An Exact Test for Meta-analysis with Binary Endpoints

O. Kuss[1], **C. Gromann**[2]
[1]Institute of Medical Epidemiology, Biostatistics, and Informatics, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany
[2]Coordinating Centre for Clinical Studies, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

Summary

Objectives: We reintroduce an exact Mantel-Haenszel (MH) procedure for meta-analysis with binary endpoints which is expected to work especially well in sparse data, e.g., in meta-analyses of safety or adverse events.
Methods: The performance of the exact MH procedure in terms of empirical size and power is compared to the asymptotic MH and to the two standard procedures (fixed effects and random effects model) in a simulation study. We illustrate the methods with a meta-analysis of postoperative stroke occurrence after off-pump or on-pump surgery in coronary artery bypass grafting.
Results: We find that in almost all situations the asymptotic MH procedure outperforms its competitors; especially the standard methods yield poor results in terms of power and size.
Conclusions: There is no need to use the reintroduced exact MH procedure; the asymptotic MH procedure will be sufficient in most practical situations. The standard methods (fixed effects and random effects model) should not be used in the sparse data situation.

Keywords

Meta-analysis, odds ratio, logistic models, exact test, SAS®

Methods Inf Med 2007; 46: 662–668

doi:10.3414/ME0422

Introduction

Meta-analysis is the quantitative, systematic summary of a collection of separate studies for the purpose of obtaining information that cannot be derived from any of the studies alone [1]. It is a method of considerable and still growing methodical interest in medicine. This is due to meta-analyses frequently being parts of the popular systematic reviews which enable clinicians, researchers, and policy makers a quick but comprehensive overview of clinical evidence on a specific topic. Not every systematic review contains a meta-analysis, sometimes the individual studies are judged to be too heterogeneous to be summarized quantitatively. However, if it is decided to perform a meta-analysis, the common treatment effect is generally assessed with the fixed effects (FEM) or the random effects (REM) model [2, 3]. Both models comprise a two-step method where the treatment effects from the individual studies along with their variances are estimated in the first step. In the second step the common treatment effect is estimated as a weighted mean of the individual studies' effects, using their inverse variances as weights. In the FEM an identical (homogeneous) true treatment effect is assumed across the different studies, in the REM model these true treatment effects are allowed to vary across studies, according to a normal distribution.

The weaknesses of these two standard approaches are well known. For example, statistical inference in both models relies on the fact that the individual studies' variances (and in the REM also the between-study variance) are fixed and known where they actually have to be estimated. Second, in the case of binary endpoints which we will concentrate on in the following, there is still controversy how to deal with empty cells which in some cases prevent from estimating odds ratios or relative risks in the respective study. Third, both models rely on normal distribution assumptions for treatment effects in the individual studies (FEM and REM) and the between-study variance (REM). Finally, the estimation uncertainty of the first step is not accounted for in the second step. According to these deficiencies, a number of proposals have been made to improve on the two standard procedures [4-9].

One partial solution to the problems mentioned above, known for a fairly long time, is the Mantel-Haenszel (MH) procedure [10]. This does not rely on the described two-step procedure of the standard methods and can also be calculated with studies that contain zero cells. However, similar to the FEM, the MH procedure also assumes a homogeneous treatment effect across the individual studies.

It has been largely undetected that there is also an exact version of the Mantel-Haenszel procedure which is expected to give valid results even in the case of very sparse data. The two standard references for meta-analysis in medicine, the books of Sutton ([2], p. 69), and Whitehead ([3], p. 220), only mention the procedure and refer to the paper of Emerson [11] without any further comment.

Of course, we cannot expect an exact procedure to solve all existing problems in meta-analysis. On the contrary, exact methods are necessarily conservative, keeping strictly the pre-specified level but rarely making full use of it [12]. Moreover, there

Received: April 24, 2006; accepted: November 23, 2006

are still considerable requirements on computing power and time.

It is obvious that the natural domain of an exact method for meta-analysis would be in situations with sparse data, e.g., if we are interested in safety or adverse events [13]. In non-sparse applications we expect the standard methods and the asymptotic MH procedure to be sufficiently valid and computational costs might make the use of exact methods impossible.

