International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 5, No. 1 (May-2018)

http://dx.doi.org/10.12785/ijcts/050106

A Comparative Study on Methods for Measuring Heterogeneity in Binomial Proportion – Meta Analytic Approach

S.Sakthivel[1], M. Subbiah[2] and R. Ramakrishnan[3]

[1]Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
[2]Department of Mathematics, Presidency College, Chennai, Tamil Nadu, India
[3]National Institute of Epidemiology (ICMR), Ayapakkam, Chennai, Tamil Nadu, India

Received November 13, 2017, Revised March 5, 2018, Accepted April 15, 2018, Published May 1, 2018

Abstract: Statistical inference on a common quantity of interest arising from multiple related studies is quite pervasive in medical, social, statistical genetics, clinical trials, and epidemiological research. However, a scope is identified to study the issues related to pooling proportion across many Bernoulli type of events. The Main objective of the study is to exploit the asymptotic approximation of proportion and logit transformation in random effects model. Seven procedures for estimating between-variance in frequentist inferential methods have been used. Number of Bernoulli trials, number of studies, and events occurred at boundaries of the parameter are prime data characteristics considered in the study. From the results it is possible to device a methodology to choose more appropriate methods under different problem situations. All the procedures are implemented in R and major functions are presented in Appendix.

Keywords: Binomial distribution, Heterogeneity, Proportion, Random effects model.

1. Introduction

Combining information from multiple studies is one of the wide spread statistical practices in medical, social, statistical genetics, clinical trials, and epidemiological research. Most prominent such approaches are Multi-centric studies, randomized clinical trial, and meta-analysis. Objective and purpose of each method may be different, yet underlying statistical model is mostly confined to fixed or random effects model (FEM or REM). Extensive literature on methodology, computation, and applications are available; partial and recent list may include [29, 30, 31, 32, 38, 42, 45, 48, 54, 57 & 56].

Researchers actively discuss on the suitability of FEM or REM for the given situation and their impact on the outcome of interest. Agresti and Hartzel [14] and references [19, 21, 39, 40, 41, 44, 47, 53 & 55] provide necessary details on these topics. These and similar other research have focused on the estimation of variance components in the model and few data issues such as zero and / or sparse data that are influencing the analyses. Especially, between variance in REM has drawn more research attention under frequentist as well as Bayesian inferential procedures.

Another important observation is data type and associated summary measures in FEM and REM. Majority of the studies have confined to binary outcome of two variables represented in a 2 x 2 contingency table and associated summary measures risk difference, risk ratio and odds ratio. Relatively few works consider count data models or measurable variables with mean or mean difference as summary measure. On the other hand, Bernoulli type of experiments with pooling proportion of success from multiple sources as quantity of interest has witnessed limited research activities.

This paper has identified a scope to investigate the meta-analytic approach in combining binomial proportion from multiple studies. The main objective is to understand the impact of between-variance estimate in a typical REM when proportion is the underlying summary measure. Along with this, zero success / failure and / or small sample problems posit interesting research objectives. Another important, but could be less demanding in the present era is the limited

E-mail address: sakthiapj@gmail.com, sisufive@gmail.com, drramakrishnan.nie@gmail.com

availability of selected procedures in wide spread statistical software; though [52] and [46] provide few procedures related to proportions.

Hence, the present work has aimed to provide a comprehensive list of procedures for meta-analytic approaches; especially, on estimators of between variance under REM and their impact on point and interval estimates of pooled proportion. This includes identifying predominant seven frequentist procedures and their performances under different data scenarios. Also aims to short list illustrative datasets that focus the hypothesis of pooling proportions from multiple studies or strata; most appropriate datasets (six out of 22) are considered for the presentation. Results and subsequent discussion have helped to understand possible factors that could influence the analysis and recommend suitable methods for future practice. R codes for three important functions are made available.

2. EXAMPLES

The study is motivated from problems on pooling proportion from a binomial model detailed in medical or epidemiological literature. Six datasets (numbered D I to D VI) are extracted from four such studies that provide meta analysis, clinical trial, and statistical genetics data; D I and D II from [34], D III and D IV from [50], D V and D VI are respectively from [27] and [33]. All these studies have focused on approaches for combining proportions. It ranges from crude proportions $\sum \frac{x_i}{n_i}$ to Meta analytic approaches but limited to DerSimonian and Laird method and Cochran Q statistic.

Also, heterogeneity has been estimated using Chi-square as well as Mantel–Haenszel tests to proceed with FEM [27 & 50] for datasets that are not in our choice. Nevertheless, a preliminary investigation with Chi square test on the selected datasets has rejected the null hypothesis of homogeneity. This augments the research interest to investigate REM together with prevalent constraints on data characteristics. In this case, they are number of successes (x), number of Bernoulli trial (n) in each of the independent but related k studies. Sparse nature (low x), cases with x near or at boundary (x ~ 0 or x ~ n; ~ denotes close or equal to), and size of k are also pose research questions on suitability of inferential procedures.

Table 1 provides summary statistics of these six datasets. It can be observed from Table 1, k varies from 3 to 41; D III and D VI illustrate cases for x = n and x = 0 respectively; D II and D IV are showing the incidence of highly polarized n with k = 3 and k =10. D II resembles D I in terms of k but in D II n and x have nearly same range but not in D I. D III has higher ranges in both n and x with k = 7. More about D VI is that, it is combination of all these attributes except any incidence of x = n

3. RANDOM EFFECT MODELS

If there are k independent studies with an effect parameter θ_i which is subjected to have a sampling variance ε_i where $\varepsilon_i \sim N(0, \sigma_i^2)$. Let Y_i be an estimate for the corresponding true effect size θ_i with the within-study variance σ_i^2, then the random effects model is

$$Y_i \sim N(\theta_i, \sigma_i^2)$$

Further θ_i is assumed from a population with an effect size μ and an error δ_i where $\delta_i \sim N(0, \tau^2)$ so that

$$\theta_i \sim N(\mu, \tau^2)$$

where τ^2 is the total amount of heterogeneity (between-study variance). If θ_i are estimated from data directly then the objective remains to estimate μ and τ^2

In the present case, Y_i is the logit of p_i the proportion of success in i^{th} stratum (i = 1, 2…k);
That is $Y_i = \log(p_i/(1-p_i))$ and $x_i \sim$ Binomial(n_i, p_i).

