_i)$ . Standard calculations show that

$$\begin{pmatrix} \tilde{\theta} \\ \tilde{\eta} \end{pmatrix} = \begin{pmatrix} \theta_0 \\ \eta_0 \end{pmatrix} + D^{-1} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \psi(\theta, \eta, Y_i) \\ \frac{1}{n} \sum_{i=1}^n \gamma(\theta, \eta, Y_i) \end{pmatrix} + o_p(1) \tag{43}$$

where

$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} = \begin{pmatrix} D_{\theta}(\psi) & D_{\eta}(\psi) \\ D_{\theta}(\gamma) & D_{\eta}(\gamma) \end{pmatrix}.$$

Using the above notations and rewriting Eq. (33) we have

$$\bar{\theta}_B = \theta_0 + \frac{1}{n} \sum_{i=1}^n D_{11}^{-1} \psi(\theta, \eta, Y_i) - D_{11}^{-1} D_{12} (\bar{\eta} - \eta_0) + o_p(1). \tag{44}$$

As demonstrated in the proof of Theorem 3.2 the two terms above are asymptotically independent so we can re-express  $\bar{\theta}_B$  as

$$\bar{\theta}_B = \theta_0 + \frac{1}{n} \sum_{i=1}^n \{ D_{11}^{-1} \psi(\theta, \eta, Y_i) - D_{11}^{-1} D_{12} Q_i \} + o_p(1) \tag{45}$$

where  $Q_i$  are IID  $\mathcal{N}(\mathbf{0}, W_1 \Upsilon_{\eta\eta} W_1^T + \rho W_2 \Sigma W_2^T)$  are random variables which are independent of  $\mathcal{D}$ . Further note that by (43)

$$\begin{aligned} \tilde{\theta} &= \theta_0 + D^{11} \frac{1}{n} \sum_{i=1}^n \psi(\theta, \eta, Y_i) + D^{12} \frac{1}{n} \sum_{i=1}^n \gamma(\theta, \eta, Y_i) + o_p(1), \\ \tilde{\eta} &= \eta_0 + D^{21} \frac{1}{n} \sum_{i=1}^n \psi(\theta, \eta, Y_i) + D^{22} \frac{1}{n} \sum_{i=1}^n \gamma(\theta, \eta, Y_i) + o_p(1), \end{aligned}$$

where  $D^{ij}$  is the appropriate submatrix of  $D^{-1}$ . Substituting the formulas above into Eq. (13) for  $\bar{\theta}_C$  and simplifying we find that

$$\bar{\theta}_C = \theta_0 + \frac{1}{n} \sum_{i=1}^n \{ (D^{11} - \Upsilon_{\theta\eta} \Upsilon_{\eta\eta}^{-1} D^{21}) \psi(\theta, \eta, Y_i) + (D^{21} - \Upsilon_{\theta\eta} \Upsilon_{\eta\eta}^{-1} D^{22}) \gamma(\theta, \eta, Y_i) \}$$

$$+\Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}\mathbf{Q}_i\} + o_p(1). \quad (46)$$

Therefore comparing the estimators  $\bar{\theta}_B$  and  $\bar{\theta}_C$  amounts to comparing their influence functions implicit in (45) and (46), i.e.,

$$\mathbf{D}_{11}^{-1}\boldsymbol{\psi}(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{Y}_i) - \mathbf{D}_{11}^{-1}\mathbf{D}_{12}\mathbf{Q}_i$$

and

$$(\mathbf{D}^{11} - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}\mathbf{D}^{21})\boldsymbol{\psi}(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{Y}_i) + (\mathbf{D}^{21} - \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}\mathbf{D}^{22})\boldsymbol{\gamma}(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{Y}_i) + \Upsilon_{\theta\eta}\Upsilon_{\eta\eta}^{-1}\mathbf{Q}_i,$$

respectively. Also note that

$$\boldsymbol{\Upsilon} = (\mathbf{D}^{-1}) \begin{pmatrix} \mathbb{E}_0(\boldsymbol{\psi}\boldsymbol{\psi}^T) & \mathbb{E}_0(\boldsymbol{\psi}\boldsymbol{\gamma}^T) \\ \mathbb{E}_0(\boldsymbol{\gamma}\boldsymbol{\psi}^T) & \mathbb{E}_0(\boldsymbol{\gamma}\boldsymbol{\gamma}^T) \end{pmatrix} (\mathbf{D}^{-1})^T$$

so although in principle it is possible to always compare the above influence functions in practice this comparison is very difficult unless some further simplifying assumptions are imposed.

