

**A PSEUDO-LIKELIHOOD APPROACH FOR THE
META-ANALYSIS OF HOMOGENEOUS TREATMENT
EFFECTS: EXPLOITING THE INFORMATION
CONTAINED IN SINGLE-ARM AND
DOUBLE-ZERO STUDIES**

ROMAIN PIAGET-ROSSEL and PATRICK TAFFÉ

Unisanté-Centre Universitaire de Médecine Générale
et Santé Publique

Route de la Corniche 10

1010 Lausanne

Switzerland

e-mail: romain.piaget-rossel@unil.ch

patrick.taffe@chuv.ch

Abstract

Mantel-Haenszel is a fixed-effect meta-analysis method, which performs quite well under the assumption of a homogeneous treatment effect, even in the presence of very rare events. However, this method fails to account for the information contained in single-arm and double zero studies. In this paper, we developed a pseudo-likelihood approach, which allows the inclusion of both single-arm and double-zero studies in the combined effect size estimate. Using Monte-Carlo simulations, we evaluated the behaviour of these two methods when subject to an increasing proportion of single-arm and double-zero studies. We found that the exclusion of double-zero studies did not impact the performance of the Mantel-Haenszel method, whereas the exclusion of single-

2010 Mathematics Subject Classification: Primary 92B15; Secondary 62P10.

Keywords and phrases: double-zero studies, Mantel-Haenszel, meta-analysis, Monte-Carlo simulations, pseudo-likelihood approach, single-arm studies.

Received April 12, 2019; Revised May 3, 2019

arm studies reduced its efficiency compared to the pseudo-likelihood approach. We thus recommend using the pseudo-likelihood approach when the meta-analysis includes single-arm studies. With only double-zero studies, the Mantel-Haenszel can safely be used.

1. Introduction

Meta-analysis is concerned with the synthesis of information contained in independent but related studies, called “primary studies” (Normand [12]). With homogeneous treatment effects (i.e., under the “fixed-effect model”), each primary study seeks to estimate the same population parameter, commonly referred to as “effect size” (ES), and the objective of a meta analysis is to combine all the available evidence into one single and more precise estimate. Under this framework, Piaget-Rossel and Taffé [13] have shown that the Mantel-Haenszel (MH) method without continuity correction (CC) performs very well, even in the presence of very rare events. However, this method excludes double-zero (i.e. studies reporting zero event in both control and treatment arms; DZ) and single-arm (i.e., studies that report results for only one arm; SA) studies from the computation of the combined ES.

DZ studies are typically encountered in rare events settings, where the probability of the event of interest can be so small that it might be unfeasible to design a study with proper sample size, (i.e., so that at least a couple of events are observed). Using a CC allows the inclusion of DZ studies (Sweeting et al. [15]). However, this method is usually not recommended (Efthimiou [4]). Indeed, not only do the resulting estimates depend on the choice of correction used (Kuss et al. [7]; Keus et al. [6]), but this also introduces a bias in the estimates (Piaget-Rossel and Taffé [13]).

Although almost never discussed in the literature, the issue of SA studies is an important one, especially when considering the meta-analysis of observational studies. For example, a systematic review on the surgical management of phyllodes tumors of the breast found that 3 out of 11 studies only included patients with a resection margin above or equal to 10mm, (i.e., these studies did not have any patients with a margin below 10mm) (Toussaint [16]).

Intuitively, SA and DZ studies do carry some information regarding the probability of the event. Consequently, it is useful to devise statistical methods, which allows the inclusion of this information in the computation of the ES estimate. In this paper, to tackle this issue, we have adopted a pseudo-likelihood (PL) approach, which allows the inclusion of SA and DZ studies, without the use of a CC. The basic idea was to adopt a working model for the counts in each arm and treat the (heterogeneous) baseline prevalences as random nuisance parameters. The distribution of the nuisance parameter was not assumed to be known, hence the denomination “pseudo-likelihood”.

In this paper, we investigated the performance of this PL approach, which explicitly includes SA and DZ studies, and compared it with the MH method. By means of Monte-Carlo simulations, we evaluated the behaviour of these two methods when subject to an increasing proportion of DZ and SA studies. ES of interest were the odds ratio (OR), the relative risk (RR), and the risk difference (RD), which are the three most commonly-used ES in meta-analyses of binary data. We focused on the specific framework of a homogeneous treatment effect, as the MH method can be seen as the gold standard when there are no SA and DZ studies (Piaget-Rossel and Taffé [13]). Baseline prevalences were assumed to be heterogeneous, as the setting of homogeneous baseline prevalences is more restrictive. The remainder of this paper is structured as follows. In Section 2, we present the PL approach. Section 3 describes our simulation model and presents the results obtained. Section 4 outlines the main findings and makes some recommendations regarding the best method to use when conducting a FE meta-analysis in the presence of SA or DZ studies.