In almost all cases, within variance σ_i^2 is assumed to be known or estimated using asymptotic approach. Delta method [18] may be used to find variance of function of parameter. Especially in univariate case, say $f(\phi)$ it can be observed that $Var[f(\phi)] = \nabla f(\phi)^2 . Var[\hat{\phi}]$ where $\hat{\phi}$ is the maximum likelihood estimate (MLE) of ϕ

Accordingly, when $x_i \sim$ Binomial(n_i, p_i), then $V[\hat{p}_i] = \frac{p_i(1-p_i)}{n_i}$.

Hence, $Var[f(\phi)] = Var\left[logit(p_i)\right]$

$$= \frac{1}{[p_i(1-p_i)]^2} \frac{p_i(1-p_i)}{n_i}$$

$=\frac{1}{n_i p_i (1-p_i)}$; replace the parameter with its MLE we get $\frac{n_i}{x_i(n_i-x_i)}$. Suitable corrections on x should be made when x = 0 or x = n and hence y_i and its variance can be estimated.

Testing $\tau^2 = 0$ against $\tau^2 \neq 0$ (precisely $\tau^2 > 0$) is the most important statistical objective; failing to reject the former leads to FEM otherwise to REM. For the problem of pooling proportions, FEM reduces to sample mean which is a uniformly minimum unbiased variance. In some cases, weighted averages are also discussed [27]. However, this paper deals mainly REM for pooling proportions using different methods to estimate between-variance. The duality between test of hypothesis and confidence intervals is applied in variety of applications and can also be found in theoretical context, for example Casella and Berger [18].

Seven frequentist methods for estimating between variance τ^2 are considered based on the recommendations from the extensive literature [29, 30 & 56]. Identified procedures are maximum-likelihood (ML), Q profile restricted maximum-likelihood (QPR_REM), DerSimonian-Laird (DL), Sidik-Jonkman (SJ), Hedges (HE), Hunter-Schmidt (HS), and Paule-Mandel (PM).

4. DATA ANALYSIS

All seven procedures are written in R language using suitable packages. Major functions are presented in Appendix. Estimates of overall proportion (μ) and between variance (τ^2) using seven methods are presented in Tables 2 and 3 respectively. Table 4 is the point and 95% confidence interval estimates for the individual (strata wise) studies in dataset VI for the purpose of illustration; this also eases pictorial representation of D VI. Forest plots provide corresponding graphical summaries; frequentist estimates for overall proportion, between-variance from seven methods and study specific estimates for D I to D VI are depicted in Figures 1, 2, and 3 respectively.

From Table 2 it is evident that point estimates of overall proportion is quite similar between all seven frequentist procedures across six datasets. Also from Figure 2 it could be observed that behaviour of frequentist methods is comparable in their width; HS yields the least width followed by ML intervals except in D V. Also the difference is minimal for large k and n.

Frequentist estimates of between variance (Table 3 and Figure 2) reveal some interesting performance of identified seven methods. It is obvious to note the difference in point estimates in all the datasets. HS yields uniformly least estimates except D VI where HE has distinctively shown zero as a measure of heterogeneity though corresponding interval is relatively wider. When k is small except HS and ML, other methods are producing very large between variance. This difference starts decreasing when k becomes larger.

Secondly, in terms of interval estimators, the lower limit due to ML is consistently negative which is a case of aberration for variance estimator; upper limit of SJ and QPR_REM are always higher when k is small and closer to other methods when k is large with varying n and x. HE and DL also follow in this behaviour but not so appreciably higher values for their upper limits. PM plays a compromising role in measuring between variance when compared to HS and other five frequentist procedures. Figures 3, presents the study wise estimates from frequentist methods for individual proportion parameter for all six datasets.

Further, this study has considered a systematic simulation plan in order to understand the performance of the estimators for overall proportion and between variance in a random effects model. In this process, logit transformation is used with corresponding within variance being estimated using delta method and then the assumption of normality on these logits are used. Hence the perceived simulation plan follows these steps in its design; R language is used for the implementation. The plan has three layers in terms of choosing parametric values for the associated quantities in the model; the first step is to fix the value of k and N (overall sample size) which is the sum of individual sample sizes n_i.

Subsequently, population proportion of the binomial model is fixed in three different scenarios representing near boundaries (zero and one) and covering the point of symmetry (0.5) as well as compliment to the first two intervals in the range [0, 1] for the proportion parameter. This would help in simulating data sets which could represent zero and / or low counts. More importantly this simulation works with a constraint that $N = \sum n_i$ which is achieved by appropriately using the random numbers from a multinomial distribution

Hence, the simulation plan is

Step 1: Fix K – 5, 10, 20, 50

Step 2: Fix N – 50, 100, 200, 500, 750, 1000, 1500

Step 3: Fix p_i – Generating from Uniform (a, b)

Scenario 1: a = 0; b = 0.1

Scenario 2: a = 0.1; b = 0.9

Scenario 1: a = 0.9; b = 1

Step 4: Generate n_i from multinomial (N, p) where p ~ Uniform (0, 1)

Step 5: Generate x_i from Binomial (n_i, p_i)

Step 6: Apply seven frequentist methods for REM to obtain overall proportion and between variance

Repeat Steps 5 and 6 for 500 times and summarize the results using a five point summary and plotting the simulated estimates in a box-whisker-plot to visually study the variations in the Monte-Carlo simulation. Further, this simulation study includes one more constraint for the individual sample sizes to be equal so that Step 4 of the above plan is modified with the fixed size of N/k; cases which result in a non-integer N/k are not included in the plan for obvious reasons.

Results from these extensive combinations (48 cases in each of two conditions of n_i) are compared to study the performance of the measures. An illustration is provided in Tables 5 and 6 and Figure 4 for the combination k = 20, N = 1000 and p_i ~ Uniform (0, 0.1) due to paucity of space, all the simulation results are not presented. With regard to the methods for estimating between variance, it is observed in Figure 4 that HS yields uniformly lower values with lesser τ^2; followed by ML and DL whereas other four methods (QP_REM, PM, HE, SJ) do not provide comparatively consistent estimates by comparing their variations.

