Received: 8 July 2019 | Revised: 8 March 2021 | Accepted: 27 March 2021

DOI: 10.1111/sjos.12530

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Combining information across diverse sources: The II-CC-FF paradigm

Céline Cunen | **Nils Lid Hjort**

Department of Mathematics, University of Oslo, Oslo, Norway

Correspondence
Céline Cunen, Department of Mathematics, University of Oslo, Oslo, Norway.
Email: cmlcunen@math.uio.no

Funding information
Norges Forskningsråd, Grant/Award Number: 235116

Abstract

We introduce and develop a general paradigm for combining information across diverse data sources. In broad terms, suppose ϕ is a parameter of interest, built up via components $\psi_1, \ldots, \psi_k$ from data sources $1, \ldots, k$. The proposed scheme has three steps. First, the independent inspection (II) step amounts to investigating each separate data source, translating statistical information to a confidence distribution (CD) $C_j(\psi_j)$ for the relevant focus parameter ψ_j associated with data source j. Second, confidence conversion (CC) techniques are used to translate the CDs to confidence log-likelihood functions. Finally, the focused fusion (FF) step uses relevant and context-driven techniques to construct a confidence distribution for the primary focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$, acting on the combined confidence log-likelihood. In traditional setups, the II-CC-FF strategy amounts to versions of meta-analysis, and turns out to be competitive against state-of-the-art methods. Its potential lies in applications to harder problems, however. Illustrations are presented, related to actual applications.

KEYWORDS

combining information, confidence distributions, confidence likelihoods, focused fusion, hard and soft data, meta-analysis

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2021 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

1 | COMBINING INFORMATION AND THE II-CC-FF SCHEME

Our paper concerns the statistical task of combining information across different and perhaps very diverse data sources. This is of course a long-standing theme in statistics, with papers going back to Karl Pearson (cf. Simpson and Pearson, 1904); see Schweder & Hjort (2016, ch. 13) for background and a general discussion of themes traditionally sorted under the bag-word meta-analysis, along with further basic references. The present paper aims at proposing and developing a certain paradigm, which we call the II-CC-FF method, meant to be powerfully applicable for ranges of situations far beyond the usual simpler setups. We will explain the role and nature of the independent inspection (II), confidence conversion (CC), focused fusion (FF) steps below.

A special case worth considering first is the textbook setup where $y_1, \ldots, y_k$ are independent estimators of the same quantity ψ, and where $y_j \sim \mathrm{N}(\psi, \sigma_j^2)$, with known standard deviations (SDs) σ_j. An easy exercise in minimizing variances shows that the optimally balanced overall estimator for the focus parameter is

$$\widehat{\psi} = \frac{\sum_{j=1}^k y_j/\sigma_j^2}{\sum_{j=1}^k 1/\sigma_j^2} \sim \mathrm{N}\left(\psi, \left(\sum_{j=1}^k 1/\sigma_j^2\right)^{-1}\right). \tag{1}$$

A natural extension, though harder to analyze to full satisfaction, is when $y_j \sim \mathrm{N}(\psi_j, \sigma_j^2)$, with the individual means ψ_j differing according to a $\mathrm{N}(\psi_0, \tau^2)$ distribution. For this type of random effects models, one wishes clear inference strategies for both the overall mean ψ_0 and level of variation τ. We return to this particular problem in Sections 6.1, 7.1, and 8.1.

Many problems of modern statistics involving combination of information are much more complicated than the situations sketched above, however. Sometimes one needs to combine "hard" data, with clear measurements from controlled experiments, and so forth, with "soft" data, associated with information more loosely connected to the parameters of primary interest, perhaps via measurement errors or surrogate variables. In addition, there might be prior distributions available, via subject matter experts, but only for some of the parameters at play, not enough to make it into a clear Bayesian analysis. For our development of II-CC-FF, we have attempted to think fundamentally and generally about combination of information problems. Our framework encompasses known meta-analysis methods, but we aim at tackling new and more challenging problems as well. Parts of the meta-analysis literature are quite narrow, with specific methods for specific problems. In that light, we hope our more general approach will be useful.

In reasonably general terms, assume there is a parameter ϕ of clear interest, related to parameters $\psi_1, \ldots, \psi_k$, either via a deterministic function $\phi = \phi(\psi_1, \ldots, \psi_k)$ or via some type of random effect distribution, where such a ϕ might be a parameter related to a background distribution of the ψ_j. Suppose further that data source y_j provides information pertaining to ψ_j. For the sake of clear presentation, we let the ψ_j be one-dimensional here. Our II-CC-FF approach for reaching inference statements for the overall focus parameter ϕ can then be schematically set up as follows:

- ◇ II, *Independent Inspection*: Data source y_j is used, via appropriate models and analyses, to yield a confidence distribution (CD) $C_j(\psi_j, y_j)$ for the main interest parameter associated with study j.
- ◇ CC, *Confidence Conversion*: The CD is converted into a log-likelihood function for this main parameter of interest for study j, say $\ell_{\mathrm{conv},j}(\psi_j)$.

◇ FF, *Focused Fusion*: In the fixed effect case, the combined confidence log-likelihood function $\ell_{\rm fus}(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{{\rm conv},j}(\psi_j)$ is used to reach FF inference for $\phi = \phi(\psi_1, \ldots, \psi_k)$. With random effects, the fusion involves the computation of an integral.

We do make use of certain subscripts in our paper, meant as helpful signposting. The subscript "conv" is for likelihood functions coming out the CC-step; "fus" relates to the FF-step; while "prof" and "cprof" are used for the profile log-likelihood and its corrected version (see Section 5.2).

The extent to which some or all of these steps will be relatively straightforward or rather complicated to carry out depends to a high degree on the special features of the given source combination problem. The steps are not "isolated" or fully separated, but often related. In Section 5.1, we provide a standardized version of II-CC-FF, with a generic recipe to follow, but we will see that in many cases one should be more careful about the various steps. In situations where the statistician has all the raw data and the particular models used for analyzing the different sources of information, the CC step is in a conceptual sense not difficult, as the required log-likelihood parts may be worked out from first principles. In various situations confronting the modern statistician this is rather more difficult, however, as one might have to base one's analysis on summary measures, directly or indirectly given via other people's work, reports, and publications. The II-CC-FF paradigm is meant to be powerfully applicable in such situations too.

A pertinent question is whether or why there is a need for specific methods for combination of information in the first place; in a suitable sense, all of statistics concerns combination of information. One might therefore ask why there even exist subfields such as meta-analysis, and specific framework aimed at combination of information such as our own. So isn't meta-analysis just analysis? Two related responses are as follows. (i) Sometimes the full sets of data are not available, with access only to summaries or partial summaries. Issues here are storage, the practicalities of other people's files, privacy concerns, and so forth. (ii) Sometimes it might be easier, conceptually or practically, to analyze the different sources or studies separately first, and then combine these pieces of summarized information. In addition, a statistical prediction is that modern statistics to an increasing degree will be concerned with such issues and challenges, finding and organizing bits and pieces of information across different sources, with a need to reach conclusions based on these pieces.

After a motivating illustration, below, we start in Section 2 with a brief review of CDs, which are essential for the II part of the framework. We then proceed with giving details related to the basics of CC in Section 3 and FF in Section 4. In Section 5, we provide a standard version of our II-CC-FF framework, and investigate some pitfalls and solutions. In Section 6, we investigate the use of our II-CC-FF scheme in well-established meta-analysis situations. Further performance and comparison issues are examined in Section 7 via simulations. There we find that II-CC-FF methods are competitive in several traditional meta-analysis settings. Finally, the three-step II-CC-FF machinery is seen in action through four applications laid out in Section 8. We also have appendixes A, B, C, D. In Appendix D connections with other CD-based approaches are explored, and some of these methods are compared with II-CC-FF through decision-theoretic risk functions.

1.1 | Motivating illustration

The following concrete illustration, which has certain features placing it outside the usual meta-analysis setups and methods, shows the three steps of the II-CC-FF at work, but with a

minimum of details. The nonstandard aspect of this illustration is partly that different studies of the same statistical question have reported different summary measures—six studies (call them type A) have reported summary statistics based on continuous outcomes, while five other studies (which we call type B) reported summaries based on a binary outcome. More crucially, the focus parameter β in question, a regression coefficient related to the difference between treatments, is not identifiable, and hence cannot be estimated directly, for the type B studies. In fact, these type B studies only inform us about a certain β/σ, where also σ is not identifiable, or estimable, from those studies. In addition, the raw data, for these studies, are not available. The data employed here were first analyzed in Whitehead et al. (1999); related problems have been treated in Dominici and Parmigiani (2000) and Liu et al. (2015).

We have 11 randomized trials investigating the use of oxytocic drugs during labor and its potential effect on postpartum blood loss. Each study has two groups of patients, a treatment group receiving oxytocic drug and a control group receiving no drugs of that type. Taking $y_{i,j}$ to be the blood loss for patient i in study j, we may use the simple model $y_{i,j} = \alpha_j + \beta z_{i,j} + \varepsilon_{i,j}$, with the $\varepsilon_{i,j}$ independent and $\mathrm{N}(0, \sigma^2)$, and with $z_{i,j}$ an indicator variable, equal to 0 for patients in the control group and 1 for patients in the treatment group. Here, β is the treatment effect and the parameter of main interest.

For the six type A trials, we have the mean and the empirical SD of the blood loss in the two groups of patients. With the simple normal model above, these four summary statistics are sufficient for each trial, and we thus have access to the full log-likelihood $\ell_{A,j}(\beta, \alpha_j, \sigma)$ for each continuous trial j. For the five type B trials, however, we merely have counts of the number of patients in each group having a blood loss of more or less than 500 ml. These numbers constitute a nonsufficient summary; we thus have less information in these studies compared with the continuous ones, and the log-likelihood functions do not inform on β directly, only on β/σ. More specifically, based on the normal model above, we obtain a probit-type log-likelihood for these binary trials, say $\ell_{B,j}(\theta, \gamma_j)$, with $\gamma_j = (500 - \alpha_j)/\sigma$ and $\theta = \beta/\sigma$.

Having made these modeling assumptions, the steps in the II-CC-FF recipe follow straightforwardly. Using the log-likelihood functions described above, we can, by methods described in the following section, construct *confidence curves* for the parameter of interest for each of the studies.

The CC step is simple in this case, with no extra work required, since the log-likelihood functions were used in the construction of the confidence curves for each study. In other situations we might have to carry out the conversion from confidence statements to log-likelihood functions in ways described in Section 3. Via arguments explained in more detail in the following section, we reach log-likelihood contributions $\ell_{A,\mathrm{prof},j}(\beta, \sigma)$ for the continuous studies and $\ell_{B,\mathrm{prof},j}(\beta/\sigma)$ for the binary studies. In the FF step these are summed, to reach

$$\text{FF:} \quad \ell_{\mathrm{fus}}(\beta, \sigma) = \sum_{j=1}^{6} \ell_{A,\mathrm{prof},j}(\beta, \sigma) + \sum_{j=1}^{5} \ell_{B,\mathrm{prof},j}(\beta/\sigma).$$

Next, we profile out σ and obtain the final combined confidence curve by

$$\mathrm{cc}^*(\beta, \text{all data}) = \Gamma_1(2\{\ell_{\mathrm{fus}}(\widehat{\beta}, \widehat{\sigma}) - \ell_{\mathrm{fus}}(\beta, \widehat{\sigma}(\beta))\}),$$

with $\Gamma_1(\cdot)$ the c.d.f. of a χ^2_1. In Figure 1, the thick red curve is this combined confidence curve. It is clearly narrower than all the individual curves and placed roughly in the middle of them, as we would expect.

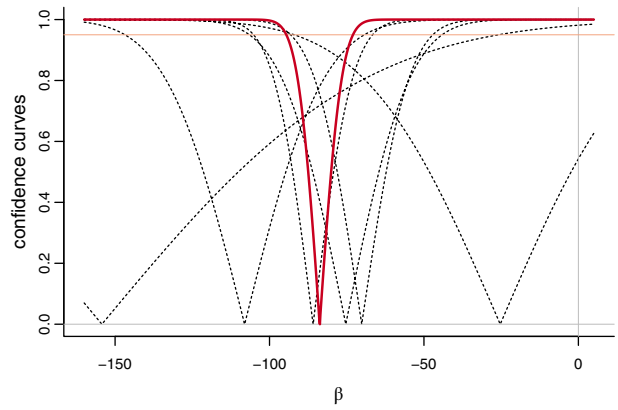


FIGURE 1 Confidence curves for the treatment effect in the six continuous trials (dashed, black). In red, the confidence curve combining all the 11 studies. The horizontal red line marks the 95% confidence level. The median confidence estimate is −83.7 ml, with 95% interval [−94.4, −73.1] [Color figure can be viewed at wileyonlinelibrary.com]

The combined inference clearly demonstrates that oxytocic drugs reduce postpartum blood loss, which is in agreement with the conclusions in Whitehead et al. (1999). Here, we have zoomed in on β as the focus parameter, to pinpoint precisely how much the two groups differ in blood loss. For clinicians, it might be of more direct interest to consider the probabilities for having a postpartum blood loss greater than a threshold, like 500 ml, for the two groups, and then focus on the odds ratio, say ρ. Our approach can easily accommodate such an analysis too, with ρ rather than β in the FF step, yielding a figure similar to Figure 1, but now for ρ. The II, CC, FF steps are not intended to form a unique recipe, as there are individual variations, depending on the application at hand. In the FF step above the log-likelihood contributions for type A and type B information was arrived at via profiling of the fuller log-likelihoods constructed under the auspices of the regression model we started out with. In other applications this would not be possible, and there would be a need to convert CC information to log-likelihoods, a theme we examine in Section 3.