2. The Pseudo-likelihood Approach

Under the assumption of heterogeneous baseline prevalences, the binomial likelihood writes:

$$L = \prod_{k=1}^K \prod_{j \in \{t, c\}} \binom{n_{kj}}{x_{kj}} \pi_{kj}(\alpha_k)^{x_{kj}} (1 - \pi_{kj}(\alpha_k))^{n_{kj} - x_{kj}},$$

where n_{kj} is the sample size of arm j ($j = c$ for control and t for treatment) in study k , x_{kj} is the number of events occurring in arm j of study k , $\pi_{kj}(\alpha_k)$ is the inverse of the link function (i.e., the probability of the event) in arm j of study k , and α_k is a nuisance parameter. To deal with the nuisance parameters, one can either treat them as fixed or random quantities. Particularly with rare events, it is advantageous to treat these quantities as random parameters to limit as much as possible the number of parameters to be estimated and allow SA and DZ studies to contribute to the estimation:

$$L = \prod_{k=1}^K \int_{-\infty}^{+\infty} \left[\prod_{j \in \{t, c\}} \binom{n_{kj}}{x_{kj}} \pi_{kj}(\alpha_k)^{x_{kj}} (1 - \pi_{kj}(\alpha_k))^{n_{kj} - x_{kj}} \right] f(\alpha_k) d\alpha_k, \quad (1)$$

where $\pi_{kj}(\alpha_k)$ is the probability of event in study k and treatment arm j defined as a function of α_k , and $f(\alpha_k)$ is the density function of the random variable α_k . Usually, the density function $f(\alpha_k)$ is unknown and to cope with it we have adopted a pseudo-likelihood approach.

For estimating the OR, the following pseudo-likelihood may be used:

$$L^P = \prod_{k=1}^K \int_{-\infty}^{+\infty} \left[\prod_{j \in \{t, c\}} \binom{n_{kj}}{x_{kj}} \pi_{kj}(\alpha_k)^{x_{kj}} (1 - \pi_{kj}(\alpha_k))^{n_{kj} - x_{kj}} \right] \phi(\alpha_k | \alpha, \sigma^2) d\alpha_k, \quad (2)$$

with link function $\text{logit}(\pi_{kj}(\alpha_k)) = \alpha_k + \beta T_{kj}$, where T_{kj} is an indicator for treatment arm (i.e., $T_{kj} = 1$ if $j = t$ and 0 if $j = c$), and $\phi(\cdot | \alpha, \sigma^2)$ is the normal density with mean α and variance σ^2 . Note that β corresponds to the $\log(\text{OR})$. This is a pseudo-likelihood since in Equation (1), $\phi(\cdot | \alpha, \sigma^2)$ is not assumed to be the true density function, it is only a “working” density function.

To estimate the RR, one may adopt the log link function $\log(\pi_{kj}(\alpha_k)) = \alpha_k + \beta T_{kj}$. However, to insure probabilities contained within 0-1, one has to constrain the α_k parameter during the optimization procedure (Marschner and Gillett [10]). The main drawback of this approach is that the imposition of the parameter constraint may lead to biased estimates, particularly when the risk level is either low or high. Therefore, to cope with this issue, we proposed to approximate the binomial distribution by a Poisson distribution with parameter $\lambda_{kj} = \pi_{kj} * n_{kj}$ and used the pseudo-likelihood function

$$L^P = \prod_{k=1}^K \int_{-\infty}^{+\infty} \left[\prod_{j \in \{t, c\}} \frac{\lambda_{kj}(\alpha_k)^{x_{kj}} \exp(-\lambda_{kj}(\alpha_k))}{x_{kj}!} \right] \phi(\alpha_k | \alpha, \sigma^2) d\alpha_k. \quad (3)$$

The numerical advantages of this approach are obvious as the log link is canonical for the Poisson likelihood. In this model, β corresponds to the $\log(\text{RR})$.

Similarly, for the RD, we used the canonical identity link function and approximated the binomial distribution of the counts by a normal distribution for the proportions:

$$L^P = \prod_{k=1}^K \int_{-\infty}^{+\infty} \left[\prod_{j \in \{C, T\}} \frac{1}{\sqrt{2\pi * \tau_{kj}^2}} \exp\left(\frac{p_{kj} - \pi_{kj}(\alpha_k)}{2\tau_{kj}^2}\right) \right] \phi(\alpha_k | \alpha, \sigma^2) d\alpha_k, \quad (4)$$

where $p_{kj} = x_{kj} / n_{kj}$, $\pi_{kj} \equiv E(p_{kj}) = \alpha_k + \beta T_{kj}$, and $\tau_{kj}^2 \equiv \text{Var}(p_{kj}) = \pi_{kj} * (1 - \pi_{kj}) / n_{kj}$. Given that the variances τ_{kj}^2 are unknown and difficult to estimate (the above model is highly nonlinear and computation of the integral with good precision is challenging, particularly with rare events), we decided to use instead the robust sandwich estimate of the variance-covariance matrix (White [17]). Moreover, to achieve appropriate empirical coverage rates, we selected the 98.5th quantile of the standard normal distribution to compute the

Wald CI for the RD. This yielded a “calibrated” CI in the spirit of that obtained by using the modified Hartung-Knapp-Sidik-Jonkman method (Röver et al. [14]).

3. Simulation Study

3.1. Model

Given our objective (study the impact of SA and DZ studies), we decided to consider large meta-analyses to avoid fluctuations issues related to scarcity of primary studies. Therefore, we set the number of primary studies at $K = 20$. Although this scenario might seem optimistic, as many published meta-analyses, notably in the Cochrane Library, included less primary studies, it is nevertheless a realistic one (see, for instance, Moher et al. [11], which found a median number of 23 studies out of 88 systematic reviews analyzed).