Further, among the latter four methods SJ seems to be less consistent across all combinations; this can be observed from its distinct behaviour with respect to the estimate of between variance. This behaviour is also reflected in the overall proportion in which case, all methods except SJ provide similar estimate with notably lesser variability. Another observation is the behaviour of DL is quite similar with HS when overall sample size (N) increases irrespective of k and proportion parameter.

5. CONCLUSION

Statistical inference on a common quantity of interest arising from multiple related studies is quite pervasive in medical, social, statistical genetics, clinical trials, and epidemiological research. However, a scope is identified to study the issues related to pooling proportion across many Bernoulli type of events.

The Main objective of the study is to exploit the asymptotic approximation of proportion and logit transformation in random effects model. Seven procedures for estimating between-variance in frequentist inferential methods have been used. Number of Bernoulli trials, number of studies, and events occurred at boundaries of the parameter are prime data characteristics considered in the study.

Estimates from multiple studies are combined to understand a measure of interest and its variability. Random effects model is a most useful statistical inferential procedure to perform this task. Literature in this topic mainly focuses on measures pertaining to binary classification variables. But it forms the basis for including other suitable data types and summary measures.

This work has exploited one such problem pertaining to pooling multiple proportions from independent but related Bernoulli type of events. Frequentist approach, one of the two major paradigms in statistical inference is applied in six datasets extracted from published research. They are to illustrate the factors that may influence the analysis and inference on these data types; sample size, number of successes, and zero occurrences are few of them.

Seven frequentist methods are considered based on the recommendations derived from existing studies. Normal approximation method has been followed for estimating within variance in frequentist REM.

Comparative analysis has shown the distinct behaviour of the chosen seven between-variance estimator when Bernoulli proportion is the underlying summary measure. In particular it is possible to highlight the consistent performance of HS in different form of datasets; this is closely followed by ML and DL; PM and QPR_REM may be used with caution and SJ could be avoided. These observations are characterized by different attributes;

- HS may be adopted for any size of k, N, x with or without zeros
- DL could be the choices for larger preferably k and n
- PM and QPR_REM may not be preferred while k is small
- ML is an option if aberration is considered as a minor issue
- SJ could be avoided for any combinations of k and N

In conclusion it can be emphasized that the choice of an optimum procedure is prevalent in any statistical investigation. In that sense, HS, DL and ML are the methods of choice in frequentist frame work. R codes are provided in the Appendix for replicating similar analyses. Also, a systematic simulation based analysis has been carried out as a comparative study on pooling of proportion from multiple studies. This attempt is quite similar to the observations from the six data sets representing various scenarios on model parameters.

From the results of a comparative data analysis and simulation study it is possible to device a methodology to choose more appropriate methods under different problem situations. All the procedures are implemented in R and major functions are presented in Appendix

One of the aspect not included in this work is, the impact of alternate continuity correction schemes for zero successes in one or more studies. This topic will be focused in our future research. Another aspect for possible future research is arc-sine or double arc-sine transformation of proportion which too has attracted research debate [38, 43 & 52].

References

[1] Cochran, W.G. (1937). Problems arising in the analysis of a series of similar experiments. Journal of the Royal Statistical Society, 4, 102–118.

[2] Cochran, W.G. (1954). The combination of estimates from different experiments. Biometrics, 10, 101–129.

[3] Hedges, L.V. (1983). A random effects model for effect sizes. Psychological Bulletin, 93, 388–395.

[4] Morris, C.N. (1983). Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association, 78, 47–55.

[5] Raudenbush, S.W., & Bryk, A.S. (1985). Empirical Bayes meta-analysis. Journal of Educational Statistics, 10, 75–98.

[6] DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials, 7, 177–188.

[7] Morris, C.N., & Normand, S.L. (1992). Hierarchical models for combining information and for meta-analyses. Bayesian statistics, 4, 321-344.

[8] Switzer, F.S., Paese, P.W., & Drasgow, F. (1992). Bootstrap estimates of standard errors in validity generalization. Journal of Applied Psychology, 77 (2), 123–129.

[9] Raudenbush, S.W. (1994). Random effects models, In The Handbook of Research Synthesis, Cooper H, Hedges LV (eds). Russell Sage Foundation: NY, 301–321.

[10] Berkey, C.S., Hoaglin, D.C., Mosteller, F., & Colditz, G.A. (1995). A random-effects regression model for meta-analysis. Statistics in Medicine, 14, 395–411.

[11] Hardy, R.J., & Thompson, S.G. (1996). A likelihood approach to meta-analysis with random effects. Statistics in Medicine, 15 (6), 619–629.

[12] Biggerstaff, B.J., & Tweedie, R.L. (1997). Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Journal of Statistics in Medicine, 16 (7), 753–768.

[13] Thompson, S.G., & Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. Statistics in Medicine, 18, 2693–2708.

[14] Agresti, A., & Hartzel, J. (2000). Tutorial in biostatistics: strategies for comparing treatments on a binary response with multi-center data. Statistics in Medicine, 19, 1115-1139.

[15] Turner, R.M., Omar, R.Z., Yang, M., Goldstein, H., & Thompson, S.G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Statistics in Medicine, 19 (24), 3417–3432.

[16] Brockwell, S.E., & Gordon, I.R. (2001). A comparison of statistical methods for meta-analysis. Statistics in Medicine, 20, 825–840.

[17] Bohning, D., Malzahn, U., & Dietz, E et al. (2002). Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. Journal of Biostatistics, 3, 445–457.

[18] Casella, G., & Berger, R.L. (2002). Statistical inference. Pacific Grove, CA: Duxbury, Vol. 2.

[19] Deeks, J.J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Statistics in Medicine, 21 (11), 1575-1600.

[20] Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21, 1539–1558.

[21] Van Houwelingen, H.C., Arends, L.R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine, 21 (4), 589-624.

[22] Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine, 22, 2693–2710.

[23] Bohning, D., Sarol, J., Rattanasiri, S., Viwatwongkasem, C., & Biggeri, A. (2004). A comparison of non-iterative and iterative estimators of heterogeneity variance for the standardized mortality ratio. Journal of Biostatistics, 5, 61–74.