In the following we reintroduce the exact MH procedure with some more mathematical rigor, comment on its computation, and show the results of a simulation study which compares the two standard methods (FEM, REM), the standard (asymptotic) MH, and the exact MH procedure in a variety of situations. The methods are illustrated by a recently published meta-analysis on the effect of off-pump surgery on postoperative stroke occurrence in coronary artery bypass grafting.

Table 1 a) Empirical levels from two-sided tests for the four different test procedures under the null hypothesis of no treatment effect (OR = 1), depending on the underlying true model, the number of patients per treatment arm, and the number of studies. Results are based on 10,000 simulation runs.

Number of patients	Number of studies	FEM	REM	Asymptotic MH	Exact MH
Fixed Effects					
20	5	0.015	0.014	0.051	0.035
20	10	0.014	0.013	0.048	0.038
20	20	0.013	0.013	0.047	0.042
50	5	0.026	0.023	0.050	0.041
50	10	0.027	0.023	0.050	0.044
50	20	0.026	0.024	0.048	0.045
Random Effects, $\tau = 0.05$					
20	5	0.026	0.024	0.074	0.053
20	10	0.038	0.037	0.100	0.084
20	20	0.074	0.073	0.154	0.139
50	5	0.076	0.068	0.118	0.102
50	10	0.126	0.116	0.179	0.168
50	20	0.218	0.209	0.284	0.273
Random Effects, $\tau = 0.10$					
20	5	0.040	0.038	0.103	0.079
20	10	0.073	0.070	0.150	0.133
20	20	0.141	0.138	0.236	0.221
50	5	0.127	0.113	0.179	0.161
50	10	0.209	0.195	0.271	0.256
50	20	0.329	0.320	0.398	0.389

An Exact Test for Meta-analysis with Binary Endpoints

We consider the following situation. Given are K (k = 1, …, K) independent studies to compare two treatments A and B. The outcome is binary (Yes/No). Thus, the results from a single study can be displayed in a 2 × 2 table (to enhance readability we initially omit the study index k):

		Outcome: Yes	Outcome: No	
Treatment	A	n_{11}	n_{12}	n_{1+}
	B	n_{21}	n_{22}	n_{2+}
		n_{+1}	n_{+2}	n_{++}

To measure the treatment effect we use the odds ratio (OR) $p_A(1 - p_B)/p_B(1 - p_A)$, where p_A, p_B are the outcome probabilities for treatments A, B.

Under the assumption that the row and column sums ($n_{1+}, n_{2+}, n_{+1}, n_{+2}$) are fixed (that is, conditional on these parameters), n_{11} is known to have a hypergeometric distribution with parameters n_{++}, n_{1+}, and n_{+1}. In the case of no treatment effect (OR = 1) the expectation and variance of this distribution are given by $E(n_{11}) = n_{1+}n_{+1}/n_{++}$ and $Var(n_{11}) = n_{1+}n_{2+}n_{+1}n_{+2}/n_{++}^2(n_{++} - 1)$. This yields an asymptotical test for no treatment effect in a single 2 × 2 table with test statistic $\chi^2_{MH} = (n_{11} - E(n_{11}))^2/Var(n_{11})$ and an asymptotical χ^2-distribution with 1 df under the null hypothesis. It can be shown that $\chi^2_{MH} = (n_{++} - 1)/n_{++} * \chi^2$, where χ^2 is the common Chi-square statistic in a 2 × 2 table. An exact p-value can be calculated by summing up the probabilities of all possible 2 × 2 tables (with the same margins as the actually observed table) which lead to the same or a more extreme value of n_{11}. This is, of course, the one-sided Fisher test.

This principle can be easily generalized to the meta-analytic situation. Under the assumption of no treatment effect in all K studies ($OR_1 = \ldots = OR_K = 1$) we use the test statistic S which is the sum over all n_{11k} from the individual studies ($S = \Sigma_k n_{11k}$). Using the independence between studies the moments for S easily generalize to $E(S) = \Sigma_k E(n_{11k}) = \Sigma_k n_{1+k}n_{+1k}/n_{++k}$ and $Var(S) = \Sigma_k Var(n_{11k}) = \Sigma_k n_{1+k}n_{2+k}n_{+1k}n_{+2k}/n_{++k}^2(n_{++k} - 1)$. The parallel asymptotic test with test statistic $\chi^2_{MH} = (S - E(S))^2/Var(S)$ with an asymptotical χ^2-distribution with 1 df under the null hypothesis is the standard Mantel-Haenzel test. An exact p-value now can be calculated by summing up the probabilities of all possible meta-analyses (with the same margins as the actually observed meta-analysis) which lead to the same or a more extreme value of S. This is again a one-sided test.