In each primary-study, treatment arms’ sample sizes ranged from 50 to 150 (i.e., $n_{kt} \sim \text{discrete-uniform } \{50; 150\}$). Control arms’ sample sizes were generated as $n_{kc} = n_{kt} + r$, with $r \sim \text{discrete-uniform } \{-15; 15\}$. Baseline prevalences π_{kc} were obtained as random draws from a continuous uniform distribution with range $[p - p/5; p + p/5]$, p being the mean of the distribution. This distribution provided a realistic level of heterogeneity in baseline prevalences under the assumption of a homogeneous treatment effect. Probabilities in the treated group π_{kt} were derived from the control probabilities and ES considered (i.e., $\pi_{kt} = \pi_{kc} * OR / (\pi_{kc} * OR + 1 - \pi_{kc})$; $\pi_{kt} = \pi_{kc} * RR$; $\pi_{kt} = \pi_{kc} + RD$). Finally, the number of events in both arms were generated by two binomial draws with respective sample sizes and event probabilities. Notice that our simulations models are different from the models described in (2), (3), or (4), hence the terminology “pseudo-likelihood”. We investigated the impact of both DZ and SA studies on the performance of the proposed approaches. To study the impact of DZ studies, we considered four

scenarios with different mean values for the baseline prevalences $p = \{0.1, 0.007, 0.0035, 0.0015\}$. Under the null hypothesis of no treatment effect, these four probabilities yielded approximately 0, 25, 50, and 75% of DZ studies per meta-analysis, respectively. When studying the impact of SA studies, we set $p = 0.1$, which ensured almost all meta-analyses to be free of DZ studies. Each primary study was defined as SA or not-SA study by means of a Bernoulli draw with probability m . In the subset of SA studies, the arm to be removed was then designated by a Bernoulli draw with probability 0.5. We considered four different scenarios with respective m values $\{0, 0.25, 0.5, 0.75\}$.

The impact of SA and DZ studies was assessed for various values of the ES, which are reported in Table 1. Since the RD is an absolute measure, we had to derive it from the $\log(\text{RR})$ and mean baseline prevalence p , to avoid generating probabilities in the treatment group below 0. For instance, for $p = 0.15\%$ and $\log(\text{RR}) = -1.5$, one obtained RD as $\exp(-1.5) * 0.0015 - 0.0015 \cong -0.0012$.

For each scenario considered, 10,000 meta-analyses were generated. For each of the generated meta-analyses, we estimated the ES of interest, its standard error and the 95% Wald confidence interval (CI). For the RD, we computed a calibrated CI using the 98.5th quantile; for the OR and RR, the usual 97.5th quantile was used. Performance of the MH and PL methods were assessed in terms of bias, coverage rate and width of the CI. We decided to compute median instead of mean values for the bias and CI's width to avoid the influence of exceedingly large or small values. We also reported the proportion of converged runs. For the MH method, a run was reported as non-converged when either of the estimates obtained (i.e., the ES or its variance) were undefined. For the OR and the RR, this happened when either $x_{kt} = 0$ or $x_{kc} = 0, \forall k$ (or both); for the RD this happened only when both $x_{kt} = 0$ and $x_{kc} = 0, \forall k$. Non-converged runs were also reported by the MH method when the K primary studies

were SA studies. Both OR and RR estimates were analyzed on the log scale because the sampling distribution was more symmetrical on this scale.

Table 1. Size of the treatment effect considered in the simulations

p	$\log(\text{OR})^*$	$\log(\text{RR})^{**}$	RD (in %)	Description of the effect
0.0015	- 1.5 (0.22)	- 1.5 (0.22)	- 0.12	Large negative ES
0.0035	- 1.5 (0.22)	- 1.5 (0.22)	- 0.27	Large negative ES
0.007	- 1.5 (0.22)	- 1.5 (0.22)	- 0.54	Large negative ES
0.1	- 1.5 (0.22)	- 1.5 (0.22)	- 7.77	Large negative ES
0.0015	- 0.5 (0.61)	- 0.5 (0.61)	- 0.06	Moderate negative ES
0.0035	- 0.5 (0.61)	- 0.5 (0.61)	- 0.14	Moderate negative ES
0.007	- 0.5 (0.61)	- 0.5 (0.61)	- 0.28	Moderate negative ES
0.1	- 0.5 (0.61)	- 0.5 (0.61)	- 3.93	Moderate negative ES
0.0015	0 (1)	0 (1)	0	Null ES
0.0035	0 (1)	0 (1)	0	Null ES
0.007	0 (1)	0 (1)	0	Null ES
0.1	0 (1)	0 (1)	0	Null ES
0.0015	0.5 (1.65)	0.5 (1.65)	0.10	Moderate positive ES
0.0035	0.5 (1.65)	0.5 (1.65)	0.23	Moderate positive ES
0.007	0.5 (1.65)	0.5 (1.65)	0.45	Moderate positive ES
0.1	0.5 (1.65)	0.5 (1.65)	6.49	Moderate positive ES
0.0015	1.5 (4.48)	1.5 (4.48)	0.52	Large positive ES
0.0035	1.5 (4.48)	1.5 (4.48)	1.22	Large positive ES
0.007	1.5 (4.48)	1.5 (4.48)	2.44	Large positive ES
0.1	1.5 (4.48)	1.5 (4.48)	34.82	Large positive ES