[24] Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). Bayesian data analysis, Chapman and Hall/CRC

[25] Makambi, K.H. (2004). The effect of the heterogeneity variance estimator on some tests of treatment efficacy. Journal of Biopharmaceutical Statistics, 14, 439–449.

[26] Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., & Jones, D.R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine, 24, 2401–2428.

[27] Qiao, C.G., Wood, G.R., & Lai, C.D. (2005). Estimating a binomial proportion from several independent samples. New Zealand Journal of Crop and Horticultural Science, 33 (3), 293-302.

[28] Sidik, K., & Jonkman, J.N. (2005). Simple heterogeneity variance estimation for meta-analysis. Journal of the Royal Statistical Society, Series C: Applied Statistics, 54, 367–384.

[29] Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics, 30 (3), 261-293.

[30] DerSimonian, R., & Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: an update. Contemporary Clinical Trials, 28 (2), 105-114.

[31] Sidik, K., & Jonkman, J.N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. Statistics in Medicine, 26 (9), 1964-1981.

[32] Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. Statistics in Medicine, 26 (1), 37-52.

[33] Young-Xu, Y., & Chan, K.A. (2008). Pooling over dispersed binomial data to estimate event rate. BMC Medical Research Methodology, 8 (1), 58.

[34] Arbyn, M., Martin-Hirsch, P., Buntinx, F., Van Ranst, M., Paraskevaidis, E., & Dillner, J. (2009). Triage of women with equivocal or low-grade cervical cytology results: a meta-analysis of the HPV test positivity rate. Journal of Cellular and Molecular Medicine, 13 (4), 648-659.

[35] Bagos, P.G., & Nikolopoulos, G.K. (2009). Mixed-effects Poisson regression models for meta-analysis of follow-up studies with constant or varying durations. The International Journal of Biostatistics, 5 (1), article 21.

[36] Bonett, D.G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. Psychological Methods, 14, 225–238.

[37] Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), The handbook of research synthesis and meta-analysis, 2nd ed., 221–235. New York: Russell Sage Foundation.

[38] Rücker, G., Schwarzer, G., Carpenter, J., & Olkin, I. (2009). Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. Statistics in Medicine, 28 (5), 721-738.

[39] Borenstein, M., Hedges, L.V., Higgins, J., & Rothstein, H.R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods, 1 (2), 97-111.

[40] Haidich, A.B. (2010). Meta-analysis in medical research’, Hippokratia, 14 (Suppl 1), 29.

[41] Nissen, S.E., & Wolski, K. (2010). Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. Archives of Internal Medicine, 170 (14), 1191-1201.

[42] Thorlund, K., Imberger, G., Walsh, M., Chu, R., Gluud, C., Wetterslev, J., & Thabane, L. (2011). The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis-a simulation study. PloS one, 6 (10), e25491.

[43] Warton, D.I., & Hui, F.K. (2011). The arcsine is asinine: the analysis of proportions in ecology. Ecology, 92 (1), 3-10.

[44] Bhaumik, D.K., Amatya, A., Normand, S.L.T., Greenhouse, J., Kaizar, E., Neelon, B., & Gibbons, R.D. (2012). Meta-analysis of rare binary adverse event data. Journal of the American Statistical Association, 107 (498), 555-567.

[45] Jackson, D., White, I.R. & Riley, R.D. (2012). Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Statistics in Medicine, 31 (29), 3805-3820.

[46] Barendregt, J.J., Doi, S.A., Lee, Y.Y., Norman, R.E., & Vos, T. (2013). Meta-analysis of prevalence. Journal of Epidemiol Community Health, 67, 974-978.

[47] Lane, P.W. (2013). Meta-analysis of incidence of rare events. Statistical Methods in Medical Research, 22 (2), 117-132.

[48] Rukhin, A.L. (2013). Estimating heterogeneity variance in meta-analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75 (3), 451-469.

[49] Aloe, A.M. (2014). An empirical investigation of partial effect sizes in meta-analysis of correlational data. Journal of General Psychology, 141, 47–64.

[50] Dolman, L., Sauvaget, C., Muwonge, R., & Sankaranarayanan, R. (2014). Meta-analysis of the efficacy of cold coagulation as a treatment method for cervical intraepithelial neoplasia: a systematic review. BJOG. An International Journal of Obstetrics & Gynaecology, 121 (8), 929-942.

[51] Fagerland, M.W., Lydersen, S., & Laake, P. (2014). Recommended tests and confidence intervals for paired binomial proportions. Statistics in Medicine, 33, 2850–2875.

[52] Nyaga, V.N., Arbyn, M., & Aerts, M. (2014). Metaprop: a Stata command to perform meta-analysis of binomial data. Archives of Public Health, 72 (1), 39.

[53] DerSimonian, R., & Laird, N. (2015). Meta-analysis in clinical trials revisited. Contemporary Clinical Trials, 45, 139-145.

[54] Langan, D., Higgins, J., & Simmonds, M. (2015). An empirical comparison of heterogeneity variance estimators in 12,894 meta-analyses. Research synthesis methods, 6 (2), 195-205.

[55] Shuster, J.J., & Walker, M.A. (2016). Low- event- rate meta- analyses of clinical trials: implementing good practices. Statistics in Medicine, 35 (14), 2467-2478.

[56] Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. Research Synthesis Methods, 7 (1), 55-79.

[57] Langan, D., Higgins, J.,& Simmonds, M. (2017). Comparative performance of heterogeneity variance estimators in meta-analysis: a review of simulation studies. Research synthesis methods, 8 (2), 181-198.

