ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 123–125

Journal of Clinical Epidemiology

COMMENTARY

Spuriously precise results from meta-analysis. Is better statistical correction or a more critical methodological assessment warranted?

Ludovic G.P.M. van Amelsvoort[a,*], Wolfgang Viechtbauer[b], Mark G. Spigt[c]

[a]*Department of Epidemiology, Occupational Health Epidemiology unit, School for Public Health and Primary Care (CAPHRI), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

[b]*Department of Methodology and Statistics, School for Public Health and Primary Care (CAPHRI), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

[c]*Department of General Practice, School for Public Health and Primary Care (CAPHRI), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*

Meta-analysis has become the common method for summarizing the results of studies on the same subject. Meta-analyses are thought to produce valid summary estimates of overall effects. However, summarizing all of the information contained in a set of studies into an overall odds ratio and corresponding confidence interval (or *P*-value) might be an oversimplification of reality. Although for randomized controlled trials (RCTs) this has already been acknowledged by Lelorier et al. [1] in 1997 and more recent findings [2], the number of meta-analyses has steadily increased (the number of publications labeled as meta-analyses in PubMed increased from 596 in 1997 to 2,696 in 2007) and the interpretation of results of meta-analyses seems to become ever less refuted.

In this issue of the *JCE*, Salanti and Ioannidis [3] took the spuriously precise results, which are reported by meta-analyses as starting point and devised a method to come to a more valid estimate of evidence from observational research. We greatly welcome a more critical approach to meta-analysis. The paper addresses an important problem and opens the debate on how to do meta-analyses in observational studies and especially how to prevent drawing false conclusions. In view of this, we would like to raise three issues following the method proposed by Salanti and Ioannidis [3].

A first issue that needs future attention is the fact that the c value essentially constitutes a limit to the *P*-value for each study included in the meta-analysis. To illustrate this point, consider a case–control study with 1,216 subjects, equally split among cases and controls. Suppose 68 of the cases and 46 of the controls have been exposed to some risk factor, yielding an odds ratio of 1.54. The corresponding log odds ratio is $y_i = 0.43$ with sampling variance equal to $v_i = 0.040$. Following the reasoning of Salanti and Ioannidis, we can now calculate the ''credibility'' of this finding. Assuming that the observed log odds ratio is equal to the true risk, there is a 0.016 probability that the observed log odds ratio would be smaller than zero (see top part of Fig. 1). This corresponds to the one-sided *P*-value for the test of $H_0 : \theta_i \leqslant 0$, that is, whether the observed log odds ratio is significantly greater than 0 (see bottom part of Fig. 1). Therefore, using a credibility ceiling of $c_i = 0.025$ implies each finding included in the meta-analysis is reduced to nonsignificance (for $\alpha = 0.05$ two sided). Using $c = 0.25$ (a value considered to be the best estimate for one of the examples) therefore implies that the two-sided *P*-value from each study cannot be lower than 0.50. Therefore, large values of c induce a severe reduction in the power of a meta-analysis to detect significant effects.

As an example, assume that the average true risk is equal to $\mu = 0.20$ (i.e., an average odds ratio of 1.22) with heterogeneity equal to $\tau^2 = 0.01$, so that approximately 95% of the true odds ratios fall between 1.0 and 1.5. Moreover, assume that the average sample size of studies examining this risk factor is 100. Then, approximately $k = 10$ studies would be sufficient to have 80% power to detect this effect in a standard meta-analysis. On the other hand, approximately $k = 70$ studies would be needed in a meta-analysis with $c = 0.25$ to achieve the same amount of power. We realize the importance of replications, especially in the context of observational research. And the use of the credibility ceiling method, as described by Salanti and Ioannidis, would be a relevant tool to address this issue. On the other hand, the use of the credibility ceiling approach may reduce the overall power of meta-analyses, especially when relatively few but large studies are involved. This inevitably means that many more studies, money, and time will be spent before reaching a final conclusion and implementing the findings.

It is also worth pointing out that the proposed method will as well reduce the power of meta-analyses when in fact no heterogeneity is present in the true effects. However, it is unlikely that biases, which may reduce our confidence in

* Corresponding author. Fax: +31-43-3884128.

E-mail address: LGPM.vanamelsvoort@epid.unimaas.nl (L.G.P.M. van Amelsvoort).