2 | INDEPENDENT INSPECTION: CONFIDENCE DISTRIBUTIONS

Suppose Y_j denotes a set of random observations from data source j, stemming from a model with parameter θ_j, often multidimensional, and with $\psi_j = \psi(\theta_j)$. For the ease of presentation, we let ψ_j be a one-dimensional focus parameter for now, but in general combination situations it will typically be multidimensional. A *confidence distribution* $C_j(\psi_j, y_j)$ for this focus parameter from source j has the properties (i) it is a cumulative distribution function (c.d.f.) in ψ_j, for each y_j, and (ii) at the true value θ_0, with associated true value $\psi_0 = \psi(\theta_0)$, the distribution of $C_j(\psi_0, Y_j)$ is uniform on the unit interval. From this follows, under standard continuity and monotonicity assumptions, that

$$P_{\theta_0}\{C_j^{-1}(0.05, Y_j) \le \psi_0 \le C_j^{-1}(0.95, Y_j)\} = 0.90,$$

and so forth, that is, $[C_j^{-1}(0.05, y_{j,\mathrm{obs}}), C_j^{-1}(0.95, y_{j,\mathrm{obs}})]$ is a 90% confidence interval for ψ_j, where $y_{j,\mathrm{obs}}$ denotes the observed dataset. Thus, the CD $C_j(\psi_j, y_{j,\mathrm{obs}})$, qua random c.d.f., is a compact and convenient representation of confidence intervals at all levels, and indeed a powerful inference summary. A close relative is the *confidence curve*, which we tend to prefer as a postdata graphical summary of information for focus parameters, defined as

$$\mathrm{cc}_j(\psi_j, y_{j,\mathrm{obs}}) = |1 - 2\, C_j(\psi_j, y_{j,\mathrm{obs}})|. \tag{2}$$

It points to its cusp point, the median confidence point estimate $\widehat{\psi}_{j,0.50} = C_j^{-1}(\frac{1}{2}, y_{j,\mathrm{obs}})$, and the two roots of the equation $C(\psi_j, y_{j,\mathrm{obs}}) = \alpha$ form a confidence interval with this confidence level. Degrees of asymmetry are easier to spot and to convey using the confidence curve than with the cumulative CD itself; cf. illustrations in Section 8. We also note that the random $\mathrm{cc}_j(\psi_j, Y_j)$ has a uniform distribution, at the true position in the parameter space, since $|1 - 2\,U|$ is uniform when U is. Indeed

$$P_{\theta_0}\{\mathrm{cc}_j(\psi_0, Y_j) \le \alpha\} = \alpha, \quad \text{for each } \alpha \tag{3}$$

at the true parameters of the model. The confidence curve is arguably a more fundamental concept than the confidence distribution, as there are cases where a natural $\mathrm{cc}_j(\psi_j, Y_j)$ may be constructed, with a valid (3), even when confidence regions are formed by disjoint intervals (as with multimodal log-likelihood functions).

For an extensive treatment of CDs, their constructions in different types of setup, properties and uses, see Schweder and Hjort (2016), and the review paper Xie and Singh (2013), with ensuing discussion contributions. The scope and broad applicability of CDs are also demonstrated in a collection of papers published in the special issue *Inference With Confidence*of the journal *Journal of Statistical Planning and Inference*, 2018 (Hjort & Schweder, 2018). Here, we shall merely point to two important and broadly useful ways of constructing a confidence distribution, for a focus parameter ψ_j, based on data from a model with a multidimensional parameter θ_j. The first is to rely on an approximately normally distributed estimator, if available, say $\widehat{\psi}_j \sim \mathrm{N}(\psi_j, \kappa_j^2)$, and with SD well estimated with an appropriate $\widehat{\kappa}_j$. Then, with $\Phi(\cdot)$ as usual denoting the c.d.f. of the standard normal, $C_j(\psi_j, y_j) = \Phi((\psi_j - \widehat{\psi}_j)/\widehat{\kappa}_j)$ is an approximately correct CD, first-order large-sample correct under weak regularity conditions. In particular the estimator used can be the maximum likelihood one (ML), say $\widehat{\psi}_{j,\mathrm{ml}}$, but other estimators are allowed too in this simple construction. The second is based on the profiled log-likelihood function $\ell_{\mathrm{prof},j}(\psi_j) = \max\{\ell_j(\theta_j) : \psi(\theta_j) = \psi_j\}$, which leads to the deviance function

$$D_j(\psi_j) = 2\{\ell_{\mathrm{prof},j}(\widehat{\psi}_{j,\mathrm{ml}}) - \ell_{\mathrm{prof},j}(\psi_j)\} = 2\{\ell_{\mathrm{prof},j,\max} - \ell_{\mathrm{prof},j}(\psi_j)\}. \tag{4}$$

As laid out in Schweder & Hjort (2016, chs. 2, 3), the Wilks theorems with variations, see Section 4.3, then lead naturally to

$$\mathrm{cc}_j(\psi_j, y_j) = \Gamma_1(D_j(\psi_j)), \tag{5}$$

with $\Gamma_\nu(\cdot)$ denoting the c.d.f. of a χ^2 with degrees of freedom ν. We will sometimes refer to the confidence curve construction above as the *Wilks approximation* in the following. Typically, the second method (5) leads to a better-calibrated confidence curve than the the simpler method mentioned first.

3 | CONFIDENCE CONVERSION: FROM CONFIDENCE TO LIKELIHOODS

Several well-explored methods, with appropriate variations and amendments, lead from likelihood functions to CDs and confidence curves; cf. again several chapters of Schweder and

Hjort (2016). Sometimes the CC step comes almost for free, in cases where the statistician can compute say log-likelihood profiles from raw data, or from sufficient statistics, for the given models. But in some cases the CC step of the II-CC-FF paradigm requires methods for going the other way, from CDs or confidence curves to log-likelihood information, and this is more involved. Among the complications is that different experimental protocols, with ensuing different CDs, might be having the same log-likelihood functions, so the link between confidence and likelihood is not one-to-one.

Schweder & Hjort (2016, ch. 10) develop and discuss this topic at some length. For the present purposes, we shall be content with what we call the *chi-squared inversion*, associated with (5) above. Assume that all our information about a parameter ψ_j from source j comes in the form of the confidence curve $\mathrm{cc}_j(\psi_j, y_j)$. Then we can obtain a confidence log-likelihood contribution from source j by the following formula,

$$\ell_{\mathrm{conv},j}(\psi_j) = -\frac{1}{2}\Gamma_1^{-1}(\mathrm{cc}_j(\psi_j, y_j)). \tag{6}$$

If the confidence curve has been constructed via the second general method presented in the previous section, see (5), we will of course simply get back the profiled log-likelihood function. The CC step therefore only comes into play when confidence curve is constructed via nonstandard methods, as we will see in Application 8.3, or when the only available information from source j is the confidence curve itself and we do not know exactly how it was constructed.

When a CD is available, rather than a confidence curve, one can use the *normal conversion* $\ell_{\mathrm{conv},j}(\psi_j) = -\frac{1}{2}\{\Phi^{-1}(C_j(\psi_j, y_j))\}^2$. This is equivalent to the recipe in (6) when the confidence curve has been constructed via $\mathrm{cc}_j(\psi_j, y_j) = |1 - 2\, C_j(\psi_j, y_j)|$. A relevant point here is that one often constructs a confidence curve $\mathrm{cc}_j(\psi_j, y_j)$ directly, not always via (2), making (6) a more versatile tool. The normal conversion confidence likelihood is also what Efron (1993) proposed, for coming from confidence to likelihood, via different arguments and for different purposes; see also Efron & Hastie (2016, ch. 11). For details on how well the chi-squared inversion methods works, in different scenarios, see Section 4.3.

In some situations one is able to construct a CD for source j via a one-dimensional statistics T_j, instead of using the general method from (5). Then one may use *exact conversion* to obtain the confidence log-likelihood. When the statistic has a continuous distribution, the exact conversion of the CD $C_j(\psi_j, T_j)$, see Schweder & Hjort (2016, ch. 10), is $\ell_{\mathrm{conv},j}(\psi_j) = \log\ |\partial C_j(\psi_j, t)/\partial t|$.

4 | FOCUSED FUSION: FROM FULL LIKELIHOOD TO FOCUS PARAMETER

Suppose now that the II and CC steps have been successfully carried out, leading to confidence log-likelihood contributions $\ell_{\mathrm{conv},j}(\psi_j)$ from information sources $j = 1, \ldots, k$. Depending on the application and its context we might then be interested in either a fixed effect approach, with the main focus parameter ϕ is a function of the ψ_j, or a random effect approach, where we introduce an additional layer of heterogeneity through a model for the ψ_j.

4.1 | Fixed effects fusion

Assuming the information sources to be independent, the overall confidence log-likelihood function is $\ell_{\rm fus}(\psi_1, \ldots, \psi_k) = \sum_{j=1}^k \ell_{{\rm conv},j}(\psi_j)$. When focused inference is wished for, for a focus parameter $\phi = \phi(\psi_1, \ldots, \psi_k)$, the natural way forward is, again, via profiling:

$$\ell_{\rm fus,prof}(\phi) = \max\{\ell_{\rm fus}(\psi_1, \ldots, \psi_k) : \phi(\psi_1, \ldots, \psi_k) = \phi\}.$$

By the Wilks theorem directly, or by variations of the arguments and details used to prove such theorems (cf. Schweder & Hjort, 2016, Appendix), the overall deviance function

$$D^*(\phi) = 2\{\ell_{\rm fus,prof}(\widehat{\phi}) - \ell_{\rm fus,prof}(\phi)\}$$

tends, at the true parameter position and with increasing information volume, to a χ^2_1. Here, $\widehat{\phi}$ is the ML, maximizing the profiled log-likelihood. Hence

$$\mathrm{cc}^*(\phi, \text{all data}) = \Gamma_1(D^*(\phi)) \tag{7}$$

is the outcome of the three-step II-CC-FF machine, a confidence curve for the focus parameter. In Section 4.3, we will come back to some discussion on the meaning of "increasing information volume" in a combination context. In situations where the ψ_j represent the same focus parameter, common across sources, the scheme above simplifies.

4.2 | Random effects fusion

In our II-CC-FF setting, we use the term "random effects" when we wish to introduce an extra layer of heterogeneity in the fusion step. This is more easily presented when assuming that $\psi_1, \ldots, \psi_k$ are scalars. In the random effects case, we do not assume that all ψ_j are equal but rather that they come from some underlying distribution. In the most canonical case, this distribution will be governed by some overall mean parameter ψ_0 and some spread parameter τ; specifically we could have $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. The parameter of main interest may be either the overall mean, or the spread, or perhaps a quantile, depending on the context.

We propose the following general solution for II-CC-FF with random effects. Suppose the ψ_j are modeled as coming from a background density $f(\psi_j, \kappa)$, say, where the κ could be a center and a spread parameter, as for (ψ_0, τ) in the normal case. Then, using the confidence log-likelihoods $\ell_{{\rm conv},j}(\psi_j)$ from each source, we define the fusion log-likelihood to be

$$\ell_{\rm fus}(\kappa) = \sum_{j=1}^k \log\left[\int \exp\{\ell_{{\rm conv},j}(\psi_j)\} f(\psi_j, \kappa)\, {\rm d}\psi_j\right]. \tag{8}$$

We would usually need to profile again, depending on what we are interested in, say the center ψ_0 or spread τ for the case of a normal model for the ψ_j. To produce our final confidence curve we will often use the Wilks approximation. This II-CC-FF solution requires the computation of integrals. Sometimes numerical integration routines in R work well enough, other times we will make use of the so-called template model builder package (TMB) and its Laplace approximations in order to compute the integral (Kristensen et al., 2016).

4.3 | Wilks theorems for conversion and fusion

There are chi-squared approximation methods at work at sometimes several levels in our II-CC-FF scheme. For some applications the chi-squared inversion method (6) is crucial, as for the nonparametric CDs for quantiles in Application 8.3, and for other situations what matters more might be the chi-squared approximation of the FF step (7). Limit distribution results securing such χ^2_1 limits are collectively referred to as Wilks theorems, with different setups of regularity conditions. We refrain from setting up lists of precise regularity conditions here, as applications of the theory would involve different types of situations, but we give brief pointers to relevant methods and literature, as follows.

First, regarding (6) and conversion to log-likelihoods, both the exact log-likelihood and the inversion approximation are guaranteed to be close to the negative quadratic $-\frac{1}{2}(\psi_j - \widehat{\psi}_{j,\mathrm{ml}})^2/\widehat{\kappa}_j^2$, for the appropriate $\widehat{\kappa}_j$, by arguments associated with classical large-sample calculus. This would include asymptotic normality of the ML estimator and indeed the traditional Wilks theorem, see Schweder & Hjort (2016, ch. 2 and appendix). The resulting approximations are typically good also when the data information volume is small, as long as the underlying models are smooth in their parameters. Second, the arguments and methods pointed to also entail that the FF inference method (7) is large-sample close to that of minimizing the relevant $\sum_{j=1}^k (\psi_j - \widehat{\psi}_{j,\mathrm{ml}})^2/\widehat{\kappa}_j^2$ under $\phi = \phi(\psi_1, \ldots, \psi_k)$ constraints. For such minimum chi-squared methods, precise Wilks theorems are given in Ferguson (1996, Section 23). Regularity conditions there are of the type where the number of information sources k is kept moderate and fixed, but with steadily more data for each. Importantly, limiting normality of estimators, along with limiting χ^2_1 results for deviances, can also be derived in the rather different setups with small data volume for each source, but where the number k increases. A case in point is where $\phi = b^{\mathrm{t}}\psi$ is linear in the ψ_j, with estimator $\widehat{\phi} = b^{\mathrm{t}}\widehat{\psi}$, and the FF step is large-sample equivalent to $D^*(\phi) = (\phi - \widehat{\phi})^2/\sum_{j=1}^k b_j^2\widehat{\kappa}_j^2$. This tends to a χ^2_1 for increasing k, under mild regularity conditions.