As, in general, there is a huge number of possible meta-analyses with identical margins as the observed one, it is essential that computational methods avoid the explicit enumeration of those. The most easily ac-

Table 1 b) Empirical levels from two-sided tests for the four different test procedures under the alternative hypothesis of a moderate treatment effect (OR = 0.666), depending on the underlying true model, the number of patients per treatment arm, and the number of studies. Results are based on 10,000 simulation runs.

Number of patients	Number of studies	FEM	REM	Asymptotic MH	Exact MH
		Fixed Effects			
20	5	0.044	0.042	0.125	0.095
20	10	0.082	0.078	0.194	0.170
20	20	0.166	0.163	0.347	0.320
50	5	0.146	0.132	0.235	0.209
50	10	0.299	0.281	0.425	0.403
50	20	0.568	0.554	0.700	0.687
		Random Effects, $\tau=0.05$			
20	5	0.047	0.045	0.139	0.105
20	10	0.109	0.105	0.231	0.204
20	20	0.227	0.222	0.388	0.365
50	5	0.184	0.170	0.275	0.248
50	10	0.329	0.316	0.433	0.414
50	20	0.546	0.537	0.633	0.623
		Random Effects, $\tau=0.10$			
20	5	0.058	0.054	0.157	0.120
20	10	0.127	0.122	0.250	0.223
20	20	0.252	0.246	0.410	0.387
50	5	0.207	0.192	0.297	0.273
50	10	0.364	0.351	0.464	0.445
50	20	0.548	0.540	0.622	0.614

cessible method to this task uses the principle of testing for conditional independence in a $2 \times 2 \times K$ table. This is possible as a $2 \times 2 \times K$ table is equivalent to a meta-analysis as we have described it. Conditional inference is based on the distribution of the sufficient statistic for the interesting parameter, conditional on the sufficient statistics for all other parameters (nuisance parameters) in the model. It turns out [12] that S is the sufficient statistic for the treatment effect in our case, given the row and column totals (n_{1+k}, n_{2+k}, n_{+1k}, n_{+2k}) in all K studies. It was further shown that under the assumption of homogeneous odds ratios this test is UMPU [14]. Efficient algorithms for computing this distribution were given by Vollset et al. [15] and are implemented in SAS® PROC FREQ and StatXact®. As the $2 \times 2 \times K$ table is a special case of a logistic regression model we could also use software for exact logistic regression like SAS® PROC LOGISTIC or LogXact®. SAS® code for fitting the exact model with PROC LOGISTIC is given in the appendix.

A Simulation Study

There is very limited information on the performance of the exact MH method from simulation studies. Sankey et al. [16], Deeks et al. [17], and Sweeting et al. [18] performed simulations in the sparse data situation, but only compared asymptotic methods. To our knowledge, the only study that assessed the exact method is due to Mehta and Walsh [19]. They compared the exact, a mid-p, and the asymptotic MH procedure for estimating odds ratios across several 2×2 tables, where the mid-p-value is the exact null probability of more extreme results plus half the exact null probability of the observed event [20]. Mehta and Walsh found the exact method necessarily to be conservative whereas the mid-p method preserved coverage. The asymptotic MH method yielded no results in some extreme situations with very sparse data and large odds ratios, a hybrid MH method with exact confidence intervals when the asymptotic method failed gave satisfactory results. However, Mehta and Walsh only considered homogeneous (or: fixed effects) odds ratios, that is, they only investigated situations where the assumptions of the two MH methods were actually fulfilled.