Appendix

This section provides important R functions that are used for applying REM. To avoid the redundancy functions with similar syntax are not presented. Similarly graphical tool is avoided due to its simplicity and other possibilities to generate graphs of interest.

```
f_mu_OALL=function(x, n)
# This function is to obtain point and interval estimates of overall proportion
{
require(metafor)
require(boot)
k=length(x)
p1=escalc(xi=x,ni=n,measure = "PLO")#point estmate - logit(p) with va(logit(p))
p2_ML=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "ML")
p2_R=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "REML")
p2_DL=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "DL")
p2_SJ=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "SJ")
p2_HE=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "HE")
p2_HS=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "HS")
p2_PM=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "PM")
mu_OALL=round(rbind(inv.logit(p2_ML$b),inv.logit(p2_R$b),inv.logit(p2_DL$b),inv.logit(p2_SJ$b),inv.logit(p2_HE$b),inv.logit(p2_HS$b),inv.lo
git(p2_PM$b)) ,4)
```

```
mu_OALL_LL=round(rbind(inv.logit(p2_ML$ci.lb),inv.logit(p2_R$ci.lb),inv.logit(p2_DL$ci.lb),inv.logit(p2_SJ$ci.lb),inv.logit(p2_HE$ci.lb),inv.lo
git(p2_HS$ci.lb),inv.logit(p2_PM$ci.lb)) ,4)
mu_OALL_UL=round(rbind(inv.logit(p2_ML$ci.ub),inv.logit(p2_R$ci.ub),inv.logit(p2_DL$ci.ub),inv.logit(p2_SJ$ci.ub),inv.logit(p2_HE$ci.ub),in
v.logit(p2_HS$ci.ub),inv.logit(p2_PM$ci.ub)) ,4)
dif1=mu_OALL_UL-mu_OALL_LL #Width of CI
OVERALL_P=round(cbind(mu_OALL,mu_OALL_LL,mu_OALL_UL,dif1),4)
row.names(OVERALL_P)=c("ML","QPR_REM","DL","SJ","HE","HS","PM")
colnames(OVERALL_P)=c("mu","mu_LL","mu_LL","Width_CI")
print(OVERALL_P)
}

f_tau2=function(x,n)
#To obtain point and interval estimates of between-variance using QPR_REM
{
require(metafor)
k=length(x)
p1=escalc(xi=x,ni=n,measure = "PLO")
p2_R=rma(yi=p1$yi,vi=p1$vi,measure = "PLO",method = "REML"
QPO1=confint(p2_R)
QPR_REM=round(cbind(Esti=QPO1$random[1],LL=QPO1$random[1,2],UL=QPO1$random[1,3]),4)
list(QPR_REM)
}
```

TABLE 1. DESCRIPTION OF THE SIX DATASETS CONSIDERED IN THE STUDY; DETAILS INCLUDE PROPORTION OF NO SUCCESSES (PNS) COMPUTED AS THE RATIO OF NUMBER OF ZEROS TO NUMBER OF STUDIES IN A DATASET; SIMILARLY ALL SUCCESSES (PAS) IS PRESENTED

Dataset	No of studies	No of successes (x)			No of trials (n)			PNS	PAS
		Sum	Min	Max	Sum	Min	Max		
D I	3	1680	88	1239	3684	543	1924	0.000	0.000
D II	3	1187	263	498	2198	653	844	0.000	0.000
D III	7	593	10	445	620	11	459	0.000	14.286
D IV	10	2362	14	1343	2523	15	1453	0.000	20.000
D V	15	378	2	69	3734	175	312	0.000	0.000
D VI	41	111	0	13	3002	12	186	39.024	0.000

TABLE 2. OVERALL PROPORTION ESTIMATES AND 95 % CONFIDENCE INTERVAL LIMITS BY SEVEN FREQUENTIST METHODS

Methods	Datasets					
	D I	D II	D III	D IV	D V	D VI
ML	0.344 (0.155, 0.600)	0.561 (0.355, 0.747)	0.929 (0.868, 0.963)	0.929 (0.891, 0.954)	0.074 (0.044, 0.121)	0.043 (0.033, 0.056)
QPR_REM	0.344 (0.126, 0.655)	0.561 (0.313, 0.781)	0.929 (0.861, 0.965)	0.929 (0.888, 0.956)	0.074 (0.043, 0.123)	0.043 (0.033, 0.055)
DL	0.344 (0.126, 0.655)	0.561 (0.307, 0.786)	0.929 (0.854, 0.967)	0.929 (0.891, 0.954)	0.078 (0.053, 0.115)	0.044 (0.034, 0.056)
SJ	0.344 (0.127, 0.655)	0.561 (0.314, 0.781)	0.929 (0.859, 0.965)	0.933 (0.877, 0.964)	0.073 (0.042, 0.124)	0.040 (0.030, 0.054)
HE	0.344 (0.126, 0.655)	0.561 (0.313, 0.781)	0.923 (0.875, 0.961)	0.933 (0.877, 0.964)	0.073 (0.042, 0.125)	0.053 (0.045, 0.064)
HS	0.345 (0.168, 0.578)	0.561 (0.350, 0.752)	0.929 (0.868, 0.963)	0.928 (0.899, 0.949)	0.079 (0.055, 0.114)	0.045 (0.035, 0.057)
PM	0.344 (0.126, 0.655)	0.561 (0.313, 0.781)	0.929 (0.868, 0.963)	0.930 (0.884, 0.959)	0.073 (0.042, 0.124)	0.045 (0.036, 0.057)

TABLE 3. BETWEEN-VARIANCE ESTIMATES AND 95 % CONFIDENCE INTERVAL LIMITS BY SEVEN FREQUENTIST METHOD

Methods	Datasets					
	D I	D II	D III	D IV	D V	D VI
ML	0.857 (-0.525, 2.238)	0.544 (-0.337, 1.426)	0.395 (-0.427, 1.218)	0.269 (-0.152, 0.691)	1.053 (0.224, 1.881)	0.235 (-0.041, 0.510)
QPR_REM	1.288 (0.344, 51.147)	0.819 (0.218, 32.581)	0.506 (0.028, 4.029)	0.343 (0.063, 3.495)	1.146 (0.572, 3.287)	0.254 (0.004, 0.557)
DL	1.288 (0.023, 5.336)	0.871 (0.017, 3.084)	0.657 (0.000, 2.361)	0.272 (0.000, 0.989)	0.586 (0.170, 1.238)	0.193 (0.000, 0.491)
SJ	1.285 (0.348, 50.765)	0.816 (0.221, 32.235)	0.546 (0.227, 2.649)	0.751 (0.355, 2.503)	1.224 (0.656, 3.043)	0.407 (0.274, 0.666)
HE	1.288 (0.031, 4.675)	0.819 (0.019, 3.110)	0.281 (0.000, 2.028)	0.729 (0.000, 2.621)	1.275 (0.438, 2.529)	0.000 (0.000, 0.675)
HS	0.709 (0.004, 1.574)	0.575 (0.003, 1.414)	0.394 (0.000, 0.980)	0.131 (0.000, 0.250)	0.524 (0.129, 1.027)	0.173 (0.000, 0.419)
PM	1.288 (0.021, 4.616)	0.871 (0.018, 3.209)	0.657 (0.000, 2.635)	0.272 (0.000, 1.222)	0.586 (0.182, 1.221)	0.193 (0.000, 0.638)