5 | II-CC-FF VERSIONS

The overall objective of the II-CC-FF is to construct a valid confidence curve for each parameter ϕ of particular interest, typically of the form $\phi = \phi(\psi_1, \ldots, \psi_k)$, incorporating the relevant information in all the sources. The framework we have presented so far has not intended to provide a single clear-cut recipe for doing such analyses in practice. Below one such concrete recipe is presented, however, which we call the standard II-CC-FF method, and which may be used for a wide range of models and data. The standard framework has limitations, which we will discuss, and which will serve as a starting point for the presentation of some partial solutions, and more fine-tuned versions of the II-CC-FF scheme.

5.1 | Standard II-CC-FF

The scheme to be described now requires that we have the full data available, or sufficient summaries, from all sources. The statistical work starts by deciding on one or more parameters of particular interest, involving relevant parameters $\psi_1, \ldots, \psi_k$ from the k sources. These might be

parameter vectors (i.e., need not be one-dimensional), they might differ from source to source, but may also contain common parameters across sources.

- ◇ II, *Independent Inspection*: analyze each source j separately. Assume a parametric model for the observations and put up the likelihood function. Profile out the source-specific nuisance parameters, and obtain $\ell_{\text{prof},j}(\psi_j)$.
- ◇ CC, *confidence conversion*: in this case we already have the log-likelihood profiles from each source, so the CC is simple, using the $\ell_{\text{prof},j}(\psi_j)$ directly.
- ◇ FF, *focused fusion*: here we want to obtain a confidence curve for the parameter of overall interest ϕ. Depending on the situation, (i) if ϕ is assumed to be the same across sources or a function of some source-specific parameters, sum the $\ell_{\text{conv},j}$ and then profile again if necessary; (ii) if some component of the ψ_j are assumed to come from some common distribution, use the random effects solution presented above, and then profile again if needed. We then obtain $\ell_{\text{fus,prof}}(\phi)$, and in both cases we use the Wilks approximation to produce the final, combined confidence curve $\text{cc}^*(\phi, \text{data})$.

As for confidence curves in general, we consider the method to work if the final combined confidence curve $\text{cc}^*(\phi)$ has the right coverage properties, either exactly or approximately, as per (3). If the final combined confidence curve does not have the correct coverage properties, this may be due to two related problems: (1) the profiling in either the II or the FF step has gone wrong; and (2) the distribution of the deviance (based on the profile log-likelihood) is far from a χ^2_1, that is, the Wilks approximation is not valid.

Problem (2) is related to the issues discussed in Section 4.3 and will usually disappear when either k or the n_j increase. In situations with little data, the Wilks approximation can sometimes be ameliorated using relatively simple tools, like the Bartlett correction; see, for instance, Schweder & Hjort (2016, chs. 7, 8) for a discussion on such fine-tuning methods and second-order approximations. Furthermore, in some situations one may be able to derive and simulate the distribution of the deviance exactly, and thus bypass the use of the Wilks approximation altogether. We will see examples of such II-CC-FF versions in Sections 6.1 and 8.1.

Problem (1) is related to the profiling and the presence of nuisance parameters. In situations with nuisance parameters using the profile log-likelihood can lead to "inefficient and even inconsistent estimates" (McCullagh & Tibshirani, 1990). As we see above, the standard II-CC-FF method may often require two rounds of profiling: first in the II step where we might profile out the *source-specific* nuisance parameters, and sometimes in the FF step where we might profile out *shared* nuisance parameters (which are shared by the k sources). If we have "large sources," that is, the sample size n_j of each source is large, we can safely profile in the II step. If some or all the sources are small, however, one should be more careful. Specifically, the profiling might go wrong and we illustrate this situation with a famous example in Appendix C, the Neyman–Scott problem.

Shared nuisance parameters can be of different kinds, and here we will particularly concern ourselves with nuisance parameters arising from the random effect distribution in the FF step. For example, if we have $\psi_j \sim \text{N}(\psi_0, \tau^2)$ and our focus parameter is ψ_0, then τ is a shared nuisance parameter of that type. If the number of sources k is large we can safely profile, while if k is small we may need to resort to some of the corrections described next. Often we have both source-specific and shared nuisance parameters. In that case, we ideally need to have a large number of large sources to produce valid CDs with the default profiling-based method. Note that in these cases, if k is too small, large sources will not necessarily help. Conversely, if the n_j are

too small, a large k will not in general be able to remedy the mistakes coming from profiling in the II step.

5.2 | Corrections to the log-likelihood profile

There is a large literature concerning corrections or modifications of the profile likelihood. The different corrections appearing in the literature have varying performance and complexity; see, for instance, Barndorff-Nielsen (1986), Cox and Reid (1993), Diciccio and Efron (1992), Stern (1997), DiCiccio et al. (1996). There is also a whole subfield of integrated likelihood methods with partly similar aims, see Berger et al. (1999). A thorough investigation of all these methods is outside the scope of this article, and we will therefore only present one rather simple, somewhat limited solution. Alternative methods might work better, or at least in a more general setting, but these are often more complicated to compute.

In Cox and Reid (1987), the authors present what we will term the simple Cox–Reid correction. This is possibly the easiest correction to compute among those mentioned above. It can be considered a special case of the correction in the general modified profile likelihood of Barndorff-Nielsen, but the simple Cox–Reid correction is limited to situations with orthogonal parameters (i.e., that the off-diagonal terms in the expected information matrix are equal to zero). Assume we have a scalar parameter of interest ψ and some vector of nuisance parameters γ. As usual, the profile log-likelihood for ψ is defined as $\ell_{\text{prof}}(\psi) = \ell(\psi, \widehat{\gamma}(\psi))$, where $\widehat{\gamma}(\psi)$ is the ML estimate of γ for each fixed ψ value. The simple Cox–Reid correction gives the following modification of the profile log-likelihood,

$$\ell_{\text{cprof}}(\psi) = \ell_{\text{prof}}(\psi) - \frac{1}{2}\log[\det J_{\gamma\gamma}(\psi, \widehat{\gamma}(\psi))] \tag{9}$$

where $J_{\gamma\gamma}(\psi, \widehat{\gamma}(\psi)) = -\partial^2\ell(\psi, \gamma)/(\partial\gamma\partial\gamma^{\mathrm{t}})$ is the observed information for the γ components, evaluated at $(\psi, \widehat{\gamma}(\psi))$. The simple Cox–Reid correction can be used both in the II and FF steps, for models with orthogonal parameters. We illustrate the use of the Cox–Reid correction in the II step in the Appendix C. In the FF step, corrections may be necessary when there are shared nuisance parameters arising from a random effect distribution. In particular, we propose that this correction should readily be applied when then random effect distribution is assumed to be normal. Here, the correction should be particularly notable for small k. We will present some Cox–Reid corrections in a classic model for random effect meta-analysis in Section 6.

5.3 | Optimal CD methods

For some parameters in exponential families, we can bypass the standard II-CC-FF, and its potential problems, by making use of the following alternative method which is much more powerful, producing *optimal CDs*. In our II-CC-FF setting, this method might come into play both in the II step and the FF step, see the application in Section 6.2.

For ease of presentation, we present the optimal confidence method in the case where all the k sources inform on a common focus parameter $\psi = \psi_1 = \ldots = \psi_k$. This constitutes a situation where the method is used in the final FF step. Suppose again that ψ is the focus parameter, and that we have m nuisance parameters $\gamma_1, \ldots, \gamma_m$, which may be both source-specific or shared

across all k sources. Suppose also that the log-likelihood function at work, based on information sources $y_1, \ldots, y_k$, can be written in the form

$$\ell(\psi, \gamma_1, \ldots, \gamma_m) = \psi A + \gamma_1 B_1 + \ldots + \gamma_m B_m - d(\psi, \gamma_1, \ldots, \gamma_m) + h(y_1, \ldots, y_k), \tag{10}$$

where A and $B_1, \ldots, B_m$ are statistics, that is, functions of the data collection, with observed values $A_{\rm obs}$ and $B_{1,\rm obs}, \ldots, B_{m,\rm obs}$, and with m often bigger than k. Then, under mild regularity conditions, there is an overall most powerful CD, namely

$$C^*(\psi, y) = P_\psi\{A \geq A_{\rm obs} \mid B_1 = B_{1,\rm obs}, \ldots, B_m = B_{m,\rm obs}\}.$$

That this $C^*(\psi, y)$ indeed depends on ψ but not on the γ_j parameters is part of the result and the construction.

To illuminate the exact meaning of "most powerful" in this setting, one needs to consider the theory for loss and risk functions for CDs developed in Schweder & Hjort (2016, ch. 5). Confidence power is measured via the risk function

$$r(C, \psi, \gamma) = \mathrm{E}_{\psi,\gamma} \int L(\psi_{\rm cd} - \psi)\, \mathrm{d}C(\psi_{\rm cd}, Y), \tag{11}$$

for any convex nonnegative $L(\cdot)$ with $L(0) = 0$. The random mechanism involved in the expectation here is a two-stage operation; first data y, governed by the (ψ, γ) held fixed, are used to generate the CD $C(\psi, y)$, and then $\psi_{\rm cd}$ is a random draw from this distribution. Intuitively, a low confidence risk means that the CD in question is tight around true value of ψ, while CDs which are less concentrated around the true value will have a higher risk. A CD with low confidence risk should therefore be expected to produce narrow confidence intervals (but keeping the correct coverage), and point estimates with little bias.

6 | META-ANALYSIS

As mentioned in the small start example (1), some common meta-analysis methods flow more or less directly from the II-CC-FF framework. In addition, the framework also invites more general, principled, and nonstandard solutions some which we will explore in this section. Furthermore, we will investigate connections between II-CC-FF and a couple of widely encountered meta-analysis methods. We will start with a discussion of the basic random effect model, before we go on to the famous case of meta-analysis of 2×2 tables.

6.1 | The basic random effect model

The most canonical type of random effect meta-analysis, which we term the basic random effect model, starts with k independent estimators $y_1, \ldots, y_k$ aiming at the parameters $\psi_1, \ldots, \psi_k$, with $y_j \mid \psi_j \sim \mathrm{N}(\psi_j, \sigma_j^2)$, and $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. Usually the source-specific SDs σ_j are assumed known. The literature treating this model is enormous, see, for instance, Langan et al. (2019) and Partlett and Riley (2017) and references therein. Note that likelihood-based method for the basic random effect model, even exploring higher order corrections, have been investigated earlier, see, for

instance, Hardy and Thompson (1996) and Noma (2011). For a more general likelihood approach see O'Rourke (2008).

When assuming known σ_j, the integral in (8) has an explicit solution and the standard II-CC-FF solution will rely on the following log-likelihood function

$$\ell_{\text{fus}}(\psi_0,\tau)=\sum_{j=1}^{k}\left\{-\frac{1}{2}\log(\sigma_j^2+\tau^2)-\frac{1}{2}\frac{(y_j-\psi_0)^2}{\sigma_j^2+\tau^2}\right\},$$

where we profile out either ψ_0 or τ depending on which parameter is of main interest. The confidence curves for ψ_0 and τ will point at the standard ML estimators, and the ensuing confidence intervals will be very similar to solutions which have been investigated in some of the references mentioned above. These solution are reasonably good when k is not too small, see also Section 7.1, but may be improved upon. Langan et al. (2019) find, for instance, that the ML estimator for τ has relatively poor performance in terms of bias and mean squared error.

We can attempt to improve on the standard II-CC-FF solution using the Cox–Reid correction. First we will consider the case where ψ_0 is the parameter of main interest, then the full combined profile likelihood from the FF step becomes

$$\begin{aligned}\ell_{\text{fus,cprof}}(\psi_0)=&\sum_{j=1}^{k}\left\{-\frac{1}{2}\log(\sigma_j^2+\hat\tau^2(\psi_0))-\frac{1}{2}\frac{(y_j-\psi_0)^2}{\sigma_j^2+\hat\tau^2(\psi_0)}\right\}\\&-\frac{1}{2}\log\sum_{j=1}^{k}\left\{-\frac{1}{2}\frac{1}{(\sigma_j^2+\hat\tau^2(\psi_0))^2}+\frac{(y_j-\psi_0)^2}{(\sigma_j^2+\hat\tau^2(\psi_0))^3}\right\}.\end{aligned}\tag{12}$$

The first part of the formula is the ordinary profile log-likelihood, the second part the simple Cox–Reid correction. We obtain the combined confidence curve using the Wilks approximation. We have not seen this solution anywhere in the literature, and we investigate its performance in Section 7.1.