Note. p = mean baseline prevalence;

*ORs reported between brackets;

**RRs reported between brackets.

3.2. Results

3.2.1. Double-zero studies

Reducing the mean baseline prevalence increased the proportion of DZ studies and, thus, the proportion of excluded studies by the MH method when computing the OR (Table 2). On the other hand, with the PL approach DZ studies are allowed and no study is excluded. We found that even when the proportion of discarded studies was very high (75%), the MH method still performed very well, and provided unbiased estimates and CIs with proper coverage rates. Likewise, the PL approach performed very well and provided results quite similar to those produced by the MH method. Except in the large negative ES scenario with extremely small baseline prevalences (i.e., $p = 0.0015$), where both methods reported large biases (median value > 0.75 for a $\log(\text{OR})$ of -1.5), estimates obtained by the two methods were good across all scenarios investigated. Both methods provided conservative CIs in case of extremely rare events. Finally, the proportion of converged runs indicated that PL was more computationally involved and could run into numerical issues, especially with very rare events.

Regarding the RR, results obtained by the MH and PL methods were comparable to those obtained for the OR (Table 3). Both methods provided unbiased estimates and valid CIs in almost all settings investigated. Again, the fact that MH discarded DZ studies did not alter the performance of this method.

For the RD, both methods included DZ studies and, thus, no study was excluded (Table 4). Similar to the OR and RR cases, biases obtained were small and coverage rates nominal. However, CIs provided by the PL approach were wider than those obtained with the MH method.

Table 2. Impact of double-zero studies for the estimation of the OR

<i>p</i> (in %)	Excluded studies (in %)		Median bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
10	0.02	0	0.001	0.001	95.20	95.18	0.64	0.64	100.00	98.97
0.70	43.69	0	- 0.015	- 0.022	97.09	97.26	2.50	2.50	95.69	80.77
0.35	64.26	0	0.145	0.125	96.33	96.40	3.28	3.28	79.19	73.67
0.15	79.83	0	0.747	0.782	96.81	96.78	4.45	4.40	46.35	44.97
Moderate negative ES										
10	0.00	0	- 0.001	- 0.000	95.24	95.20	0.47	0.47	100.00	99.21
0.70	34.42	0	- 0.005	- 0.016	95.91	95.92	1.75	1.76	99.96	80.68
0.35	57.62	0	- 0.018	- 0.024	97.59	97.74	2.53	2.54	98.50	87.21
0.15	76.71	0	0.090	0.090	99.50	99.63	3.92	3.92	79.27	76.05
Null ES										
10	0.00	0	0.002	0.001	95.04	95.04	0.41	0.41	100.00	99.29
0.70	27.05	0	- 0.005	- 0.005	96.01	96.20	1.52	1.52	100.00	80.80
0.35	50.78	0	0.001	- 0.000	96.91	96.87	2.19	2.19	99.76	84.93
0.15	73.02	0	0.001	0.000	99.71	99.68	3.40	3.40	89.73	85.30

Table 2. (Continued)

<i>p</i> (in %)	Excluded studies (in %)		Median bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
10	0.00	0	- 0.000	- 0.000	95.39	95.41	0.38	0.38	100.00	99.29
0.70	18.35	0	0.008	0.008	95.75	95.59	1.36	1.36	100.00	82.56
0.35	41.17	0	0.016	0.019	96.17	96.05	1.94	1.96	99.89	81.78
0.15	67.06	0	- 0.015	- 0.001	98.46	98.77	3.05	3.10	94.00	87.52
Large positive ES										
10	0.00	0	0.000	0.000	95.27	95.13	0.35	0.35	100.00	99.12
0.70	4.04	0	0.016	0.015	95.79	95.87	1.18	1.18	100.00	94.99
0.35	17.32	0	0.021	0.022	96.42	96.34	1.67	1.66	99.86	82.46
0.15	45.04	0	0.010	0.020	97.28	97.44	2.53	2.53	94.76	81.18

Note: *p* = mean baseline prevalence, ES = effect size, MH = Mantel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

Table 3. Impact of double-zero studies for the estimation of the RR

<i>p</i> (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
10	0.02	0	- 0.003	- 0.003	95.17	95.60	0.64	0.65	100.00	99.27
0.70	43.42	0	- 0.015	- 0.014	96.93	96.95	2.49	2.49	95.80	94.83
0.35	64.28	0	0.149	0.128	96.55	96.52	3.28	3.28	78.73	78.68
0.15	79.77	0	0.756	0.789	97.32	97.36	4.44	4.38	46.31	46.62
Moderate negative ES										
10	0.00	0	- 0.001	- 0.001	94.69	95.52	0.44	0.45	100.00	99.15
0.70	34.32	0	- 0.006	- 0.007	95.97	96.02	1.74	1.74	99.98	99.02
0.35	57.81	0	- 0.013	- 0.014	97.96	97.92	2.54	2.53	98.34	97.68
0.15	76.64	0	0.104	0.099	99.53	99.53	3.88	3.92	78.51	78.52
Null ES										
10	0.00	0	- 0.001	- 0.000	95.50	96.35	0.37	0.39	100.00	99.03
0.70	27.01	0	0.001	0.001	95.75	95.72	1.51	1.52	100.00	99.20
0.35	50.96	0	0.000	- 0.001	96.90	96.93	2.19	2.18	99.75	98.81
0.15	73.15	0	0.001	0.000	99.68	99.69	3.39	3.39	90.16	89.89