## References

- Anderson TW (1957) Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J Am Stat Assoc* 52:200–203
- Ansari MA, Shah FA, Butt TM (2008) Combined use of entomopathogenic nematodes and *Metarhizium anisopliae* as a new approach for black vine weevil, *Otiiorhynchus sulcatus*, control. *Entomol Exp Appl* 129:340–347
- Becker NG (2017) Analysis of infectious disease data. Routledge, London
- Benichou J, Gail MH (1989) A delta method for implicitly defined random variables. *Am Stat* 43:41–44
- Bliss CI (1939) The toxicity of poisons applied jointly. *Ann Appl Biol* 26:585–615
- Bryan JG, Hoff PD (2020). Smaller  $p$ -values in genomics studies using distilled historical information. arXiv preprint [arXiv:2004.07887](https://arxiv.org/abs/2004.07887)
- Chen YH, Chen H (2000) A unified approach to regression analysis under double-sampling designs. *J R Stat Soc Ser B* 62:449–460
- Cochran WG (1954) The combination of estimates from different experiments. *Biometrics* 10:101–129
- Cooper I, Mondal A, Antonopoulos CG (2020) A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* 139:110057
- Davidov O, Haitovsky Y (2000) Optimal design for double sampling with continuous outcomes. *J Stat Plan Inference* 86:253–263
- Davidov O, Zelen M (2004) Overdiagnosis in early detection programs. *Biostatistics* 5:603–613
- Doudich M, Ezzrari A, Van der Weide R, Verme P (2016) Estimating quarterly poverty rates using labor force surveys: a primer. *World Bank Econ Rev* 30:475–500
- Demidenko E, Miller TW (2019) Statistical determination of synergy based on the definition of the Bliss drugs independence. *PLoS ONE* 14(11):e0224137
- Edmunds WJ, O'Callaghan CJ, Nokes DJ (1997) Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc R Soc Lond Ser B* 264:949–957
- Food and Drug Administration (2006) Fixed dose combinations, co-packaged drug products, and single-entity versions of previously approved antiretrovirals for the treatment of HIV. Guidance to Industry. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Available at <https://www.fda.gov/media/72248/download>



- Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552
- Goeyvaerts N, Hens N, Ogunjimi B, Aerts M, Shkedy Z, Damme PV, Beutels P (2010) Estimating infectious disease parameters from data on social contacts and serological status. *J Ro Stat Soc Ser C* 59:255–277
- GOV.UK Department of Work and Pensions (2019) Family Resources Survey. UK Department of Work and Pensions. Available at <https://www.gov.uk/government/collections/family-resources-survey-2>
- Heyde CC (2008) Quasi-likelihood and its application: a general approach to optimal parameter estimation. Springer, New York
- Hidiroglou MA (2001) Double sampling. *Surv Methodol* 27:143–154
- Hoff PD (2019) Smaller  $p$ -values via indirect information. arXiv preprint [arXiv:1907.12589](https://arxiv.org/abs/1907.12589)
- Kanda S, Goto K, Shiraishi H, Kubo E, Tanaka A, Utsumi H, Sunami K, Kitazono S, Mizugaki H, Horinouch H (2016) Safety and efficacy of nivolumab and standard chemotherapy drug combination in patients with advanced non-small-cell lung cancer: a four arms phase Ib study. *Ann Oncol* 27(12):2242–2250
- Kanda T, Fujikoshi Y (1998) Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. *Am J Math Manag Sci* 18:161–192
- Kim JK, Rao JNK (2012) Combining data from two independent surveys: a model assisted approach. *Biometrika* 99:85–100
- Kogan Y, Halevi-Tobias K, Elishmereni M, Vuk-Pavlović S, Agur Z (2012) Reconsidering the paradigm of cancer immunotherapy by computationally aided real-time personalization. *Cancer Res* 72:2218–2227
- Kronik N, Kogan Y, Elishmereni M, Halevi-Tobias K, Vuk-Pavlović S, Agur Z (2010) Predicting outcomes of prostate cancer immunotherapy by personalized mathematical models. *PLoS ONE* 5:e15482
- Kozłowska E, Färkkilä A, Vallius T, Carpén O, Kemppainen J, Grénman S, Hautaniemi S (2018) Mathematical modeling predicts response to chemotherapy and drug combinations in ovarian cancer. *Cancer Res* 78:4036–4044
- Lee S (2010) Drug interaction: focusing on response surface models. *Korean J Anesthesiol* 58:421–434
- Lee KM, Wason J (2020) Including non-concurrent control patients in the analysis of platform trials: is it worth it? *BMC Med Res Methodol* 20:1–12
- Lee SJ, Zelen M (1998) Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *J Am Stat Assoc* 93:1271–1281
- Liu Q, Yin X, Languino LR, Altieri DC (2018) Evaluation of drug combination effect using a Bliss independence dose-response surface model. *Stat Biopharm Res* 10(2):112–122
- Moore JC, Stinson LL, Welniak LJ (2000) Income measurement errors in surveys: a review. *J Off Stat* 16:331–361
- Morrison DF (1971) Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. *J Am Stat Assoc* 66:602–604
- Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Heijne J (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5(3):74
- Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ (2010) Summarizing historical information on controls in clinical trials. *Clin Trials* 7:5–18
- Newey WK, Ramalho JJ, Smith RJ (2005). Asymptotic bias for GMM and GEL estimators with estimated nuisance parameters. *Identif Inference Econom Models* 245–281
- Niu J, Straubinger RM, Mager DE (2019) Pharmacodynamic drug–drug interactions. *Clin Pharmacol Ther* 105:1395–1406. <https://doi.org/10.1002/cpt.1434>
- Pallmann P, Schaarschmidt F (2016) Common pitfalls when testing additivity of treatment mixtures with  $\chi^2$ . *J Appl Entomol* 140:135–141
- Palmer AC, Sorger PK (2017) Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell* 171:1678–1691
- Peddada SD, Dinse GE, Kissling GE (2007) Incorporating historical control data when comparing tumor incidence rates. *J Am Stat Assoc* 102:1212–1220
- Peto R, Pike M, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG (1976) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *Br J Cancer* 34:585–612
- Piantadosi S (2017) Clinical trials: a methodological perspective. Wiley, New York
- Pierce DA (1982) The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann Stat* 10:475–478
- Pocock SJ (1976) The combination of randomized and historical controls in clinical trials. *J Chronic Dis* 29:175–188