We can consider a variation of (12) in the case of the individual sources having few measurements, which means that the σ_j estimates become uncertain. The II-CC-FF can then provide more sophisticated solutions. In the II step, we have exact CDs for each ψ_j based on the Student's t distribution, which we can convert to a confidence log-likelihood by exact conversion in the CC step. For the FF step, we use the general random effect method from (8), either with numerical integration or using the TMB package. Corrections in both the II and FF step may be considered, but we have not fully investigated these options yet.

Rather than focusing on the overall mean, as we did above, one may be interested in the overall spread τ. From the above, we have direct and Cox–Reid corrected log-likelihood profiles $\ell_{\text{fus,prof}}(\tau)=-\frac{1}{2}A_k(\tau)$ and $\ell_{\text{fus,cprof}}(\tau)=-\frac{1}{2}B_k(\tau)$, with

$$A_k(\tau)=\sum_{j=1}^{k}\left[\log(\sigma_j^2+\tau^2)+\frac{\{y_j-\hat\psi_0(\tau)\}^2}{\sigma_j^2+\tau^2}\right]\quad\text{and}\quad B_k(\tau)=A_k(\tau)+\log\left(\sum_{j=1}^{k}\frac{1}{\sigma_j^2+\tau^2}\right),$$

with the profiled ML estimator $\hat\psi_0(\tau)=\sum_{j=1}^{k}y_j/(\sigma_j^2+\tau^2)/\sum_{j=1}^{k}1/(\sigma_j^2+\tau^2)$ for each given τ. One might recognize the $-\frac{1}{2}B_k(\tau)$ as the log-likelihood associated with the restricted maximum likelihood (REML) estimator for τ, see, for instance, Langan et al. (2019). The link between the

Cox–Reid correction (and more general corrections) and the REML procedure has been known for some time, see Durban and Currie (2000) and also Cox and Reid (1987), but is possibly underappreciated in the meta-analysis literature. The above leads to two deviance functions, using the direct and the corrected log-likelihood profiles,

$$D_{k,\mathrm{ml}}(\tau) = A_k(\tau) - A_k(\widehat{\tau}_{\mathrm{ml}}) \quad \text{and} \quad D_{k,\mathrm{cml}}(\tau) = B_k(\tau) - B_k(\widehat{\tau}_{\mathrm{cml}}),$$

with $\widehat{\tau}_{\mathrm{ml}}$ and $\widehat{\tau}_{\mathrm{cml}}$ the minimizers of $A_k(\tau)$ and $B_k(\tau)$. The distribution of $y_j - \widehat{\psi}_0(\tau)$ does not depend on the underlying ψ_0, which means that $D_{k,\mathrm{ml}}(\tau)$ and $D_{k,\mathrm{cml}}(\tau)$ have distributions only depending on the candidate value τ. Thus, we have well-defined and exact confidence curves for τ,

$$\mathrm{cc}_{\mathrm{ml}}(\tau) = P_\tau\{D_{k,\mathrm{ml}}(\tau) \le D_{k,\mathrm{ml,obs}}(\tau)\} \quad \text{and} \quad \mathrm{cc}_{\mathrm{cml}}(\tau) = P_\tau\{D_{k,\mathrm{cml}}(\tau) \le D_{k,\mathrm{cml,obs}}(\tau)\}. \tag{13}$$

We make use of these confidence curve methods in Application 8.1. The Wilks type chi-squared approximation works well for moderate to large values of k, but not for τ small, which often might be the parameter region of primary interest, as for the application pointed to. Hence we need to compute the two confidence curves via simulations, with a high number of $D_{k,\mathrm{ml}}(\tau)$ and $D_{k,\mathrm{cml}}(\tau)$ generated for each candidate value τ.

6.2 | Meta-analyses of 2×2 tables

In meta-analyses of 2×2 tables, each study seeks to compare the probability of observing a certain binary event in the control group and in the treatment group. The counting variables $Y_{0,j}$ and $Y_{1,j}$ indicate how many patients have experienced the event in each group in study j. These variables are usually modeled as pairs of binomials, $Y_{0,j} \sim \mathrm{binom}(m_{0,j}, p_{0,j})$ and $Y_{1,j} \sim \mathrm{binom}(m_{1,j}, p_{1,j})$, with subscript "1" indicating treatment and "0" control, and the sample sizes in each group denoted by $m_{0,j}$ and $m_{1,j}$. The most common measure for the treatment effect is the odds ratio, or equivalently, the log odds ratio ψ_j. For that effect measure it is convenient to express the event probabilities in the control and treatment groups as $p_{0,j} = \exp(\theta_j)/(1+\exp(\theta_j))$ and $p_{1,j} = \exp(\theta_j+\psi_j)/(1+\exp(\theta_j+\psi_j))$. Each study, or source, has a specific nuisance parameter θ_j, governing the event probability in the control group. We will first treat the fixed effect case where the log odds ratios are assumed common across all sources, $\psi_1 = \cdots = \psi_k = \psi$, before we come to the random effect case in the next paragraph. The information available in each source depends on the sample sizes $m_{0,j}$ and $m_{1,j}$, and on the event probabilities. If the number of studies increases while the size of each study stays constant, it is known that the ML estimator is inconsistent (Breslow, 1981), and we can expect that the standard II-CC-FF method will not work well. In addition, the simple Cox–Reid correction to the profile in each source is not immediately available because ψ_j and θ_j are not orthogonal. However, there exists an optimal CD for the common ψ based on the theory from Section 5.3,

$$C_{\mathrm{opt}}(\psi, \mathrm{data}) = P_\psi(B_k > b \mid z_1, \ldots, z_k) + \frac{1}{2}P_\psi(B_k = b \mid z_1, \ldots, z_k). \tag{14}$$

Here, $z_j = y_{0,j} + y_{1,j}$ and $B_k = \sum_{j=1}^k Y_{1,j}$. The CD is obtained by simulating the distribution of B_k given $z_1, \ldots, z_k$. The second part of (14) is a half-correction. Note also that we similarly have an optimal CD for ψ_j within each source,

$$C_{\text{opt},j}(\psi_j, y_{0,j}, y_{1,j}) = P_\psi(Y_{1,j} > y_{1,j}|z_j) + \frac{1}{2}P_\psi(Y_{1,j} = y_{1,j}|z_j). \tag{15}$$

This CD is simple to compute as $Y_{1,j}|Z_j$ has an eccentric hypergeometric distribution. Note that this CD is closely related to the method known as Fisher's exact test, see Lehmann & Romano (2005, ch. 4). Starting from (15) for each source in the II step, we can obtain an approximation to the optimal solution in (14) which is faster to compute and also lends itself to a natural random effect extension, as we will see. In the CC step, we use exact conversion to obtain the confidence log-likelihoods $\ell_{\text{conv},j}(\psi_j) = \log g_j(y_{1,j}, \psi_j)$, where

$$g_j(y_{1,j}, \psi_j) = \frac{\binom{m_{0,j}}{z_j - y_{1,j}}\binom{m_{1,j}}{y_{1,j}}\exp(\psi_j y_{1,j})}{\sum_{u=0}^{z_j}\binom{m_{0,j}}{z_j - u}\binom{m_{1,j}}{u}\exp(\psi u)} \quad \text{for } y_{1,j} = 0, 1, \ldots, \min(z_j, m_{1,j}) \tag{16}$$

is the probability point function of the eccentric hypergeometric distribution. We sum these confidence log-likelihoods to get $\ell_{\text{fus}}(\psi) = \sum_{j=1}^{k}\ell_{\text{conv},j}(\psi_j)$, find the ML estimate $\widehat{\psi}$ and the deviance, and use the Wilks approximation:

$$\text{cc}^*(\psi, \text{data}) = \Gamma_1(2\{\ell_{\text{fus}}(\widehat{\psi}) - \ell_{\text{fus}}(\psi)\}). \tag{17}$$

Even though there is some level of approximation in this solution, it tends to work well, see Section 7.2.

From this approximate fixed effect approach, we find a natural extension to random effects. Assume that the log-odd ratios from the different sources come from a common normal distribution, $\psi_j \sim \text{N}(\psi_0, \tau^2)$. Then we have the following fusion log-likelihood for the overall parameters,

$$\ell_{\text{fus}}(\psi_0, \tau) = \sum_{j=1}^{k}\log\left\{\int g_j(y_{1,j}, \psi_j)\frac{1}{\tau}\phi\left(\frac{\psi_j - \psi_0}{\tau}\right)\,\text{d}\psi_j\right\}, \tag{18}$$

where $g_j(y_{1,j}, \psi_j)$ is the density function of the eccentric hypergeometric distribution, pointed to above. If we are mainly interested in ψ_0, we profile out τ. Then, we use the Wilks approximation to obtain the combined $\text{cc}(\psi_0, \text{data})$. If the heterogeneity is big or the number of sources is small we add an approximate Cox–Reid correction $\log\{\widehat{\tau}(\psi_0)\}$ to the profile log-likelihood. This correction comes from the normal random effect model and is approximate in the sense that the orthogonality of ψ_0 and τ in the full (integrated) distribution does not necessarily hold. When the within-study sample sizes are sufficiently large we can expect this approximation to be good. In applications, we computed the integral in (18) using the TMB package. This approach seems promising with good coverage properties in simulations (Section 7.2).

7 | PERFORMANCE EVALUATIONS

The main benefit of our II-CC-FF framework is its general nature and wide applicability, and in Section 8we will therefore demonstrate the use of II-CC-FF in several nonstandard combination situations. Still, we also require that methods coming out of II-CC-FF should be competitive

against other methods from the literature when applied to typical combination situations, for example, meta-analysis settings. Here, we study the performance of II-CC-FF methods in two very classical situations: first in the basic random effect model and then in the meta-analysis of 2×2 tables. Both of these types of meta-analyses were discussed in Section 6, along with some notation which is used in this section too.

7.1 | The basic random effect model

We investigate some of the methods from Section 6.1 with a simulation study inspired by Langan et al. (2019). We treat the setting where the overall mean, ψ_0, is the parameter of main interest. In that setting Langan et al. (2019) found that confidence intervals computed by the Hartung–Knapp–Sidik–Jonkman (HKSJ) method were the clear winners in terms of coverage properties. That method uses the traditional inverse-variance method for the point estimator, and computes confidence intervals based on the t-distribution and the following variance formula,

$$\mathrm{Var}_{\mathrm{HKSJ}}\,\widehat{\psi}_0 = \frac{\sum_{j=1}^k (\widehat{\psi}_j - \widehat{\psi}_0)^2/(\widehat{\sigma}_j^2 + \widehat{\tau}^2)}{(k-1)\sum_{j=1}^k 1/(\widehat{\sigma}_j^2 + \widehat{\tau}^2)}.$$

This formula requires one to plug-in an estimator for τ and for this we used the REML estimator, as recommended by Langan et al. (2019). We compare this state-of-the-art method with our standard II-CC-FF method, and the II-CC-FF method using the corrected log-likelihood profile in (12). We also include the method for the basic random effect model which comes out of the CD combination framework of Xie et al. (2011). This method is implemented in the `gmeta` package (Yang et al., 2016). See Appendix A for more details.

We let $\psi_0 = 0.5$ and consider two values for the between-study heterogeneity, $\tau = 0.09$ for a scenario with little heterogeneity and $\tau = 0.44$ for a scenario with large heterogeneity. Furthermore, we choose $\sigma_j = 2/\sqrt{m_j}$ with $m_j \sim \mathrm{unif}(30, 50)$, and investigate four values for k the number of studies, 5, 10, 20, 50. We generate 10,000 meta-analyses for each k value. In each of the k studies ψ_j and σ_j are estimated using their ordinary estimators. For each method, we record the coverage rate of 95% confidence intervals, the width of these intervals, and point estimates for ψ_0.

The confidence intervals from the HKSJ method have perfect coverage rate for almost all k values in both scenarios (Figure 2), which is consistent with results in Langan et al. (2019). The `gmeta` method has very good performance in the scenario with small heterogeneity, but severe undercoverage in the scenario with large heterogeneity. The standard II-CC-FF method has a similar problem in that scenario, but the II-CC-FF version with corrected profile likelihood obtains coverage rates close to the nominal 0.95. Both II-CC-FF versions are slightly conservative in the small heterogeneity scenario, but obtain nonetheless confidence intervals of similar mean width as the HKSJ intervals. The four methods produce virtually identical point estimates for ψ_0. All in all we find the II-CC-FF method to have almost equally good performance as the state-of-the-art method. In scenarios with low heterogeneity, the correction to the profile likelihood is not important, while for scenarios with large between-study heterogeneity, it has a noticeable effect on the coverage properties of the II-CC-FF method.