Table 3. (Continued)

<i>p</i> (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
10	0.00	0	0.002	0.002	94.93	96.37	0.33	0.35	100.00	98.65
0.70	18.27	0	0.012	0.011	95.17	95.20	1.34	1.35	100.00	99.62
0.35	41.30	0	0.002	0.001	96.63	96.57	1.93	1.93	99.87	99.06
0.15	67.08	0	-0.014	-0.004	98.44	98.54	3.06	3.06	94.12	93.77
Large positive ES										
10	0.00	0	0.000	-0.000	95.15	96.92	0.28	0.31	100.00	98.64
0.70	4.02	0	0.008	0.008	95.49	95.63	1.16	1.17	100.00	99.57
0.35	17.37	0	0.014	0.016	96.18	96.15	1.65	1.65	99.92	99.67
0.15	44.98	0	-0.002	-0.004	97.08	97.04	2.51	2.51	94.49	94.06

Note: *p* = mean baseline prevalence, ES = effect size, MH = Mantel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

Table 4. Impact of double-zero studies for the estimation of RD

p (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width (in %)		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
10	0	0	-0.010	-0.004	94.66	95.44	2.917	3.344	100.00	99.99
0.70	0	0	0.005	0.007	94.21	95.05	0.802	0.900	100.00	100.00
0.35	0	0	0.005	0.006	94.16	94.14	0.567	0.632	99.97	99.95
0.15	0	0	0.011	0.006	95.75	94.52	0.369	0.411	97.28	97.22
Moderate negative ES										
10	0	0	-0.013	-0.007	94.66	95.48	3.344	3.819	100.00	99.98
0.70	0	0	0.001	-0.002	94.70	95.52	0.919	1.037	100.00	100.00
0.35	0	0	-0.003	0.002	94.73	95.63	0.648	0.726	100.00	99.99
0.15	0	0	0.007	0.004	93.79	95.13	0.423	0.473	98.90	98.86
Null ES										
10	0	0	-0.021	-0.029	94.73	95.35	3.700	4.225	100.00	99.99
0.70	0	0	-0.002	-0.003	94.90	95.67	1.026	1.160	100.00	99.99
0.35	0	0	-0.001	0.001	94.92	96.02	0.722	0.815	100.00	100.00
0.15	0	0	-0.000	0.000	94.84	97.59	0.469	0.528	99.79	99.77

Table 4. (Continued)

p (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width (in %)		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
10	0	0	- 0.004	- 0.006	94.80	95.83	4.162	4.767	100.00	99.96
0.70	0	0	- 0.012	- 0.011	94.90	95.90	1.178	1.334	100.00	99.98
0.35	0	0	- 0.001	0.001	95.10	96.01	0.833	0.939	100.00	100.00
0.15	0	0	0.001	- 0.001	94.92	95.81	0.542	0.608	99.99	99.99
Large positive ES										
10	0	0	0.017	0.006	94.41	95.67	5.071	5.776	100.00	99.96
0.70	0	0	- 0.000	- 0.002	94.93	95.98	1.683	1.914	100.00	99.95
0.35	0	0	- 0.002	- 0.005	94.89	95.55	1.199	1.356	100.00	99.98
0.15	0	0	- 0.009	- 0.008	94.60	94.92	0.784	0.885	100.00	100.00

Note: p = mean baseline prevalence, ES = effect size, MH = Mantel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

3.2.2. Single-arm studies

Whatever the ES considered, the MH method excluded SA studies, whereas this was not the case with the PL approach.

Regarding the OR, both methods provided unbiased estimates and nominal coverage rates in all scenarios considered (Table 5). However, as the proportion of SA studies increased, the MH method became less and less efficient and CIs provided by this method were wider than those obtained using the PL approach, especially for large negative ES. Finally, SA studies created less numerical issues than DZ studies (the proportion of converged runs was always above 98%).

Results obtained for the RR were quite comparable to those obtained when estimating the OR (Table 6). In terms of bias, both methods provided virtually identical results and all values observed were below 2% (in absolute terms). Coverage rates were always nominal. Again, the MH method was less and less efficient as the proportion of SA studies increased and provided CIs much wider than the PL method.

Table 7 reports the results obtained by the MH and PL methods when estimating the RD. Median values of the bias were all below 0.05% in absolute terms. Regarding coverage rates, both methods provided valid CIs across all the scenarios, except when the proportion of SA studies was 75%, in which case the coverage rate of the PL's CI was below nominal. Regarding precision of the estimates, similar to what was observed for the OR and RR, the PL approach outperformed the MH method for increasing proportions of SA studies.