TABLE 4. MEAN AND 95 % CONFIDENCE INTERVAL LIMITS FOR PROPORTION OF SUCCESS FOR THE INDIVIDUAL STUDIES IN DATASET VI

Study No	Frequentist Estimates		
	Mean	Lower Limit	Upper Limit
Study 1	0.038	0.018	0.078
Study 2	0.015	0.002	0.101
Study 3	0.030	0.004	0.186
Study 4	0.027	0.010	0.068
Study 5	0.020	0.001	0.251
Study 6	0.016	0.001	0.211
Study 7	0.024	0.002	0.287
Study 8	0.022	0.001	0.268
Study 9	0.080	0.030	0.195
Study 10	0.100	0.042	0.219
Study 11	0.026	0.002	0.310
Study 12	0.019	0.001	0.236
Study 13	0.007	0.000	0.100
Study 14	0.033	0.005	0.202
Study 15	0.029	0.002	0.336
Study 16	0.077	0.019	0.261
Study 17	0.084	0.043	0.159
Study 18	0.032	0.010	0.093
Study 19	0.003	0.000	0.041
Study 20	0.075	0.042	0.131
Study 21	0.014	0.004	0.055
Study 22	0.065	0.033	0.124
Study 23	0.018	0.003	0.116
Study 24	0.039	0.002	0.403
Study 25	0.010	0.001	0.138
Study 26	0.034	0.011	0.100
Study 27	0.010	0.001	0.143
Study 28	0.053	0.020	0.134
Study 29	0.007	0.000	0.095
Study 30	0.018	0.003	0.116
Study 31	0.059	0.031	0.109
Study 32	0.015	0.002	0.097
Study 33	0.108	0.064	0.178
Study 34	0.011	0.001	0.154
Study 35	0.006	0.000	0.087
Study 36	0.023	0.001	0.277
Study 37	0.021	0.007	0.062
Study 38	0.121	0.066	0.210
Study 39	0.044	0.014	0.128
Study 40	0.100	0.033	0.268
Study 41	0.025	0.008	0.075

TABLE 5. SUMMARY OF THE SIMULATION STUDY CORRESPONDING TO THE DESIGN K = 20, N = 1000 AND p_i ~ UNIFORM (0, 0.1). SEVEN FREQUENTIST METHODS FOR ESTIMATING BETWEEN VARIANCE ARE COMPARED IN THIS SIMULATION EXERCISE

Methods	Min.	1stQu.	Median	Mean	3rdQu.	Max.
ML	0.000	0.000	0.067	0.096	0.152	0.519
QP	0.000	0.022	0.096	0.125	0.195	0.588
DL	0.000	0.048	0.097	0.126	0.189	0.524
SJ	0.234	0.394	0.464	0.494	0.567	0.971
HE	0.000	0.000	0.070	0.156	0.238	0.847
HS	0.000	0.021	0.069	0.097	0.153	0.450
PM	0.000	0.042	0.103	0.136	0.195	0.559

TABLE 6. SUMMARY OF THE OVERALL PROPORTION FROM THE SIMULATION STUDY CORRESPONDING TO THE DESIGN K = 20, N = 1000 AND p_i ~ UNIFORM (0, 0.1)

Methods	Min.	1stQu.	Median	Mean	3rdQu.	Max.
ML	0.044	0.055	0.061	0.060	0.065	0.079
QP	0.043	0.055	0.060	0.059	0.064	0.077
DL	0.044	0.055	0.060	0.059	0.064	0.076
SJ	0.041	0.049	0.054	0.053	0.058	0.069
HE	0.043	0.056	0.060	0.060	0.064	0.084
HS	0.045	0.056	0.061	0.060	0.064	0.077
PM	0.046	0.055	0.060	0.059	0.063	0.076