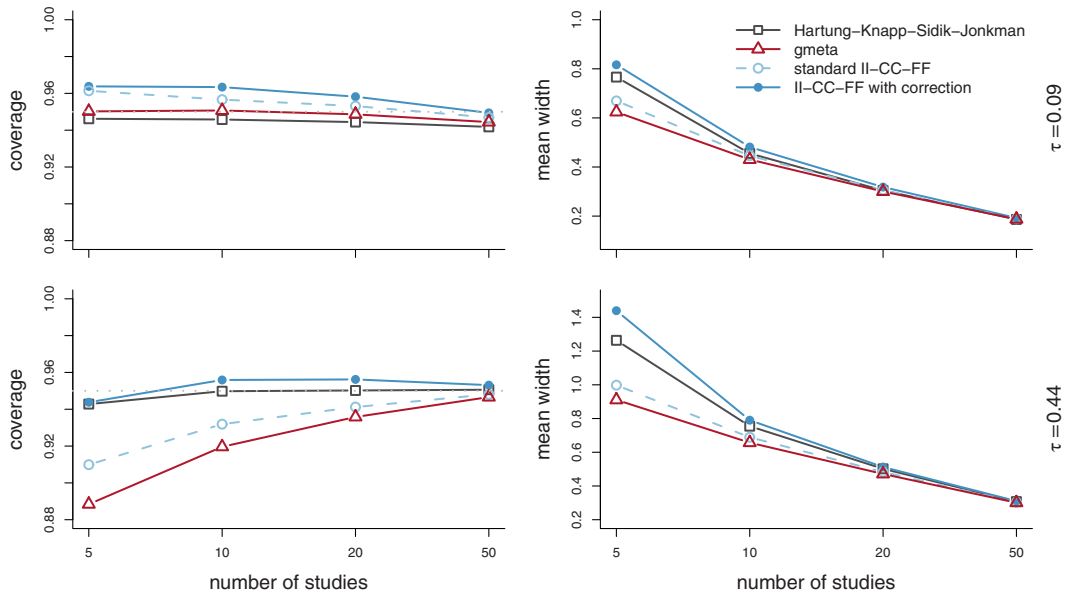


FIGURE 2 Simulation results for the basic random effect model. The left column gives the realized coverage rate of 95% confidence intervals, the right column gives the mean width of these intervals. The results in the top row belong to a scenario with small between-study heterogeneity, while in the bottom row the between-study heterogeneity is large [Color figure can be viewed at wileyonlinelibrary.com]

7.2 | Meta-analyses of 2×2 tables

We investigate both the fixed effect and random effect cases. Our simulation setups are inspired by two recent papers with extensive simulation studies: Piaget-Rossel and Taffé (2019) for the fixed effect case and Jackson et al. (2018) for the random effects case. With fixed effects, we investigate three common effect measures: the (log) odds ratio defined in Section 6, the (log) risk ratio, $\exp(\psi_j) = p_{1,j}/p_{0,j}$, and the risk difference, $\psi_j = p_{1,j} - p_{0,j}$. With random effects, we only treat the odds ratio, however.

Piaget-Rossel and Taffé (2019) focus on meta-analyses with rare events, that is, where both the treatment and control group have low event probabilities and where there may be many studies with zero events. The overall conclusion of the paper was that the Mantel–Haenszel method had the best performance among the methods that were considered. This method is applicable to all three effect measures mentioned above, and Piaget-Rossel and Taffé (2019) found that it produced confidence intervals with good coverage properties, and point estimates with relatively small bias. The Mantel–Haenszel is a well-established method, which was originally proposed in 1959 (Mantel and Haenszel (1959) and extended in Rothman et al. (2008)). The method offers explicit estimators for the log odds ratio, log risk ratio, and risk difference, as well as expressions for the variance of these estimators. Confidence intervals are computed using the Wald approximation. We will compare some variants of our II-CC-FF scheme with the Mantel–Haenszel method. We will also include methods reviewed in Liu et al. (2014). These methods are based on exact tests and fall into the unifying CD framework of Singh et al. (2005) which we describe in Appendix D. For the odds ratio and risk difference, these CD methods have been implemented in the `gmeta` package (Yang et al., 2016); for the risk ratio a related exact method is implemented

in the `exactmeta` package (Yu & Tian, 2014). See Appendix A for more details on the simulations.

For all three effect measures, we use the standard II-CC-FF procedure which consists of profiling out the nuisance parameters in each source, summing the log-likelihood profiles, and then using the Wilks approximation to obtain a confidence curve for ψ, from which we can extract confidence intervals and point estimates. For the odds ratio case we will include two additional variants that can be said to fall under the II-CC-FF umbrella: the optimal CD method given in (14) and the II-CC-FF method using exact conversion, which we give in (17). This II-CC-FF version resembles the standard II-CC-FF procedure, but avoids the profiling step by making use of the conditional log-likelihood for ψ_j, in (16).

We simulate datasets with a median event probability of 0.005 in the control group. According to the simulation study in Piaget-Rossel and Taffé (2019) we can expect that this will be a challenging setting for all methods. We generate baseline probabilities $p_{0,j} = \exp(\theta_j)/\{1+\exp(\theta_j)\}$ by drawing $\theta_j \sim \mathrm{N}(\log \frac{0.005}{1-0.005}, 0.5)$, $m_{1,j} \sim \mathrm{unif}(50, 150)$ and $m_{0,j} = r_j m_{1,j}$ with $r_j \sim \mathrm{unif}(0.5, 1.5)$. The probabilities in the treatment group are computed according to the chosen effect measure,

$$\begin{aligned}
&\text{Odds ratio:} \quad p_{1,j} = \frac{\exp(\theta_j+\psi)}{1+\exp(\theta_j+\psi)} \quad \text{with the log odds ratio } \psi = -1.5,\\
&\text{Risk ratio:} \quad p_{1,j} = \exp(\psi)p_{0,j} \quad \text{with the log risk ratio } \psi = -1.5,\\
&\text{Risk difference:} \quad p_{1,j} = \psi + p_{0,j} \quad \text{with the risk difference } \psi = 0.05.
\end{aligned}$$

We let k, the number of studies, take the values 5, 10, 20, and 50. For each setup we generate 10,000 meta-analyses, except in the risk difference setup where we only generate 1000 (because the `gmeta` package was extremely slow for this effect measure). For each method, we compute the coverage rate of 95% confidence intervals, the median width of these intervals, and the median bias of the point estimates coming out of the methods. Both the odds ratio and risk ratio were analyzed on the log scale.

Results are presented in Figure 3. When there are few studies, most method struggle with overcoverage in the odds ratio and risk ratio setups, but the Mantel–Haenszel method and the various II-CC-FF schemes come close to the nominal level as k increases. For the odds ratio, the standard II-CC-FF and the II-CC-FF with exact conversion have practically identical performance: both have some degree of undercoverage for $k = 20$, but come very close to the nominal 0.95 for $k = 50$. This seems to indicate that even though we are in a rare events setting, there is nonetheless sufficient information within each source so as to safely profile out the θ_j. The undercoverage experienced by these methods for $k = 20$ indicates that the Wilks approximation is not perfectly fine, and one could consider adjustments, like for instance, the Bartlett correction. The standard II-CC-FF has reasonably good coverage properties for the risk ratio and risk difference. For the risk ratio, there is again some degree of undercoverage for $k = 20$, but for the risk difference there is no such pattern. In that setting, the Mantel–Haenszel method and the standard II-CC-FF method have practically identical performance. The `gmeta`and `exactmeta` methods are very conservative in all these experiments and consistently produce very wide 95% intervals, with a coverage rate close to 1. When there are few events, all methods tend to underestimate the effect measure (giving a negative median bias). As k increases all methods come closer to the true effect, except for the `gmeta` method for the risk difference which appears to get worse. For the odds ratio and risk ratio, however, the `gmeta` and `exactmeta` methods

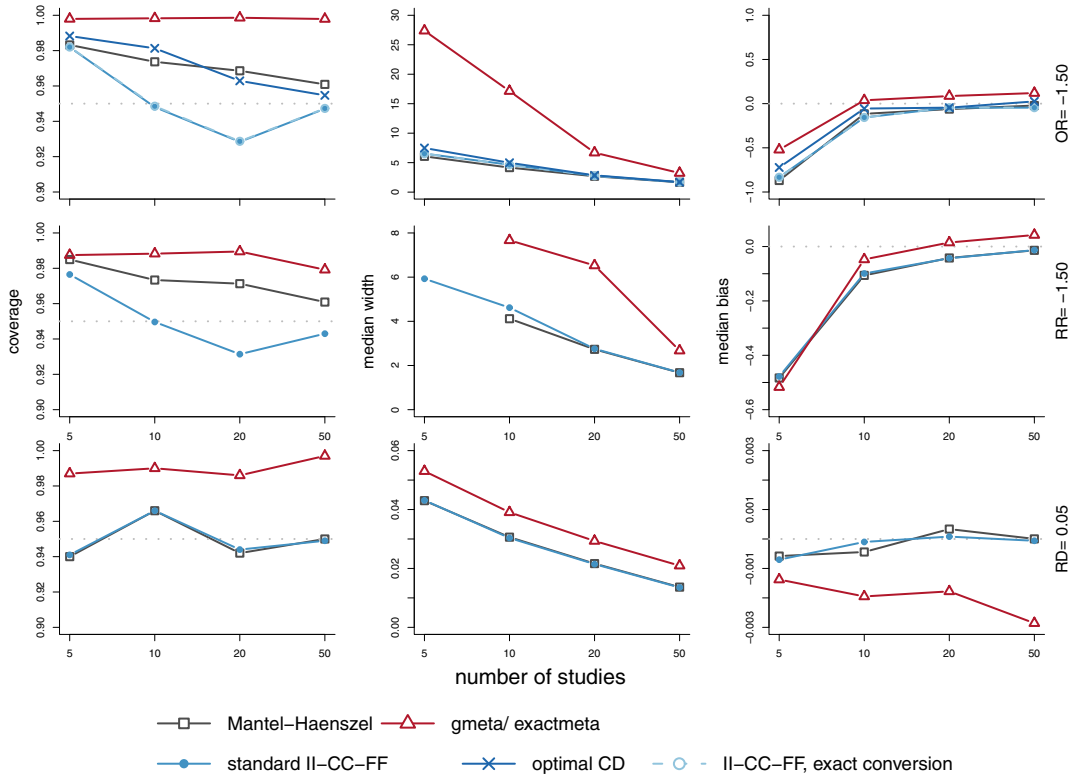


FIGURE 3 Simulation results for fixed-effect meta-analysis of 2×2 tables. The left column gives the realized coverage rate of 95% confidence intervals, the middle column gives the median width of these intervals, and the right column gives the median bias of the point estimate coming from each of the methods. The top row gives the results for the (log) odds ratio, the middle row for the (log) risk ratio, and the bottom row gives the results for the risk difference. When k is small the confidence intervals can sometimes have infinite width (which explains the missing points for the log risk ratio) [Color figure can be viewed at wileyonlinelibrary.com]

produce point estimates with somewhat smaller median bias than the other methods for some values of k.

Overall, II-CC-FF performs similarly, and at best, slightly better than the main existing competitor (according to Piaget-Rossel and Taffé (2019)). The II-CC-FF and the Mantel–Haenszel methods have generally similarly wide confidence intervals, but II-CC-FF intervals have a coverage rate sometimes coming closer to the nominal level. In some settings, the standard II-CC-FF method produces intervals with some degree of undercoverage, but the optimal CD method (which is only available for the odds ratio) does not. When considering coverage properties the `gmeta` and `exactmeta` methods perform poorly in the settings we have presented here. We have included a simulation study with less extreme event probabilities in the appendix. There the median event probability in the control group is 0.1. That situation is less challenging for all the methods, and most obtain a coverage rate close to 0.95 even for small k. The `gmeta` method has acceptable performance in that setting, at least for the odds ratio. For the risk difference, the results are still quite poor.

Some readers might be puzzled by the fact that the optimal CD method for the odds ratio does not have even better performance in the simulations. This CD is optimal in the sense of Section 5.3

and one might think that it should have exact coverage properties. For discrete data as we have here, the method of Section 5.3 requires a half-correction, as we see in (14). In settings with a few tables and few events, the statistics encountered will attain only a few possible values, and so have particularly noncontinuous distributions; this likely explains the results we see in these simulations.

Finally, note that in rare events settings it might be fruitful to assume that the event counts in the two groups are Poisson distributed rather than binomials, see, for instance, Cai et al. (2010) and Cunen and Hjort (2015). Our II-CC-FF framework can naturally be applied to such models too and would produce different results than the ones we see here.

In Jackson et al. (2018) the authors compare seven methods for odds ratio meta-analysis of 2×2 tables in a random effects setting. Many authors find random effects methods, and implicitly random effects models, to be preferable to fixed effects methods for this type of meta-analysis, since it may be more realistic to allow for some heterogeneity in the treatment effects between the studies. As is commonly done, we assume that the log odds ratios come from a normal distribution, $\psi_j \sim \mathrm{N}(\psi_0, \tau^2)$. In their simulation study, Jackson et al. (2018) found that several of the methods had similar (equally good) performance. Among others, they recommend the use of a modified version of the Simmonds and Higgins (2016) method, for its ease of use and good performance. The method uses a generalized linear mixed effect framework to estimate ψ_0 and τ (and the θ_j) by assuming that the logit of $(p_{0,j}, p_{1,j})$ are drawn from a bivariate normal distribution with expectation $(\theta_j, \theta_j + \psi_0)$ and a certain covariance matrix (with variances equal to $\tau^2/4$). The modified Simmonds-Higgins method is implemented in the `metafor` package (Viechtbauer, 2010), and we will compare our II-CC-FF methods with this. The `gmeta` and `exactmeta` do not provide a readily applicable method for these type of random effects situations.

We will investigate two II-CC-FF versions here. The first version constitutes the standard II-CC-FF solution for random effects. We use (18), profile out τ and use the Wilks approximation to obtain the combined confidence curve for ψ_0. The second version makes use of the same profile log-likelihood, but adds the approximate Cox–Reid correction as suggested in the text under (18). The idea is that this correction will account for the error from having profiled out τ. We compute the integral in (18) using the TMB package. Note that we do not compute the correction in rounds where τ is estimated to be very close to zero (below 0.0001), since in that case the correction blows up, as is apparent from its form, $\log\{\widehat{\tau}(\psi_0)\}$.