- Pukelsheim F (2006). Optimal design of experiments. *Soc Ind Appl Math*
- Raftery AE, Alkema L, Gerland P (2014) Bayesian population projections for the United Nations. *Stat Sci* 29:58–68
- Randles RH (1982) On the asymptotic normality of statistics with estimated parameters. *Ann Stat* 1:462–474
- Ridder G, Moffitt R (2007) The econometrics of data combination. *Handbook Econ* 6:5469–5547
- Roig MB, Krotka P, Burman CF, Glimm E, Gold SM, Hees K, Jacko P, Koenig F, Magirr D, Mesenbrink P, Viele K (2022) On model-based time trend adjustments in platform trials with non-concurrent controls. *BMC Med Res Methodol* 22:1–16
- Rukhin AL (2007) Estimating a common vector parameter in interlaboratory studies. *J Multivar Anal* 98:435–454
- Russ D, Kishony R (2018) Additivity of inhibitory effects in multidrug combinations. *Nat Microb* 3. <https://doi.org/10.1038/s41564-018-0252-1>
- Tamhane AC (1978) Inference based on the regression estimator in double sampling. *Biometrika* 65:419–427
- Tamma PD, Cosgrove SE, Maragakis LL (2012) Combination therapy for treatment of infections with gram-negative bacteria. *Clin Microbiol Rev* 25:450–470
- Tasseva IV (2019) The changing education distribution and income inequality in great Britain. *Euromod Working Paper Series EM 16/19*, University of Essex, available at <https://www.euromod.ac.uk/sites/default/files/working-papers/em16-19.pdf>
- Thompson ME (1997) *Theory of sample surveys*. Chapman and Hall, London
- United Nations, Department of Economic and Social Affairs, Population Division (2014) *World Population Prospects: The 2012 Revision, Methodology of the United Nations Population Estimates and Projections*. ESA/P/WP.235
- Van der Vaart AW (2000) *Asymptotic statistics*. Cambridge University Press, Cambridge
- Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Micallef S (2014) Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat* 13:41–54
- Wallinga J, Teunis P, Kretzschmar M (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol* 164:936–944
- Yaari R, Dattner I, Huppert A (2018) A two-stage approach for estimating the parameters of an age-group epidemic model from incidence data. *Stat Methods Med Res* 27:1999–2014

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