As we wanted to check the performance, or more precisely, the robustness of these methods also in the heterogeneous (or: random effects) odds ratio situation, we set up our own simulation study. For each parameter setting we generated 10,000 meta-analyses under the null hypothesis of no treatment effect (OR = 1) as well as under the alternative of OR = 0.666 and OR = 0.333. The baseline (or control group) event probability was kept fixed to 0.1 as we aimed to assess the performance of the different methods in the sparse data situation. We further varied the underlying model (fixed effects or random effects, where in the random effects situation the between study variance was set to 0.05 or 0.1, respectively, on the LogOR scale), the number of patients within the individual studies (20 or 50 per treatment arm), and the number of studies in the respective meta-analysis (5, 10, or 20). Cross-classification of these factors (treatment effect, true underlying model, number of patients per treatment arm, and number of studies) resulted in 54 ($= 3 \times 3 \times 3 \times 2$) different parameter settings.

The simulation was programmed in SAS® 9.1.3. Tests were calculated in PROC FREQ (FEM and asymptotic MH), PROC MIXED (REM [21]), and PROC LOGISTIC (exact MH). FEM and REM estimates were calculated according to the standard approach (see [2], p. 58 for the FEM, and p. 74f for the REM (inverse weighted variance method)), the between-study variance τ^2 was estimated by the methods of moments (PROC GLM, see [3], p. 91). In case of an empty cell in an individual trial, 0.5 was added to all four cells in the respective trial for the computation of the FEM and the REM. No such correction is necessary for the asymptotic and the exact MH approach. We decided to use the 0.5-correction as we feel this is still considered the standard procedure, which is, for example, also implemented in SAS® PROC FREQ.

All tests were two-sided, in the exact MH approach we used mid-p values to judge statistical significance. In Tables 1a-c we report the empirical levels for the four approaches in the described situations.

Table 1 c) Empirical levels from two-sided tests for the four different test procedures under the alternative hypothesis of a strong treatment effect (OR = 0.333), depending on the underlying true model, the number of patients per treatment arm, and the number of studies. Results are based on 10,000 simulation runs.

Number of patients	Number of studies	FEM	REM	Asymptotic MH	Exact MH
		Fixed Effects			
20	5	0.207	0.193	0.441	0.377
20	10	0.498	0.483	0.745	0.712
20	20	0.861	0.851	0.961	0.954
50	5	0.683	0.652	0.833	0.806
50	10	0.948	0.942	0.984	0.982
50	20	0.999	0.999	1.000	1.000
		Random Effects, $\tau=0.05$			
20	5	0.213	0.199	0.438	0.372
20	10	0.488	0.471	0.716	0.684
20	20	0.816	0.808	0.930	0.921
50	5	0.658	0.629	0.801	0.777
50	10	0.920	0.913	0.964	0.961
50	20	0.992	0.991	0.997	0.996
		Random Effects, $\tau=0.10$			
20	5	0.229	0.214	0.446	0.389
20	10	0.480	0.464	0.701	0.668
20	20	0.784	0.775	0.898	0.889
50	5	0.637	0.610	0.767	0.741
50	10	0.884	0.876	0.937	0.931
50	20	0.976	0.975	0.986	0.985

Results

Under the null hypothesis of no treatment effect (Table 1a) we find that in the fixed effects model only the asymptotic MH procedure keeps the nominal level of 0.05. The exact MH procedure is, as expected, slightly conservative. The two standard procedures show a very conservative behavior, their empirical levels always lying below 0.03. Under the random effects model all procedures exceed the nominal level in most of the situations. The larger the number of patients in the individual studies and the larger the number of studies in the meta-analysis, the larger is the deviation from the nominal 5%. Note that also the REM procedure, which is the only procedure that should be able to deal with random effects, performs very poorly.

Under the alternative hypothesis of a treatment effect (Tables 1b and 1c) we find the asymptotic MH procedure always to have the best power, followed by the exact procedure. The power of the two standard procedures (FEM and REM) to detect a treatment effect is always considerably lower.

An Example: Off-pump versus On-pump Surgery in Coronary Artery Bypass Grafting

To illustrate the described procedures with a real life data set we consider an example from a recent meta-analysis of the risks of off-pump and on-pump surgery in patients undergoing coronary artery bypass grafting (CABG) [22]. Table 2 gives the data from 21 studies on the effect of off-pump versus on-pump surgery on postoperative stroke occurrence. We note that only a small number of events ($n_{+1+} = 20$) was observed in total, in 11 studies there wasn't even a single outcome observed. The studies are very homogeneous, the standard Cochran Q-test for homogeneity gives a p-value of p = 0.999.