Jackson et al. (2018) include many different scenarios in their simulations, but we limit ourselves to one main scenario, where we let the number of studies, k, vary as we did in the other simulations. We use $\theta_j \sim \mathrm{N}(\log \frac{0.2}{1-0.2}, 0.3^2)$ giving median baseline probabilities of 0.2, $m_{1,j} \sim \mathrm{unif}(10, 50)$ and $m_{0,j} = m_{1,j}$. Furthermore, we let $\psi_j \sim \mathrm{N}(0, 0.168)$, which gives considerable heterogeneity in the treatment effects. We generate 10,000 meta-analyses for each value of k (5, 10, 20 and 50).

For each method, we compute the coverage rate of 95% confidence intervals, the median width of these intervals, and we recorded the point estimates coming out of the methods. In Figure 4, we display the coverage rate and median width results. Both the modified Simmonds and Higgins and standard II-CC-FF methods produce confidence intervals with some degree of undercoverage. For the modified Simmonds and Higgins method, this is consistent with the results in Jackson et al. (2018) for scenarios with considerable heterogeneity in the treatment effects and small within-study sample sizes (as we have here). Surprisingly, the coverage rate of the intervals from the standard II-CC-FF seems to worsen as k increases. The performance of the corrected II-CC-FF method in terms of coverage rate is good.

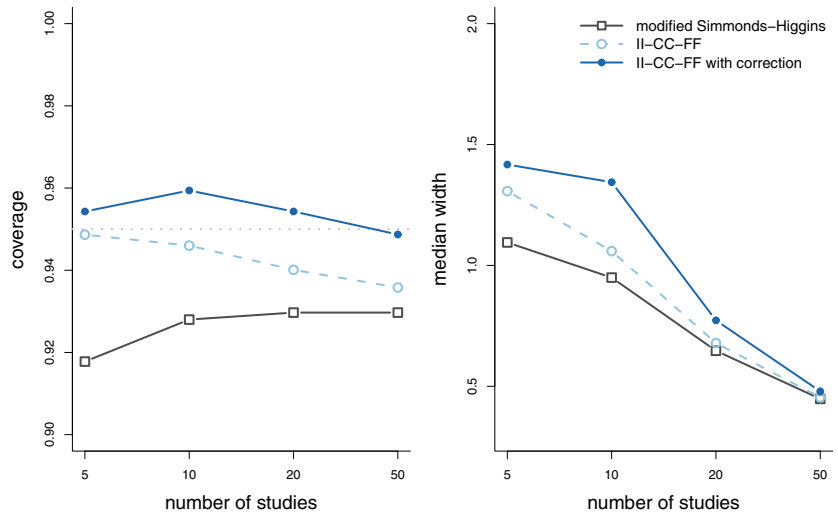


FIGURE 4 Simulation results for random effect meta-analysis of 2×2 tables. The left plot gives the realized coverage rate of 95% confidence intervals, the right plot gives the median width of these intervals [Color figure can be viewed at wileyonlinelibrary.com]

The three methods produce very similar estimates of ψ_0 and τ. The bias for ψ_0 is small, but all three methods tend to underestimate τ, which is a feature they share with all the methods investigated in Jackson et al. (2018). Remember that the II-CC-FF methods are in this case focusing on ψ_0, and if τ had been of primary interest we would have used the confidence curve given in (13).

In Appendix A we include a similar figure for a scenario with less heterogeneity and higher within-study sample sizes. In that case, the modified Simmonds-Higgins method has close to the correct coverage rate for all values of k, while both II-CC-FF methods are slightly conservative. When τ is small the correction term seldom changes the confidence curve, and the two II-CC-FF confidence curves are therefore often very similar in that scenario.

Again, we find II-CC-FF to have an overall good performance. In situations with large heterogeneity, the corrected II-CC-FF method outperforms the recommended method from the literature in terms of coverage properties. The scenario we have investigated here has quite large event probabilities, and it is conceivable that the results would be somewhat different in a rare event setting. We hope to investigate this issue further in a separate paper.

Our standard II-CC-FF solution has actually been suggested previously, in Stijnen et al. (2010), and also in Van Houwelingen et al. (1993), as the hypergeometric-normal model. This model is in fact among the seven studied in Jackson et al. (2018), and while it obtains reasonably good performance in their simulations, the authors report numerical problems and estimation failure. This could be related to a different implementation than ours and particularly to the Laplace approximations used in the TMB package. We did not find that the standard II-CC-FF had a high probability of failure in the scenarios we investigated.

8 | APPLICATIONS

Below we illustrate the capacity for the II-CC-FF paradigm to solve problems in four rather different application settings. The first application concerns an interesting archaeological dataset. Here,

we use the so-called basic random effect model which was discussed in Sections 6.1 and 7.1, but perhaps atypically our parameter of main interest is the spread parameter, not the overall mean parameter. The annual growth rate of humpback whales is the focus of our second application story. There we illustrate how to construct confidence curves based on nonsufficient summary statistics; we only have access to a point estimate and a highly nonsymmetric confidence interval. In this example, we also demonstrate how partial prior information can be incorporated into our II-CC-FF framework. Our third story concerns the development over time of the median Body Mass Index for Olympic speedskaters, where part of the challenge is to construct and then convert accurate nonparametric CDs for sample medians to parametric log-likelihood terms. Finally, we illustrate the combination of "hard" and "soft" data with a grand question from the field of peace and conflict research; is there evidence for The long peace, and in that case, when did it start?

8.1 | Anthropometry

In their fascinating anthropometrical study of the inhabitants of Upper Egypt, from the earliest prehistoric times to the Mohammedan Conquest, Thomson and Randall-Maciver (1905) report on skull measurements for more than a thousand crania. A subset of their data is reported on and analyzed in Claeskens & Hjort (2008, chs. 1 and 9), see in particular their figures 1.1 and 9.1. This pertains to four cranium measurements, say $y=(y_1,y_2,y_3,y_4)^{\rm t}$, for 30 skulls, from each of five Egyptian time epochs, corresponding to $-4000,-3300,-1850,-200,150$ on our A.D. scale. We model these vectors as

$$Y_{j,i} \sim {\rm N}_4(\xi_j,\Sigma_j) \quad \text{for } i=1,\ldots,30,$$

for each of the five epochs j. There is a variety of parameters worth recording and analyzing, where the emphasis is on identifying the necessarily small changes over time, related to the history of emigration and immigration in ancient Egypt; see also Schweder & Hjort (2016, Example 3.10). For the present illustration, we choose to focus on the variance matrices, not the means, and consider

$$\psi = \{\max \text{eigen}(\Sigma)\}^{1/2}/\{\min \text{eigen}(\Sigma)\}^{1/2},$$

the ratio of the largest root-eigenvalue to the smallest root-eigenvalue of the variance matrix of the four skull measurements. This is the ratio of the largest to the smallest SDs of linear combinations $a^{\rm t}Y$ of the four skull measurements, normalized to have coefficient vector length $\|a\|=1$. This parameter is one of several natural measures of the degree to which the skull distribution is "stretched." The question is whether the stretch parameter ψ has changed over time. We assess the degree of change, if any, via the spread parameter τ in the natural model taking $\psi_1,\ldots,\psi_5 \sim {\rm N}(\psi_0,\tau^2)$. Rather than merely providing a test of the implied hypothesis $H_0:\psi_1=\cdots=\psi_5$, which is equivalent to $\tau=0$, with its inevitable p-value and a yes-no answer, we aim at giving a full CD for τ, applying the II-CC-FF scheme.

Table 1 gives point estimates

$$\widehat{\psi}_j = \{\max \text{eigen}(\widehat{\Sigma}_j)\}^{1/2}/\{\min \text{eigen}(\widehat{\Sigma}_j)\}^{1/2}$$

for the five-time epochs, along with estimated standard deviations σ_j for these estimators, the latter obtained via parametric bootstrapping from the estimated multinormal distributions. For

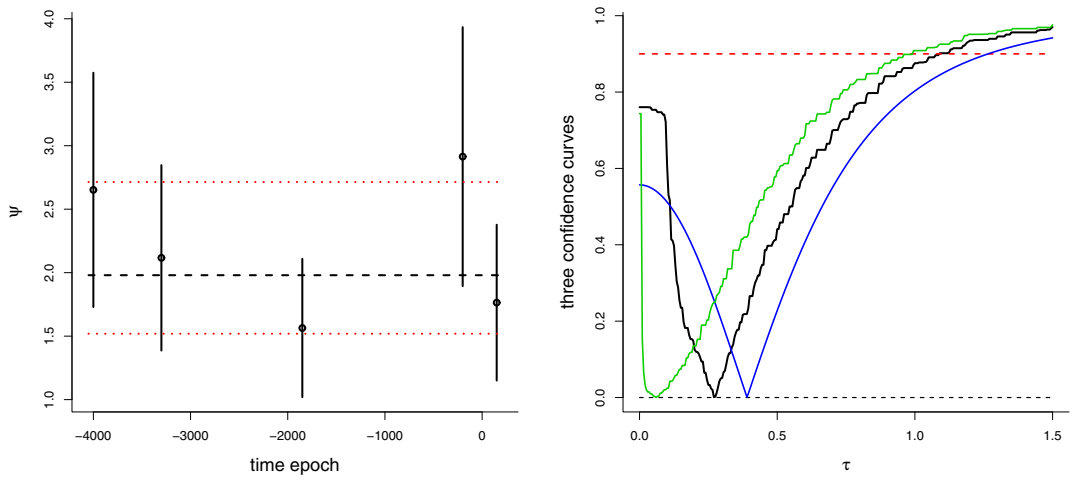


FIGURE 5 Left panel: Point estimates $\widehat{\psi}_j$ with 90% confidence intervals, for the skull stretch parameter ψ, across five time epochs (see Table 1). The dashed black line and the red dotted lines indicate the overall point estimate and 90% confidence interval for ψ_0. Right panel: Three confidence curves for the spread parameter τ, with median confidence estimates 0.006 (the direct profile method), 0.272 (the corrected profile method), 0.390 (the $Q_k(\tau)$ method) [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Skulls: For each of the five time epochs, the table gives the estimate $\widehat{\psi}$ and its estimated SD $\widehat{\sigma}$

epoch	$\widehat{\psi}$	$\widehat{\sigma}$
−4000	2.652	0.561
−3300	2.117	0.444
−1850	1.564	0.331
−200	2.914	0.620
150	1.764	0.373

Note: See Section 8.1 and Figure 5.

our present purposes the underlying distributions for the estimators are approximately normal, with the SDs σ_j approximately known. Figure 5 displays point estimates with 0.90 confidence intervals (left panel), for the five epochs.

Using log-likelihood fusion methods developed in Section 6.1, involving profiling and corrections, we may compute the confidence curves $\mathrm{cc}_{\mathrm{ml}}(\tau)$ and $\mathrm{cc}_{\mathrm{cml}}(\tau)$, as per (13). These involve simulation of a high number of deviance statistics for each candidate value of τ. The resulting confidence curves are shown in Figure 5 (right panel). The direct profile method can be shown to have a clear negative bias, particularly so for smaller values of k. For the present case of $k = 5$ the corrected version $\mathrm{cc}_{\mathrm{cml}}(\tau)$, with median confidence estimate 0.272, is better than the direct version $\mathrm{cc}_{\mathrm{ml}}(\tau)$, with median confidence estimate 0.006. A third and simpler to compute CD for τ is via

$$Q_k(\tau) = \sum_{j=1}^{k} \frac{\{\widehat{\psi}_j - \widehat{\psi}_0(\tau)\}^2}{\sigma_j^2 + \tau^2} \quad \text{and} \quad C(\tau, \text{data}) = 1 - \Gamma_{k-1}(Q_{k,\text{obs}}(\tau)),$$

TABLE 2 Abundance assessment of a humpback population, from 1995 and 2001, summarized as 2.5%, 50%, 97.5% confidence quantiles; from Paxton et al. (2009)

	2.5%	50%	97.5%
1995	3439	9810	21,457
2001	6651	11,319	21,214

Note: See Section 8.2 and Figure 6.

the point being that $Q_k(\tau)$ for a given true value of τ has the χ^2_{k-1} distribution; see Schweder & Hjort (2016, ch. 13). We note that these confidence curves have point masses at zero, hence also the associated CDs, via $\text{cc}(\tau) = |1 - 2C(\tau)|$. For each, the $C(0)$ has a clear interpretation as the p-value for testing $\tau = 0$ against $\tau > 0$. For the corrected profile CD, we find $C(0) = 0.123$, not small enough to warrant a claim that this particular ψ parameter has changed over the 4000 years of Egyptian history. An accurate 0.90 interval for τ, using the corrected profile, also indicated in the figure, is $[0, 1.085]$, with median confidence estimate 0.272.

In other applications the overall mean ψ_0 of the background distribution of the ψ_j might be of high importance. For the skulls analysis, the primary question is whether the ψ_j parameters have changed over the course of 4000 years, and the precise value of ψ_0 is of secondary importance. We report, though, that the corrected log-profile methods of Section 6.1, see (12), lead to an overall point estimate 1.980, with an accurate 90% interval stretching from 1.662 to 2.480. These intervals are not symmetric around the point estimate; see left panel of Figure 5.