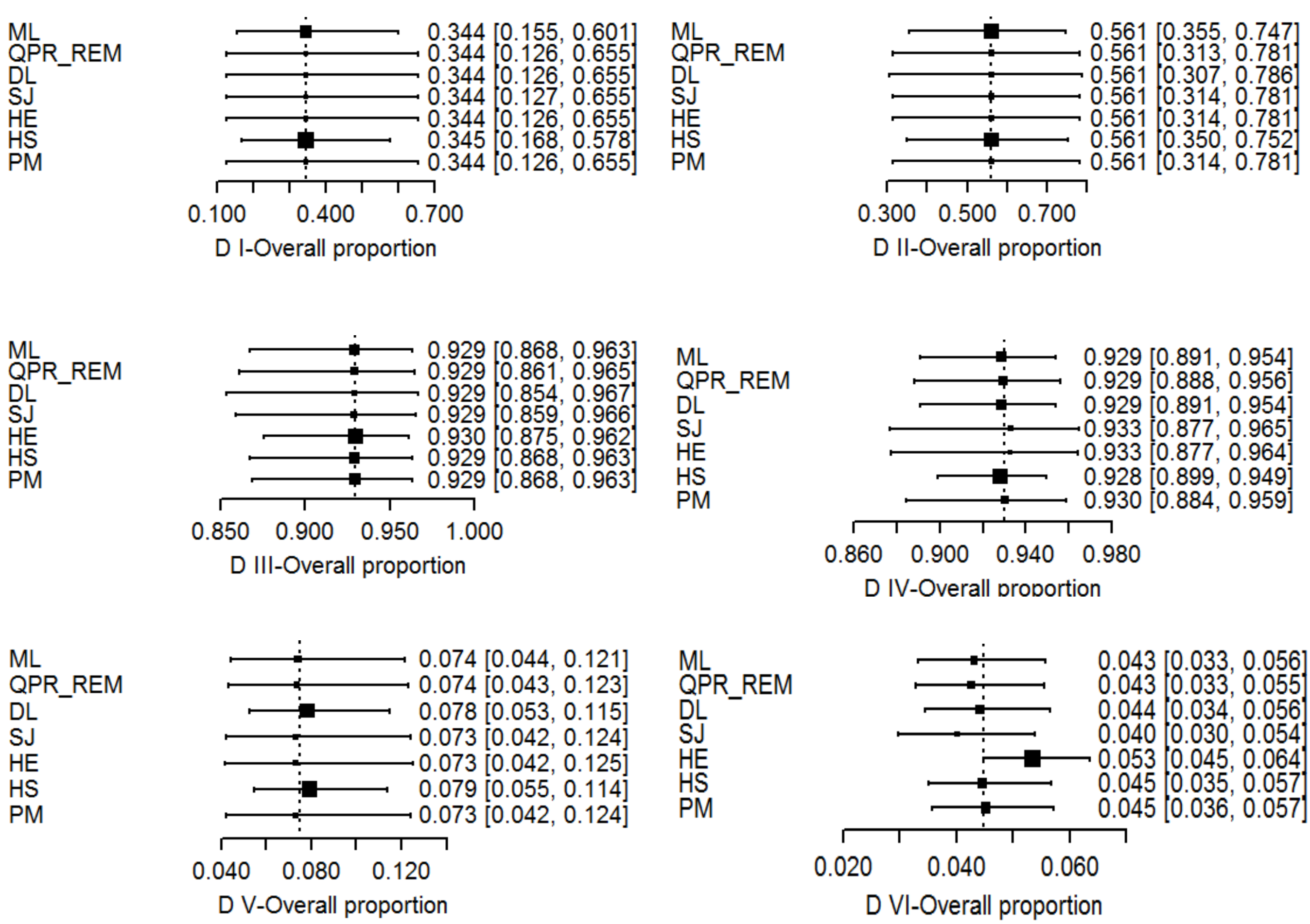

Figure 1. Estimates of overall proportion for six datasets (D I to D VI) obtained using seven frequentist methods. Vertical dotted line represents mean proportion

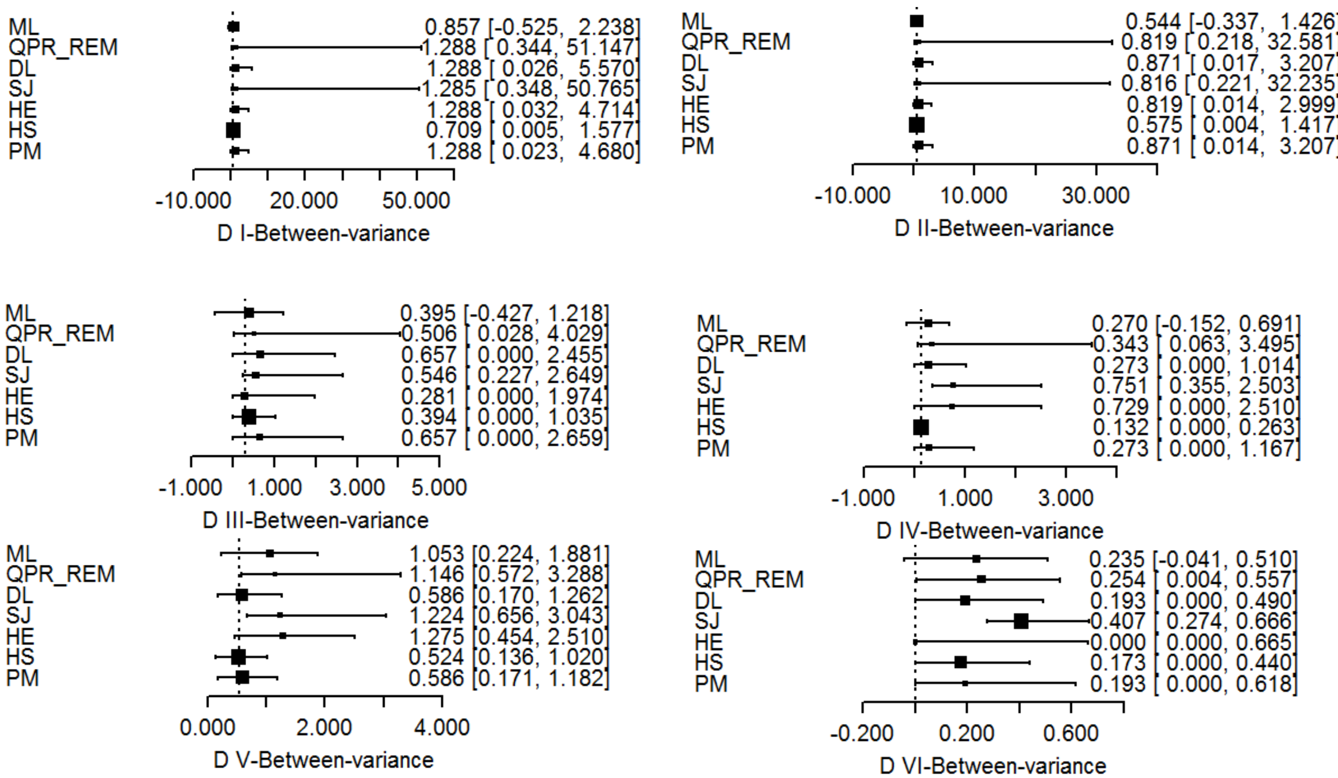

Figure 2. Estimates of between-variance for six datasets (D I to D VI) obtained using seven frequentist methods. Vertical dotted line represents minimum variance