Table 5. Impact of single-arm studies for the estimation of OR

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
0	0.03	0	- 0.004	- 0.004	95.14	95.15	0.64	0.64	100.00	98.79
25	24.98	0	- 0.007	- 0.006	94.99	95.06	0.74	0.69	100.00	98.79
50	49.98	0	- 0.001	- 0.002	95.30	95.09	0.92	0.75	100.00	98.63
75	74.94	0	- 0.015	- 0.007	95.85	94.97	1.32	0.84	99.26	98.39
Moderate negative ES										
0	0.00	0	0.002	0.002	95.03	95.05	0.47	0.46	100.00	99.22
25	25.03	0	- 0.002	- 0.003	94.98	95.00	0.54	0.50	100.00	99.26
50	49.97	0	- 0.003	0.000	95.06	95.13	0.66	0.55	100.00	99.19
75	75.06	0	- 0.007	- 0.002	95.25	95.00	0.95	0.62	99.64	98.81
Null ES										
0	0.00	0	0.002	0.002	95.27	95.19	0.41	0.41	100.00	99.26
25	25.01	0	0.001	- 0.001	94.96	95.06	0.48	0.45	100.00	99.32
50	50.09	0	0.000	0.001	94.93	94.63	0.59	0.49	100.00	99.29
75	74.82	0	0.002	- 0.001	95.39	94.95	0.84	0.55	99.76	99.02

Table 5. (Continued)

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
0	0.00	0	0.001	0.001	94.85	94.94	0.38	0.38	100.00	99.29
25	24.76	0	0.003	0.001	94.94	94.57	0.44	0.41	100.00	99.24
50	50.11	0	0.001	0.002	94.96	95.03	0.54	0.45	100.00	99.42
75	74.77	0	-0.000	-0.000	95.41	95.41	0.77	0.51	99.65	99.15
Large positive ES										
0	0.00	0	0.003	0.003	95.01	94.99	0.35	0.35	100.00	99.29
25	24.89	0	-0.001	-0.001	95.18	94.95	0.40	0.38	100.00	99.32
50	49.86	0	-0.001	-0.001	95.37	94.89	0.49	0.41	100.00	99.31
75	74.94	0	0.004	0.002	94.55	94.47	0.70	0.47	99.76	99.25

Note: SA = proportion of single-arm studies, ES = effect size, MH = Mantel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

Table 6. Impact of single-arm studies for the estimation of RR

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
0	0.03	0	- 0.001	- 0.001	95.01	95.36	0.64	0.65	100.00	99.21
25	25.02	0	- 0.005	- 0.004	95.63	95.66	0.74	0.70	100.00	99.42
50	50.07	0	- 0.003	- 0.003	94.73	95.32	0.91	0.76	100.00	99.52
75	74.87	0	- 0.015	- 0.007	95.61	95.72	1.32	0.84	99.32	99.53
Moderate negative ES										
0	0.00	0	- 0.000	0.000	94.62	95.53	0.43	0.45	100.00	99.12
25	25.14	0	- 0.001	- 0.001	95.32	95.97	0.50	0.49	100.00	99.55
50	50.13	0	- 0.001	0.000	95.12	95.91	0.62	0.53	100.00	99.54
75	74.79	0	0.001	0.001	95.37	95.96	0.88	0.59	99.69	99.56
Null ES										
0	0.00	0	- 0.001	- 0.001	94.98	96.06	0.37	0.39	100.00	99.11
25	25.03	0	- 0.002	- 0.002	95.35	96.21	0.43	0.42	100.00	99.45
50	50.01	0	0.002	0.001	95.29	95.85	0.53	0.46	100.00	99.41
75	74.95	0	- 0.005	- 0.001	95.19	95.43	0.76	0.52	99.61	99.62

Table 6. (Continued)

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
0	0.00	0	0.000	0.001	95.50	96.89	0.33	0.35	100.00	99.07
25	24.95	0	0.001	0.001	94.66	96.01	0.38	0.38	100.00	99.18
50	49.94	0	0.000	0.001	94.96	96.27	0.47	0.42	100.00	99.35
75	74.97	0	0.001	-0.001	95.29	95.83	0.67	0.46	99.64	99.43
Large positive ES										
0	0.00	0	0.001	0.001	94.66	96.48	0.28	0.31	100.00	98.53
25	25.03	0	-0.001	-0.000	95.25	96.92	0.32	0.33	100.00	98.99
50	49.99	0	0.002	0.001	95.31	96.52	0.40	0.36	100.00	98.71
75	74.84	0	0.004	0.003	94.91	95.75	0.57	0.40	99.64	98.80

Note: SA = proportion of single-arm studies, ES = effect size, MH = antel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