8.2 | Abundance of humpback whales

The II-CC-FF paradigm readily lends itself to combination of information from published sources, where we may not have access to the full data, but only summary measures. Paxton et al. (2009) provide estimates of the abundance of humpback whales in the North Atlantic in the years 1995 and 2001. The two estimates are based on different surveys and can be considered independent. The authors also provide 95% confidence intervals, via a somewhat complicated model involving aggregation of line transect data from different areas via spatial smoothing, and also includes bootstrapping. The available information is as presented in Table 2; note here that the natural 95% confidence interval is not at all symmetric around the point estimate, with an implied skewness to the right.

For this illustration, we are interested in the underlying true abundances underlying these two studies. Let ψ_1 be the population size in 1995 and ψ_2 be the size in 2001. Our main interest may lie in the annual growth rate underlying these two population sizes. We define $\rho = (\psi_2 - \psi_1)/(6\psi_1)$, a simple (and in some sense approximate) definition of annual growth rate.

The first step, *independent inspection*, requires us to construct CDs for ψ_1 and ψ_2 from the two surveys. In Schweder & Hjort, 2016 (ch. 10), certain methods are proposed and developed for constructing CDs based only on an estimate and a confidence interval. With a positive parameter, like abundance, one may use

$$\text{II}: \quad C(\psi_j, y) = \Phi\left(\frac{h(\psi_j) - h(\widehat{\psi}_j)}{s}\right)$$

with a power transformation $h(\psi, a) = \text{sgn}(a)\psi^a$; see also Schweder and Hjort (2013) for some more discussion of this approach (along with a different application, essentially also using the

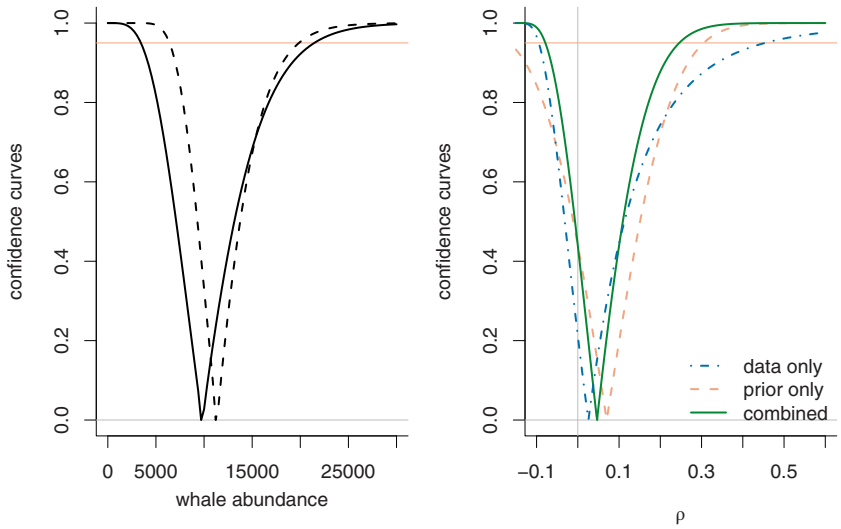


FIGURE 6 Left panel: Confidence curves for ψ_1 and ψ_2, the abundance of humpback whales in the North Atlantic in 1995 (fully drawn line) and 2001 (dashed line). Right panel: The confidence curve for $\rho = (\psi_2 - \psi_1)/(6\psi_1)$ based on the two surveys (dot-dashed, blue curve); the confidence curve based on prior information alone (dashed, orange curve); and the confidence curve combining the studies and the prior information (full, green curve). See Section 8.2 and Table 2 [Color figure can be viewed at wileyonlinelibrary.com]

II-CC-FF paradigm). To estimate the power a and the scale s the following two equations must be solved,

$$\psi_L^a - \widehat{\psi}^a = -1.96\, s \quad \text{and} \quad \psi_R^a - \widehat{\psi}^a = 1.96\, s,$$

where $[\psi_L, \psi_R]$ is the 95% confidence interval and $\widehat{\psi}$ the median confidence point estimate. For the whale abundance, we find (a, s) equal to $(0.321, 2.798)$ for 1995 and $(0.019, 0.007)$ for 2001 (a small value of a indicates that the transformation is nearly logarithmic). The corresponding confidence curves are shown in the left panel of Figure 6. In this case the confidence log-likelihoods in the *confidence conversion* step are easily obtained. For year j,

$$\text{CC}: \quad \ell_{\text{conv},j}(\psi_j) = -\frac{1}{2}\{h_j(\psi_j) - h_j(\widehat{\psi}_j)\}^2/s_j^2.$$

In the final *focused fusion* step, we sum the two confidence log-likelihoods, profile with respect to ρ, find the combined deviance function, and construct an approximative combined confidence curve by the Wilks theorem, as per Section 2:

$$\begin{aligned}\text{FF}: \quad \ell_{\text{fus,prof}}(\rho) &= \max_{\psi_1,\psi_2}\{\ell_{\text{conv},1}(\psi_1) + \ell_{\text{conv},2}(\psi_2) \,:\, (\psi_2 - \psi_1)/(6\psi_1) = \rho\},\\ \text{cc}^*(\rho) &= \Gamma_1(2\{\ell_{\text{fus}}(\widehat{\rho}) - \ell_{\text{fus}}(\rho)\}).\end{aligned}$$

Here, we obtain the blue curve in the right panel of Figure 6, with $\widehat{\rho} = 0.026$ and a 95% confidence interval $[-0.094, 0.454]$.

In some cases, there may exist some expert knowledge pertaining to at least the focus parameter under study, here the annual growth rate ρ, though not necessarily for the full parameter

vector of the combined models, here (ψ_1, ψ_2) the two population sizes. A proper Bayesian analysis requires the statistician to have such a prior for (ψ_1, ψ_2)—without this ingredient, there is no Bayes theorem leading to a posterior distribution for the model parameters, or indeed for ρ. The II-CC-FF scheme allows, however, incorporation of such partial prior information, that is, a prior for ρ without a prior for (ψ_1, ψ_2). For this illustration, we assume that whale biologists provide a normal prior with expectation equal to 0.07 and variance 0.12^2. This prior may come from knowledge of other humpback whale populations or simulation-based life-history models (see, e.g., Zerbini et al. (2010), giving a similar point estimate as we have used).

The prior can be represented as a confidence curve, supplementing the confidence curve based on the two studies. To fuse the prior knowledge and the data we simply add the prior log-likelihood $\ell_B(\rho)$ to the confidence log-likelihoods, in the following way,

$$\text{FF:} \quad \ell_{\text{fus,prof},B}(\rho) = \ell_{\text{fus,prof}}(\rho) + \ell_B(\rho).$$

We use "B" as a subscript to indicate the in this instance partial and perhaps lazy Bayesian, who does not give a full prior for the model parameters, but contributes a component, namely where it matters the most, about the focus parameter. Of course the log-prior $\ell_B(\rho)$ employed here could have been obtained in the more careful and proper Bayesian way of having started with a full prior for (ψ_1, ψ_2), and then a transformation, but we do suggest that expert knowledge concerning focus parameters is most often put forward directly, not via the full parameter vector.

Importantly, the extended deviance function coming from the log-likelihood in the fusion step still has an approximate χ^2_1 distribution, by the general approximation arguments briefly discussed in Section 4.3, unless the log-prior $\ell_B(\rho)$ is sharp and distinctly nonnormal. One may conceptually and sometimes practically interpret the log-prior as having resulted from real data in previous experiences, in which case the $\ell_B(\rho)$ would be a genuine profiled log-profile likelihood function from such a source. In addition, as the sample sizes of the studies increase the information from the two studies will dominate the prior and we can safely continue to use the Wilks theorem. As expected, the confidence curve fusing the prior information and the information from the two studies lies between the original confidence curve and the prior confidence curve (see Figure 6, right panel). It is also somewhat narrower than both.

8.3 | Olympic medians

This is an Olympic story about certain dynamic changes of the Body Mass Index distribution for speedskaters. We focus specifically on how the median of this distribution has changed over time. Our use of the II-CC-FF scheme will involve first building highly accurate nonparametric CDs for the medians, and then converting these to parametric log-likelihoods, after which a dynamic model for the medians can be fused together in the end. Our methodology will work with modest changes also in cases where interest lies in the evolution of any given quantile, say the 0.90 quantile points of income distributions over time across different strata.

The BMI is defined as weight (in kg) divided by squared height (in meters). The left panel of Figure 8 displays the median BMI, for the male participants, across the 10 last Olympics (1984, 1988, 1992, 1994, 1998, 2002, 2006, 2010, 2014, 2018), marked as small red circles. The red line adjoining these sample medians indicate that the BMI has undergone a certain evolution, with a marked increase up to perhaps 1998 Nagano or 2002 Salt Lake City, followed by a downward trend. Even though these changes perhaps do not qualify as drastic, and athletes with 24.5 do not

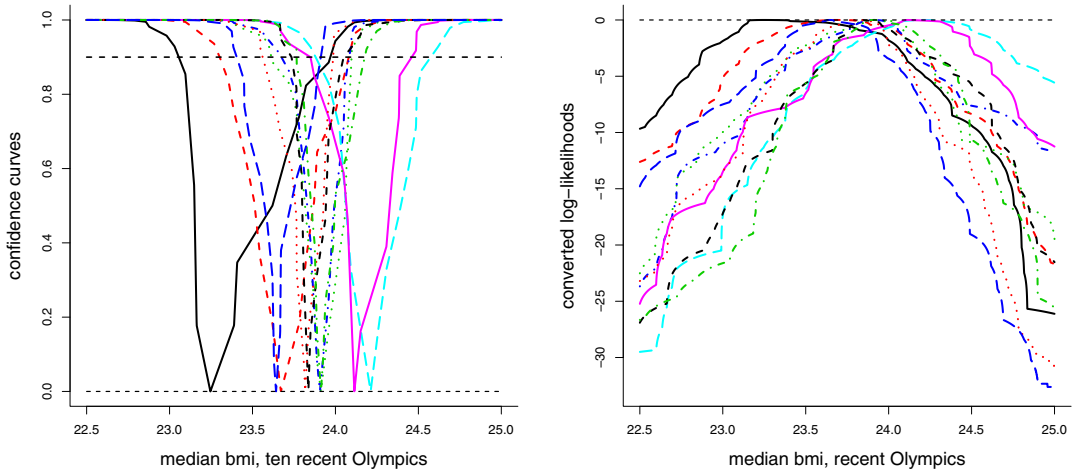


FIGURE 7 Left: Nonparametric confidence curves $\mathrm{cc}_j(\mu_j)$ for the medians of the BMI distributions, for the 10 last Olympics 1984 to 2018. Ninety percent confidence intervals are read off from the horizontal dashed line at 0.90. Right: Converted log-likelihood contributions $\ell_{\mathrm{conv},j}(\mu_j)$ for the 10 Olympic medians [Color figure can be viewed at wileyonlinelibrary.com]

differ very much from those with 23.5, they are nevertheless interesting enough to be discussed in the proper fora.

This motivates the following investigation, to look both for significant changes and for the position of a potential top-point for the evolution of the BMI distribution over time. Let μ_j be the population median at Olympics j, for occasions $j = 1, \ldots, k$. We first need CDs for these, constructed nonparametrically. With $y_{j,(1)} < \ldots < y_{j,(n_j)}$ the ordered sample from Olympics j, assumed to come from a continuous distribution with positive density f_j and cumulative F_j, we start from the exact calculation

$$P_{f_j}\{\mu_j \le y_{j,(r)}\} = P\{F_j(\mu_j) \le F_j(y_{j,(r)})\} = P\left\{\frac{1}{2} \le U_{j,(r)}\right\},$$

with $U_{j,(1)} < \ldots < U_{j,(n_j)}$ an ordered i.i.d. sample from the uniform distribution on the unit interval. But these have known Beta distributions. We therefore define the $C_j(\mu_j)$ for all values of μ_j by first setting

$$C_j(y_{j,(r)}) = 1 - \mathrm{Be}\left(\frac{1}{2}, r, n_j - r + 1\right) \quad \text{for } r = 1, \ldots, n_j,$$

featuring cumulative Beta distribution functions, and then applying linear interpolation between the ordered data points. The full confidence curves $\mathrm{cc}_j(\mu_j) = |1 - 2\,C_j(\mu_j)|$ are displayed in Figure 7 (left panel) and they are then converted to log-likelihood components $\ell_{\mathrm{conv},j}(\mu_j) = -\frac{1}{2}\Gamma_1^{-1}(\mathrm{cc}_j(\mu_j))$ (right panel). The 90% confidence intervals for the 10 medians shown in the left panel of Figure 8 are computed from the $\mathrm{cc}_j(\mu_j)$. Note that these confidence curves and intervals are fully nonparametric. They can be shown to be highly accurate, even for smaller sample sizes, and work better than alternative methods involving approximate normality with estimates of standard errors. The resulting intervals deviate from symmetry, reflecting underlying asymmetry in the distribution of the BMI.