Table 3 gives the results from a two-sided test on the common (meta-analytic) treatment effect for the off-pump data. The results mirror the evidence seen in the simulation, the exact and the asymptotic MH procedure point more strongly to a possible treatment effect than the standard methods. Also given in Table 3 are estimated odds ratios for the common effect of off-pump surgery with corresponding 95% confidence intervals. We note that the exact and the asymptotic Mantel-Haenszel methods yield very similar estimates for the common odds ratio which deviate from the FEM and REM estimates. Surprisingly, the FEM and REM estimates do not coincide in this situation, although we observed very homogeneous treatment effects. This is due to a peculiarity of PROC FREQ, the procedure by default removes all studies with no events from computations.

As a sensitivity check we calculated the treatment effect with two additional methods: 1) from a simple fourfold-table with the data collapsed over all studies, and 2) from a random effects logistic regression model. This latter method can be used in our situ-

Table 2 Data from 21 studies on the effect of off-pump surgery in coronary artery bypass grafting on postoperative stroke occurrence [22]. Given are the number of events (postoperative strokes) and the number of observations (patients) in the respective treatment arm and study. For the citations of the individual studies we refer to the original paper.

Study	Off-Pump		On-Pump	
	Number of events	Number of observations	Number of events	Number of observations
BHACAS 1	0	100	0	100
BHACAS 2	0	100	0	101
Carrier	0	28	1	37
Czerny	0	14	0	20
Diegeler	1	20	1	19
Gerola	0	80	0	80
Khan	0	54	0	49
Legare	2	150	0	150
Lingaas	0	60	2	60
Lonn	0	15	0	16
Matata	0	10	0	10
Motellabzk	0	15	1	20
Muneretto	0	88	2	88
OctoPump	1	142	2	139
OctoStent	0	136	0	131
PRAGUE-4	0	204	2	184
Puskas	1	98	2	99
Sahlman	1	24	1	26
Vural	0	25	0	25
Wandschneider	0	41	0	67
Wehlin	0	21	0	16

ation of a binary response as the available data can be interpreted to arise from a meta-analysis of original or individual data, the outcome and the treatment of each single patient being readily identifiable. This considerably enlarges the possible meta-analytic models and estimation methods but also poses new statistical problems of analysis and interpretation [23]. The results were (p-value; OR (95%-CI)) 0.075; 0.429 (0.164; 1.119) for the collapsed table, and 0.041; 0.429 (0.017; 0.842) for the random effects logistic model.

We note that these additional results are very similar to the asymptotic and the exact MH results and we judge both standard procedures (FEM and REM) to show a somewhat deviant behavior. In the original report, Cheng reported the results from the fixed effects model which might give a slight underestimation of the treatment effect.

Discussion

The simulation study largely supports the recommendations which were already given by Emerson [11] in 1994, and were later confirmed by others [16-18]. In meta-analyses for binary endpoints when the odds ratio is chosen to describe the treatment effect the standard fixed effects and random effects models should not be used, at least with moderate to small sample sizes in the individual studies. The asymptotic (or standard) Mantel-Haenszel procedure should be used instead. In principle, this fact has already been known since the theoretical work of Breslow in 1981 [24].

Our own results add two important points: First, it is not necessary to use the reintroduced exact MH procedure; the asymptotic procedure will be sufficient in most practical situations. Second, in the light of the severe anticonservatism of all four methods in the random effects situation, none of the methods should be used if there is heterogeneity between the studies. This confirms the results of Sankey et al. [16] which also found anticonservative behavior for the standard methods as well as the asymptotic MH procedure in the heterogeneous sparse data situation.

Which method should be used instead is still an open question. We believe that the equivalence between meta-analyses of binary data and the logistic regression model for clustered responses (where the individual studies represent the clusters) [23, 25, 26] is not yet made full use of.

A reviewer pointed to the fact that it is controversial to use the situation of no treatment effect (OR = 1), but treatment heterogeneity ($\tau = 0.05$ or $\tau = 0.10$) in the simulation. Following Senn [27], the assumption of no heterogeneity actually belongs to the null hypothesis of no treatment. As such the results from the lower two panels of Table 1a should actually be interpreted as empirical levels under a different hypothesis of no treatment and no heterogeneity and not as those under the standard null hypothesis of no treatment. It seems that this subtlety has been overlooked in today's applications of meta-analysis; other simulation studies have also used this design.