Table 7. Impact of single-arm studies for the estimation of RD

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width (in %)		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Large negative ES										
0	0.00	0	0.004	0.009	94.83	95.66	0.029	0.033	100.00	99.98
25	25.03	0	0.017	0.007	94.69	94.86	0.034	0.036	100.00	99.96
50	50.08	0	0.006	0.013	95.11	94.05	0.041	0.039	100.00	99.89
75	74.86	0	0.008	0.035	94.88	90.92	0.059	0.043	99.66	98.89
Moderate negative ES										
0	0.00	0	0.001	-0.011	95.34	95.66	0.033	0.038	100.00	99.99
25	25.11	0	-0.001	-0.011	94.81	94.90	0.039	0.041	100.00	99.96
50	50.11	0	-0.033	-0.014	94.92	94.07	0.047	0.044	100.00	99.95
75	74.87	0	0.040	0.029	94.71	91.56	0.067	0.049	99.65	98.94
Null ES										
0	0.00	0	-0.012	-0.006	95.01	95.53	0.037	0.042	100.00	99.97
25	24.93	0	-0.003	-0.017	95.22	95.60	0.043	0.045	100.00	99.96
50	49.90	0	-0.006	0.009	94.84	94.95	0.052	0.049	100.00	99.95
75	74.93	0	-0.005	0.001	94.62	90.65	0.075	0.053	99.65	99.07

Table 7. (Continued)

SA (in %)	Excluded studies (in %)		Median Bias		Coverage rate (in %)		Median CI's width (in %)		Converged runs (in %)	
	MH	PL	MH	PL	MH	PL	MH	PL	MH	PL
Moderate positive ES										
0	0.00	0	-0.004	-0.023	95.01	95.70	0.042	0.048	100.00	100.00
25	24.90	0	-0.011	-0.031	94.63	95.38	0.048	0.051	100.00	99.98
50	50.21	0	-0.020	-0.007	94.83	94.47	0.059	0.055	100.00	99.88
75	74.81	0	0.038	0.011	95.13	91.17	0.084	0.060	99.76	98.77
Large positive ES										
0	0.00	0	0.006	0.016	94.87	95.64	0.051	0.058	100.00	99.95
25	24.99	0	0.017	0.026	94.84	95.63	0.058	0.062	100.00	99.87
50	50.11	0	-0.008	-0.021	94.62	94.23	0.072	0.066	100.00	99.91
75	74.89	0	0.007	0.006	94.60	91.35	0.102	0.072	99.62	98.92

Note: SA = proportion of single-arm studies, ES = effect size, MH = antel-Haenszel method, PL = pseudo-likelihood method, CI = confidence interval. See Table 1 for the ES' values in the different scenarios.

4. Discussion

The MH method has been shown to perform extremely well under the assumption of a homogeneous treatment effect (Bradburn et al. [2]; Piaget-Rossel and Taffé [13]). It involves simple computations, can be applied to compute the three classical ESs, (i.e., OR, RR, and RD) and is readily available in most of the statistical packages. However, DZ and SA studies are excluded from the computation of the combined ES estimate. In this paper, we have developed a novel approach based on the formulation of a PL, which allows one to include both SA and DZ studies into the meta-analysis. Using simulations, we compared the performance of this PL approach to that of the MH method, in settings with increasing proportion of SA and DZ studies.

Our proposed PL method performed very well for all three ESs. For the RD, we found that the calibrated Wald's CI computed using the 98.5th quantile of the standard normal distribution provided nominal coverage rates, except in the settings with 75% of SA studies. This shows that using the normal distribution as a working distribution for the baseline prevalences does not impact the performance of this method. In additional simulations, we found that using an asymmetrical beta distribution to generate the baseline prevalences – instead of the uniform distribution described in Subsection 3.1 – did not alter this conclusion (results not shown). This finding challenges Dias and Ades's statement [3] that “unless the baseline model is correctly specified, the relative effect estimates will be biased”.

We found that both the MH and PL methods provided reliable results, whatever the proportion of DZ studies. Biases, coverage rates, and CIs' width provided by these two methods were quite similar. The only noticeable difference was that the PL's calibrated CIs for the RD were wider than the CIs obtained with the MH method. These results suggest that under the assumption of a homogeneous treatment effect, DZ studies do not contain relevant information for the meta-analysis.

This is quite unexpected given that, for instance, a DZ study of size 150 does seem to convey more information regarding the low probability of an event compared to a DZ study of size 50. Moreover, Friedrich et al. [5] argued that deleting DZ studies in balanced trials might bias the treatment effect away from the null. Nevertheless, from our results, we conclude that MH is a valid FE meta-analysis method, even in the presence of DZ studies. The PL method is a good alternative, but it has the disadvantage of being more computationally-involved and may run into numerical issues, especially when the proportion of DZ studies is high. When the meta-analysis includes only DZ studies, none of the methods work.

Regarding SA studies, results suggested that they contained more relevant information than DZ studies. Indeed, whereas bias and coverage rates obtained by the PL and MH methods were found to be similar, the latter provided CIs wider as the proportion of SA studies increased, suggesting a loss of precision related to the non-inclusion of the information contained in SA studies. Based on these results, PL should be favored in the presence of SA studies. With 100% SA studies, the MH method breaks down, whereas we found that the PL approach still performed very well (additional simulations; results not shown).

In additional simulations, we compared our PL approach to the beta-binomial model discussed by Kuss [8], which also allows including SA and DZ studies. We found that both methods performed similarly, whatever the proportion of SA and DZ studies and the ES considered (results not shown). However, the Beta-Binomial model encountered more numerical issues (e.g., the number of converged runs when estimating the OR was systematically below 90%).