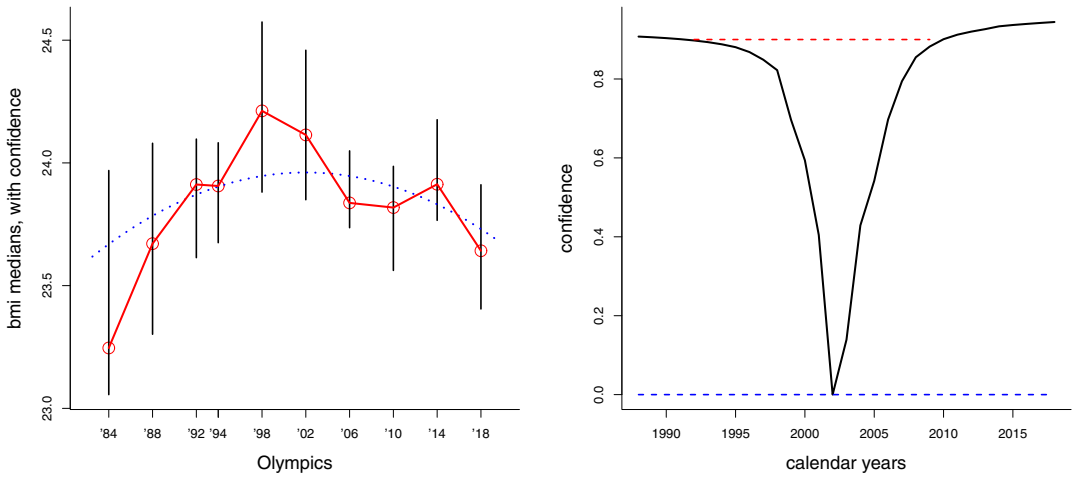


FIGURE 8 Left: Median BMI for men, 10 last Olympics 1984 Sarajevo to 2018 Sochi, with 90% confidence band, and the fitted parabola (19). Right: Focused fusion confidence curve for $x^* = -\beta_1/(2\beta_2)$, the maximum point for the parabola in the model for Olympic BMI medians [Color figure can be viewed at wileyonlinelibrary.com]

Next we model the medians μ_j dynamically as

$$\mu_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 \quad \text{for } j = 1, \ldots, k, \tag{19}$$

where $x_j = t_j - t_1$ is the time passed since the first of the Olympics to no. j. The parameters of this parabola will be such that it has a maximum point

$$x^* = x^*(\beta_0, \beta_1, \beta_2) = -\beta_1/(2\beta_2) \tag{20}$$

within the range from the first to the last of these Olympics, see again Figure 8 (left panel). Now the FF step of our general recipe leads to $\ell_{\rm fus}(\beta_0, \beta_1, \beta_2) = -\frac{1}{2}\sum_{j=1}^k \Gamma_1^{-1}({\rm cc}_j(\beta_0 + \beta_1 x_j + \beta_2 x_j^2))$, and then to the profiled version

$$\ell_{\rm fus}(x^*) = \max\{\ell_{\rm fus}(\beta_0, \beta_1, \beta_2) : (\beta_0, \beta_1, \beta_2) \text{ fitting with } x^*\}.$$

Using the FF step of our II-CC-FF is also equivalent to working with the fused deviance function $D_{\rm fus}(x^*) = 2\{\ell_{\rm fus}(\widehat{x}^*) - \ell_{\rm fus}(x^*)\}$, with $\widehat{x}^*$ the appropriate function of the overall maximizers of $\ell_{\rm fus}(\beta_0, \beta_1, \beta_2)$.

Carrying out this machinery leads first to the fitted parabola shown as the blue dashed curve in left panel of Figure 8. The linear and quadratic coefficients β_1 and β_2 are very significantly positive and negative, respectively, as shown via Wald ratios; also, the estimated top-point is at $\widehat{x}^* = 2002.4$. Second, using the distribution approximation $D_{\rm fus}(x^*) \sim \chi_1^2$ we can execute the FF step and compute the confidence curve ${\rm cc}_{\rm fus}(x^*) = \Gamma_1(D_{\rm fus}(x^*))$. It is displayed in the right panel of Figure 8, with the partial ruggedness and asymmetry reflecting the nonsmoothness of the 10 nonparametric sample median operations. We have been able to use CC, from nonparametric confidence curves to log-likelihood contributions and then back again to a FF confidence curve. The 90% confidence interval for the point of maximum median BMI for male speedskaters is from 1992.1 to 2009.7.

8.4 | The long peace

Our last illustration concerns the use of the II-CC-FF framework in a highly nonstandard setting, where one wishes to combine hard data, sources that inform directly on the focus parameter, with softer data sources, which only contain indirect or noisier information about the focus parameter. This kind of combination has wide potential in various fields where "soft" data could be based on webscraping, using Twitter accounts or other social media, but raises specific issues and challenges. The question we investigate here is the extent of statistical evidence for The long peace, the period of relative peace and stability following the second world war (and still lasting, presumably). Specifically, do we find evidence of a change-point τ when analyzing the sequence of battle deaths in interstate wars between 1823 and today?

There are many components, issues, and details involved in this application story, and a fuller version is reported on in Appendix B. Here, we outline the main statistical ingredients. First, the question has been investigated in Cunen et al. (2020) using the correlates of war dataset (Sarkees & Wayman, 2010). The authors found evidence of an abrupt change in the battle death distribution at some point after the second world war, from a distribution with a high median battle death to a distribution with a lower median (and also a less heavy tail). This involved establishing a certain three-parameter model for the battle deaths distribution, with parameters changing at time τ, itself an unknown change-point parameter. We may view the relevant statistical information as represented by the log-likelihood contribution $\ell_{B,\text{prof}}(\tau)$. Here, the aim is to extend this analysis and investigate whether there might be benefits in combining the battle death data with other sources assumed to be informed on τ.

Some political scientists consider the aforementioned decrease in battle deaths to reflect a moral and political shift within a large portion of the world's population. At some point in the 20th century, it is argued, the perception of war changed, from being seen as something natural and inevitable, sometimes even positive, to being perceived as highly negative, evil and unacceptable; cf. Pinker (2011, ch. 5). This change in norms has likely manifested itself in various ways, including cultural, artistic, and political expressions, for example through text. We have therefore collected sequences representing the usage of certain relevant words or phrases, like "anti-war" or "pacifist," and then attempted to combine the change-point inference from such an Ngram analysis (suggested to us by Steven Pinker, personal communication), with the change-point inference from the battle deaths data. Such statistical work, along with data collection and separate modeling efforts, leads to an overall log-likelihood contribution $\ell_{N,\text{prof}}(\tau)$ for the potential change-point τ. These combination efforts lead to an overall log-likelihood fusion function

$$\ell_{\text{fus}}(\tau) = w_B \ell_{B,\text{prof}}(\tau) + w_N \ell_{N,\text{prof}}(\tau),$$

with relative importance weights w_B and w_N, involving a separate discussion. Other applications, involving the combination of perhaps very different data sources, would similarly involve separate and perhaps partly subjective analyses for deciding on such relative importance weight parameters. For the present application, with battle deaths the hard data and Ngrams the soft data, more details are presented in Appendix B, along with our tentative concluding confidence curve $\text{cc}^*(\tau)$.

ACKNOWLEDGMENTS

The work reported on here has been partially funded via the Norwegian Research Council's project *FocuStat: Focused Statistical Inference With Complex Data*, led by Hjort. The

authors are grateful to comments from Steven Pinker, and to Tore Schweder and Emil Stoltenberg for always fruitful discussions related to issues and methods worked with in this article. Constructive comments and suggestions from anonymous reviewers and an associate editor have contributed to a better paper. The data on heights and weights and hence BMI for Olympic speedskaters stem from files collected over the years by N.L.H., and in this endeavor he has been helped by fellow speedskating historians Arild Gjerde and Jeroen Heijmans.

ORCID

Céline Cunen https://orcid.org/0000-0002-8215-3080

REFERENCES

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, *73*, 307–322.

Berger, J. O., Liseo, B., & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters [with discussion]. *Statistical Science*, *14*, 1–28.

Breslow, N. (1981). Odds ratio estimators when the data are sparse. *Biometrika*, *68*, 73–84.

Cai, T., Parast, L., & Ryan, L. (2010). Meta-analysis for rare events. *Statistics in Medicine*, *29*, 2078–2089.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University Press.

Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference [with discussion]. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *49*, 1–39.

Cox, D. R., & Reid, N. (1993). A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *55*, 467–471.

Cunen, C., & Hjort, N. L. (2015). Optimal inference via confidence distributions for two-by-two tables modelled as Poisson pairs: Fixed and random effects. *Proceedings of the 60th World Statistics Congress, 26-31 July 2015, Rio de Janeiro*, *Amsterdam: International Statistical Institute* (vol. I, pp. 3581–3586).

Cunen, C., Hjort, N. L., & Nygård, H. M. (2020). Statistical sightings of better angels: Analysing the distribution of battle-deaths in interstate conflict over time. *Journal of Peace Research*, *57*, 221–234.

Diciccio, T., & Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika*, *79*, 231–245.

DiCiccio, T. J., Martin, M. A., Stern, S. E., & Young, G. A. (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *58*, 189–203.

Dominici, F., & Parmigiani, G. (2000). *Combining studies with continuous and dichotomous responses: A latent-variables approach*. In *Meta-analysis in medicine and health policy* (pp. 99–118). CRC Press.

Durban, M., & Currie, I. D. (2000). Adjustment of the profile likelihood for a class of normal regression models. *Scandinavian Journal of Statistics*, *27*, 535–542.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika*, *80*, 3–26.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference*. Cambridge University Press.

Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall.

Hardy, R. J., & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, *15*, 619–629.

Hjort, N. L., & Schweder, T. (2018). Confidence distributions and related themes [introduction to the special issue, by the guest editors]. *Journal of Statistical Planning and Inference*, *195*, 1–13.

Jackson, D., Law, M., Stijnen, T., Viechtbauer, W., & White, I. R. (2018). A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Statistics in Medicine*, *37*, 1059–1085.

Kristensen, K., Nielsen, A., Berg, C., Skaug, H., & Bell, B. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, *70*, 1–21.

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*, 83–98.

Lehmann, E., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer.

Liu, D., Liu, R., & Xie, M.-G. (2014). *Exact inference methods for rare events* (pp. 1–6). *Wiley StatsRef: Statistics Reference Online*.

Liu, D., Liu, R. Y., & Xie, M.-G. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. *Journal of the American Statistical Association*, *110*, 326–340.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719–748.

McCullagh, P., & Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *52*, 325–344.

Noma, H. (2011). Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine*, *30*, 3304–3312.

O'Rourke, K. (2008). *The combining of information: Investigating and synthesizing what is possibly common in clinical observations or studies via likelihood* (Ph.D. thesis), University of Oxford, Oxford, UK.

Partlett, C., & Riley, R. D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in Medicine*, *36*, 301–317.

Paxton, C. G. M., Burt, M. L., Hedley, S. L., Víkingsson, G. A., Gunnlaugsson, T., & Desportes, G. (2009). *Density surface fitting to estimate the abundance of Humpback whales based on the NASS-95 and NASS-2001 aerial and shipboard surveys* (Vol. *7*, pp. 143–160). NAMMCO Scientific Publications.

Piaget-Rossel, R., & Taffé, P. (2019). Meta-analysis of rare events under the assumption of a homogeneous treatment effect. *Biometrical Journal*, *61*, 1557–1574.

Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Viking Books.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Lippincott Williams & Wilkins.

Sarkees, M. R., & Wayman, F. W. (2010). *Resort to war: A data guide to inter-state, extra-state, intra-state, and non-state wars, 1816-2007*. CQ Press.

Schweder, T., & Hjort, N. L. (2013). Integrating confidence intervals, likelihoods and confidence distributions. *Proceedings of the 59th World Statistics Congress, 25-30 August 2013, Hong Kong*, *Amsterdam International Statistical Institute,* (vol. I. pp. 277–282).

Schweder, T., & Hjort, N. L. (2016). *Confidence, likelihood, probability: Statistical inference with confidence distributions*. Cambridge University Press.

Simmonds, M. C., & Higgins, J. P. (2016). A general framework for the use of logistic regression models in meta-analysis. *Statistical Methods in Medical Research*, *25*, 2858–2877.

Simpson, R. J. S., & Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, *3*, 1243–1246.

Singh, K., Xie, M.-G., & Strawderman, W. E. (2005). Combining information from independent sources through confidence distributions. *Annals of Statistics*, *33*, 159–183.

Stern, S. E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *59*, 653–665.

Stijnen, T., Hamza, T. H., & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, *29*, 3046–3067.

Thomson, A., & Randall-Maciver, R. (1905). *Ancient races of the Thebaid: Being an anthropometrical study of the inhabitants of Upper Egypt from the earliest prehistoric times to the Mohammedan conquest, based upon the examination of over 1,500 crania*. Oxford University Press.

Van Houwelingen, H. C., Zwinderman, K. H., & Stijnen, T. (1993). A bivariate approach to meta-analysis. *Statistics in Medicine*, *12*, 2273–2284.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Whitehead, A., Bailey, A. J., & Elbourne, D. (1999). Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. *Journal of Biopharmaceutical Statistics*, *9*, 1–16.

Xie, M., Singh, K., & Strawderman, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, *106*, 320–333.

Xie, M.-G., & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review [with discussion and a rejoinder]. *International Statistical Review*, *81*, 3–39.

Yang, G., Shi, P., & Xie, M.-G. (2016). *gmeta: Meta-analysis via a unified framework of confidence distribution*. R package version 2.2-4.

Yu, Y., & Tian, L. (2014). *exactmeta: Exact fixed effect meta analysis*. R package version 1.0-2.

Zerbini, A. N., Clapham, P. J., & Wade, P. R. (2010). Assessing plausible rates of population growth in humpback whales from life-history data. *Marine Biology*, *157*, 1225–1236.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cunen C, Hjort NL. Combining information across diverse sources: The II-CC-FF paradigm. *Scand J Statist*. 2021;1–32. https://doi.org/10.1111/sjos.12530