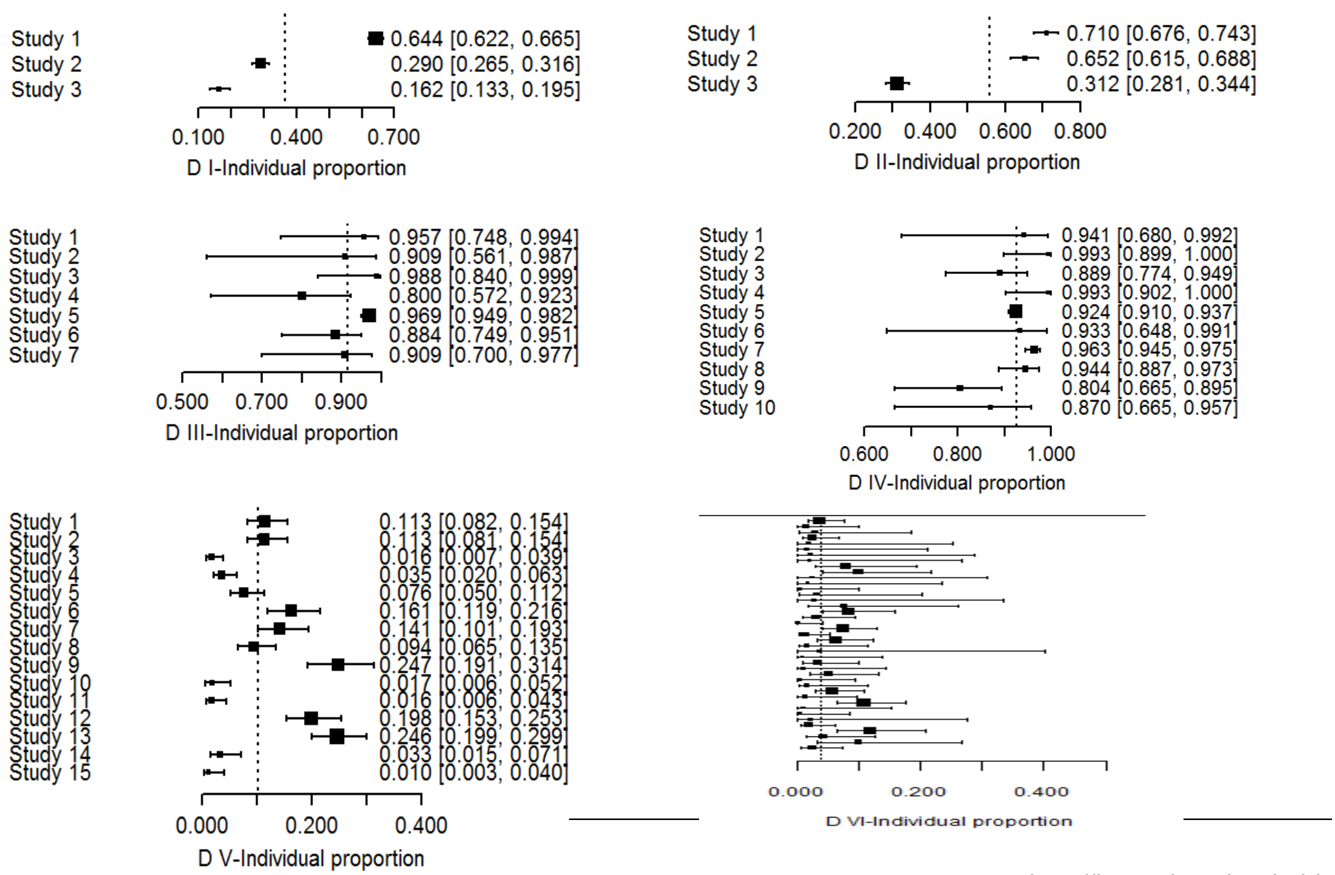

Figure 3. Estimates of study wise proportion for six datasets (D I to D VI) obtained using frequentist method. Vertical dotted line represents mean proportion

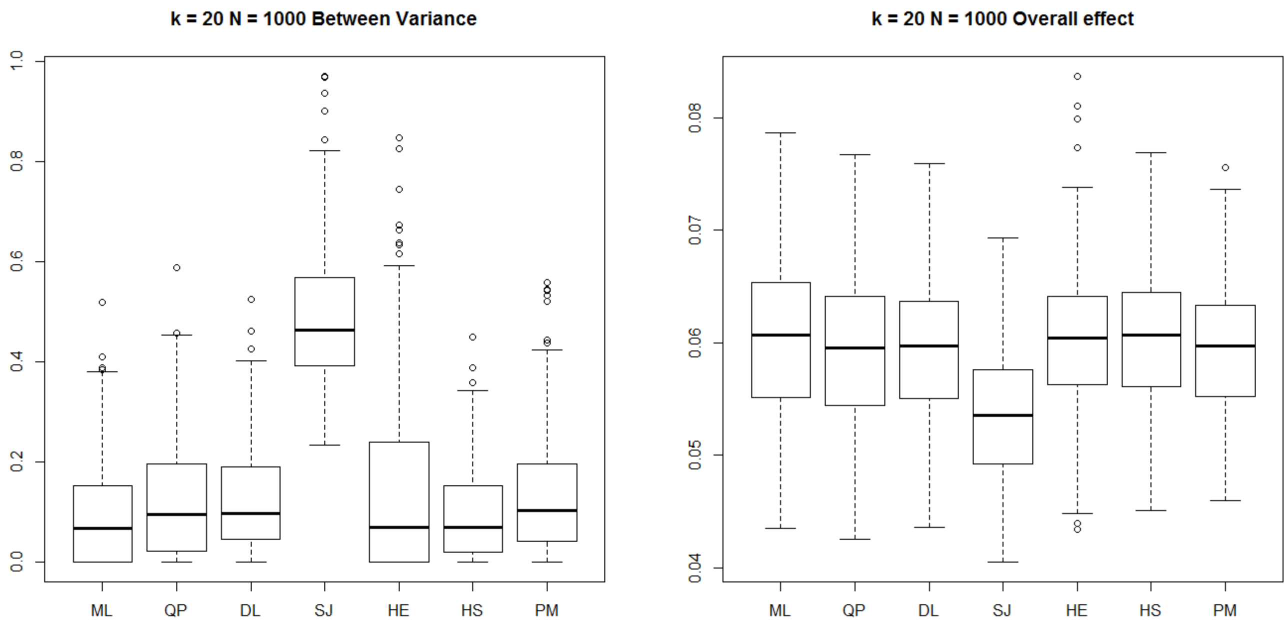

Figure 4.Box-whisker-plots from the simulation study corresponding to the design k = 20, N = 1000 and p_i ~ Uniform (0, 0.1). Seven frequentist methods for estimating between variance (left panel) and overall proportion (right panel) are compared in this simulation exercise