To sum up, in settings with DZ studies, we recommend using the MH method, although this method exclude the information contained in these studies. In settings with SA studies, we recommend using the PL approach, which was shown to be more efficient. However, when the ES

of interest is the RD and the proportion of SA studies is very high (75% or more), the meta-analyst should be aware that the CIs computed using the PL approach may have coverage rates slightly below nominal.

The main limitation to these recommendations is that they only apply to the framework of homogeneous treatment effects. In practice, there are many situations where this assumption is likely to be violated (Kontopantelis et al. [9]) and the reader must keep in mind that the MH method is not valid when treatment effects are heterogeneous (Kuss [8]). As for the PL approach, it can easily be adapted to account for heterogeneity in treatment effects by including regressors (meta-regression). Moreover, our simulations did not cover all the possible settings. For instance, we did not consider the case of study's scarcity (i.e., small K), which can impact the methods' performance.

In a future research we will focus on adapting the PL approach to the framework of heterogeneous treatment effects. It would also be worth seeking alternatives to our calibrated CIs for the RD, such as the use of the profile likelihood method (Böhning et al. [1]).

Acknowledgement

The work of RPR was supported by the Swiss National Science Foundation (Grant Number: 200021_169504/1).

References

- [1] D. Böhning, R. Kuhnert and S. Rattanasiri, *Meta-Analysis of Binary Data Using Profile Likelihood*, CRC Press, 2008.
- [2] M. J. Bradburn, J. J. Deeks, J. A. Berlin and A. Russell Localio, Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events, *Statistics in Medicine* 26(1) (2007), 53-77.
DOI: <https://doi.org/10.1002/sim.2528>
- [3] S. Dias and A. E. Ades, Absolute or relative effects? Arm-based synthesis of trial data, *Research Synthesis Methods* 7(1) (2015), 23-28.
DOI: <https://doi.org/10.1002/jrsm.1184>

- [4] O. Efthimiou, Practical guide to the meta-analysis of rare events, *Evidence-Based Mental Health* 21(2) (2018), 72-76.
DOI: <http://dx.doi.org/10.1136/eb-2018-102911>
- [5] J. O. Friedrich, N. K. Adhikari and J. Beyene, Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data, *BMC Medical Research Methodology* 7 (2007); Article 5.
DOI: <https://doi.org/10.1186/1471-2288-7-5>
- [6] F. Keus, J. Wetterslev, C. Gluud, H. G. Gooszen and C. J. van Laarhoven, Robustness assessments are needed to reduce bias in meta-analyses that include zero-event randomized trials, *American Journal of Gastroenterology* 104(3) (2009), 546-551.
- [7] O. Kuss, J. F. Gummert and J. Borgermann, Meta-analysis with rare events should use adequate methods, *The Journal of Thoracic and Cardiovascular Surgery* 136(1) (2008), 241.
- [8] O. Kuss, Statistical methods for meta-analyses including information from studies without any events – add nothing to nothing and succeed nevertheless, *Statistics in Medicine* 34(7) (2015), 1097-1116.
DOI: <https://doi.org/10.1002/sim.6383>
- [9] E. Kontopantelis, D. A. Springate and D. Reeves, A re-analysis of the Cochrane library data: The dangers of unobserved heterogeneity in meta-analyses, *Plos One* 8(7) (2013), e69930.
DOI: <https://doi.org/10.1371/journal.pone.0069930>
- [10] I. C. Marschner and A. C. Gillett, Relative risk regression: Reliable and flexible methods for log-binomial models, *Biostatistics* 13(1) (2012), 179-192.
DOI: <https://doi.org/10.1093/biostatistics/kxr030>
- [11] D. Moher, J. Tetzlaff, A. A. Tricco, M. Sampson and D. G. Altman, Epidemiology and reporting characteristics of systematic reviews, *PLoS Medicine* 4(3) (2007), e78.
DOI: <https://doi.org/10.1371/journal.pmed.0040078>
- [12] S. L. T. Normand, Meta-analysis: Formulating, evaluating, combining, and reporting, *Statistics in Medicine* 18(3) (1999), 321-359.
DOI: [https://doi.org/10.1002/\(SICI\)1097-0258\(19990215\)18:3<321::AID-SIM28>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P)
- [13] R. Piaget-Rossel and P. Taffé, Meta-analysis of rare events under the assumption of a homogeneous treatment effect, Accepted for publication in *Biometrical Journal* (2019).
- [14] C. Röver, F. Knapp and T. Friede, Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies, *BMC Medical Research Methodology* 15 (2015); Article 99.
DOI: <https://doi.org/10.1186/s12874-015-0091-1>

- [15] M. J. Sweeting, A. J. Sutton and P. C. Lambert, What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data, *Statistics in Medicine* 23(9) (2004), 1351-1375.
DOI: <https://doi.org/10.1002/sim.1761>
- [16] A. Toussaint, Width of Margins in Phyllodes Tumors of the Breast: The Controversy Drags on: A Systematic Review, *Diplôme inter-Universitaire de Sénologie et Pathologie Mammaire*, François-Rabelais University and Rennes University, Tours and Rennes, 2017.
- [17] H. White, Maximum likelihood estimation of misspecified models, *Econometrica* 50(1) (1982), 1-26.