An important advantage of both MH procedures concerns zero cells which seem to be more the rule than the exception in meta-analyses for safety or adverse event data. We appreciate the extensive work on continuity corrections for zero cells of Sankey et al. [16] and Sweeting et al. [18], but we share Sutton et al.'s opinion [13] who said that "… our preference is for the use of statistical methods that do not require continuity correction factors to combine data …". Our off-pump example emphasizes this point when 17 pseudo events have to be added to 20 actually observed events to achieve well-defined FEM and REM estimates.

The advantages of the MH procedures, however, come with a prize. The underlying

Table 3 Results from four different meta-analytic methods on the off-pump data. Given are (column 2) the p-value of a two-sided test for the common treatment effect and (column 3) an estimated odds ratio with corresponding 95% confidence interval for the common meta-analytic treatment effect. The FEM and the asymptotic MH procedure were calculated in SAS® PROC FREQ, the REM in PROC MIXED and PROC GLM, and the exact MH procedure in PROC LOGISTIC. P-values for the exact MH procedure are mid-p-values.

		Off-pump vs. On-Pump
Method	p-value	Odds ratio [95% confidence interval]
FEM	0.168	0.535 [0.216; 1.328]
REM	0.286	0.672 [0.327; 1.390]
Asymptotic MH	0.081	0.432 [0.165; 1.135]
Exact MH	0.094	0.435 [0.136; 1.213]

conditioning principle unfortunately prevents from additional modelling of study-specific covariates, that is, meta-regression becomes impossible. Moreover, focussing on the exact procedure, there are still large demands on computing time and power.

Finally, our simulation study also has limitations: We only considered meta-analyses with binary endpoints and with the odds ratio as the chosen measure for the treatment effect. There are also asymptotic MH procedures for the various other effects measured in 2×2 tables (relative risks or risk differences), but we are not aware of corresponding exact versions. However, for theoretical reasons [28] we also expect the FEM and REM estimators in these cases to be inferior to the respective MH methods. For the relative risk, this was already confirmed by simulations [29].

Appendix

SAS® code for performing the exact MH analysis in PROC LOGISTIC for the off-pump data. The first DATA STEP reads in the data, and the second unfolds the data set, so that every individual patient has an own line in the data set.

PROC LOGISTIC uses the multivariate shift algorithm of Hirji et al. [30].

```
DATA offpump;
     INPUT study treatment event n @@;
     DATALINES;
     1 1  0 100   1 0  0 100
     2 1  0 100   2 0  0 101
     3 1  0  28   3 0  1  37
     4 1  0  14   4 0  0  16
     5 1  1  20   5 0  1  20
     6 1  0  80   6 0  0  80
     7 1  0  54   7 0  0  49
     8 1  2 150   8 0  0 150
     9 1  0  60   9 0  2  60
    10 1  0  15  10 0  0  16
    11 1  0  10  11 0  0  10
    12 1  0  15  12 0  1  20
    13 1  0  88  13 0  2  88
    14 1  1 142  14 0  2 139
    15 1  0 136  15 0  0 131
    16 1  0 204  16 0  2 184
    17 1  1  98  17 0  2  99
    18 1  1  24  18 0  1  26
    19 1  0  25  19 0  0  25
    20 1  0  41  20 0  0  67
    21 1  0  21  21 0  0  16
;RUN;

DATA offpump_individual(DROP=i);
     SET offpump;
     DO i=1 TO n;
        IF i<=event THEN stroke=1;
        IF i> event THEN stroke=0;
        OUTPUT;
     END;
RUN;

PROC LOGISTIC DATA=offpump_individual DESCENDING EXACTONLY;
     CLASS study;
     MODEL stroke = study treatment;
     EXACT treatment / ESTIMATE=ODDS;
RUN;
```

References

1. Boissel JP, Sacks HS, Leizorovicz A, Blanchard J, Panak E, Peyrieux JC. Meta-analysis of clinical trials: summary of an international conference. Eur J Clin Pharmacol 1988; 34 (6): 535-538.
2. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-Analysis in Medical Research. Chichester: Wiley & Sons; 2000.
3. Whitehead A. Meta-analysis of controlled clinical trials. Chichester: Wiley & Sons; 2002.
4. Ziegler S, Koch A, Victor N. Deficits and remedy of the standard random effects methods in meta-analysis. Methods Inf Med 2001; 40 (2): 148-155.
5. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Stat Med 2001; 20 (24): 3875-3889.
6. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Stat Med 1997; 16 (7): 753-768.
7. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. Biometrics 1999; 55 (3): 732-737.
8. Knapp G, Hartung J. Combined test procedures in the meta-analysis of controlled clinical trials. Stud Health Technol Inform 2000; 77: 34-38.
9. Böhning D. Meta-analysis – A unifying meta-likelihood approach framing unobserved heterogeneity, study covariates, publication bias, and study quality. Methods Inf Med 2005; 44 (1): 127-135.
10. Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. J Natl Cancer I 1959; 22 (4): 719-748.
11. Emerson JD. Combining estimates of the odds ratio: the state of the art. Stat Meth Med Res 1994; 3 (2): 157-178.
12. Agresti A. A survey of exact inference for contingency tables (with discussion). Stat Sci 1992; 7 (1): 131-177.
13. Sutton AJ, Cooper NJ, Lambert PC, Jones DR, Abrams KR, Sweeting MJ. Meta-analysis of rare and adverse event data. Expert Rev Pharm Out 2002; 2 (4): 367-379.
14. Birch MW. Maximum-Likelihood in 3-Way Contingency-Tables. JRSS, Series B 1963; 25 (1): 220-233.
15. Vollset SE, Hirji KF, Elashoff RM. Fast Computation of Exact Confidence-Limits for the Common Odds Ratio in a Series of 2 x 2 Tables. JASA 1991; 86 (414): 404-409.
16. Sankey SS, Weissfeld LA, Fine MJ, Kapoor W. An assessment of the use of the continuity correction for sparse data in meta-analysis. Commun Stat – Simul C 1996; 25 (4): 1031-1056.
17. Deeks J, Bradburn M, Localio R, Berlin J. Much Ado About Nothing: Statistical Methods for Meta-analysis with Rare Events. Oxford, UK: Centre for Statistics in Medicine, Institute of Health Sciences; 1999.
18. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004; 23 (9): 1351-1375.
19. Mehta CR, Walsh SJ. Comparison of Exact, Mid-P, and Mantel-Haenszel Confidence-Intervals for the Common Odds Ratio Across Several 2×2 Contingency-Tables. Am Stat 1992; 46 (2): 146-150.
20. Lancaster HO. Significance tests in discrete distributions. JASA 1961; 56 (294): 223-234.
21. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002; 21 (4): 589-624.
22. Cheng DC, Bainbridge D, Martin JE, Novick RJ, Evidence-Based Perioperative Clinical Outcomes Research Group. Does off-pump coronary artery bypass reduce mortality, morbidity, and resource utilization when compared with conventional coronary artery bypass? A meta-analysis of randomized trials. Anesthesiology 2005; 102 (1): 188-203.
23. Senn S. The many modes of meta. Drug Inf J 2000; 34 (2): 535-549.
24. Breslow N. Odds Ratio Estimators When the Data Are Sparse. Biometrika 1981; 68 (1): 73-84.
25. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. Stat Med 1999; 18 (6): 643-654.
26. Gao S. Combining binomial data using the logistic normal model. J Stat Comput Simul 2006; 74 (4): 293-306.
27. Senn S. Controversies concerning randomization and additivity in clinical trials. Stat Med 2004; 23 (24): 3729-3753.
28. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics 1985; 41 (1): 55-68.
29. Lui KJ. A Monte Carlo evaluation of five interval estimators for the relative risk in sparse data. Biometrical J 2006; 48 (1): 131-143.
30. Hirji KF, Mehta CR, Patel NR. Computing distributions for exact logistic regression. JASA 1987; 82: 1110-1117.

Corresponding author:
Dr. Oliver Kuss
Institut für Medizinische Epidemiologie, Biometrie und Informatik
Martin-Luther-Universität Halle-Wittenberg
Magdeburger Str. 27
06097 Halle (Saale)
Germany
E-mail: Oliver.Kuss@medizin.uni-halle.de