0895-4356/09/$ – see front matter
doi: 10.1016/j.jclinepi.2008.08.011

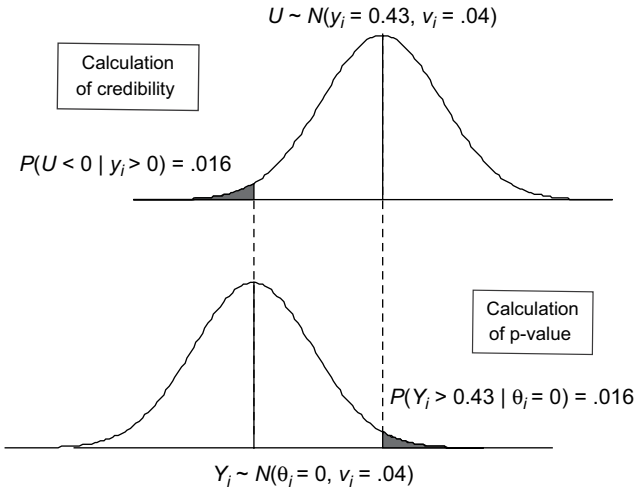

Fig. 1. Example to illustrate the relationship between the ''credibility'' and the one-sided *P*-value of a particular study.

the validity of the observed effects, are operating in such a way that their combined influence yields a relatively homogeneous set of studies. The large values of I^2 observed in the examples considered by the authors may indeed be a reflection of heterogeneity introduced by various biases. However, when I^2 is small, then this may actually strengthen our confidence in the findings from any particular study, in which case we would be hesitant to apply the proposed method.

Secondly, the method proposed by Salanti and Ioannidis assumes that bias always results in an overestimation of the true effect, which needs adjusting by the introduction of the *c* value. But what if the observed effect is a valid estimate, or even an underestimation of the true effect? Take for example, the issue of adverse effects of methylmercury exposure from fish and seafood, as described by Budtz-Jørgensen et al. [4]. In this described case, studies with poor control for the confounding factor, underestimate the true effect. When attributing the same *c* value to all studies, results from studies with a proper adjustment for confounding on average receive relatively less weight than studies without proper adjustment. This is because these studies more quickly reach their ceiling value. The solution would be to use a high *c* value for uncorrected studies and a very low *c* value for corrected studies. Then corrected studies will tend to receive their ''full'' weight, whereas the weight of the uncorrected studies will be decreased to some extent. Alternatively, one could examine the influence of correcting the size of the effect within a moderator analysis. In either case, a careful consideration of the relationship between potential biases and the observed effects is required.

Thirdly, we have a more fundamental concern regarding the proposed method as it does not provide a solution for what in our opinion is the most important problem when doing meta-analyses on observational studies. Common confounding factors, or improper adjustment for these factors, might still be present whether a statistical adjustment like the one proposed is applied or not. Therefore, a meta-analysis of observational studies should always include a careful consideration concerning the contents of the included studies.

Take, for example, the effect of beta-carotene on lung cancer and cardiovascular diseases. Meta-analyses of observational research on this subject revealed a plausible preventive association [5,6], but large RCTs on this subject could not confirm this association or even reported a more worrying inverse association [5,7]. Discrepancies like these would not have been resolved by applying the proposed method of the *c* value to these meta-analyses. In fact, doing so may have resulted in the conduct of even larger observational studies, with potentially even more lives lost. Most vulnerable in this respect are studies where the factors under study are part of a complex matrix of highly correlated behaviors and demographic attributes such as nutritional, lifestyle, and socioeconomic factors. In such cases, meta-analyses can only provide valid estimates when we are able to assess and accurately control for crucial confounders.

On the other hand, it might be possible in some cases to draw valid conclusions from a small number of valid observational studies. We therefore do not agree with the statement made by Salanti and Ioannidis that just a few observational studies cannot be sufficient to provide evidence for a causal effect. When the relationship between a unique exposure and outcome is relatively homogeneous, then it is unlikely that biases are operating and a few studies might be sufficient to firmly establish the presence of a relationship; for example, in the case of vinyl chloride monomer and angiosarcoma [8].

We fully agree with Salanti and Ioannidis that statistical adjustment to prevent spuriously precise estimates should be a method to consider when doing a meta-analysis. However, a critical view with both experts in the field of interest and in statistics and methodology still remains, in our view, the best option for aggregating findings from a set of related studies. Or, to repeat a statement made by Bollen et al. ''In particular, in the meta-analysis of observational research, just pooling of studies could lead to precise but spurious results. In meta-analyses of observational research, statistical pooling should not be a prominent feature, whereas the investigation and critical appraisal of potential sources of heterogeneity and bias should receive principal attention'' [9].

To conclude, we welcome the original and critical approach in improving the methodology of meta-analysis by Salanti and Ioannidis. Although the introduction of a new factor, the ''credibility ceiling'' or *c* value, will inevitably introduce some new problems with the solution of old ones. But the method and it's related problems can be a starting point for new research on this topic. For example, it remains difficult to decide what *c* value or range of *c* values one needs to consider. The authors state that the *c* value might depend on the quality of the studies, but it is unclear

how this would be objectively translated into a specific *c* value (range) in practice. If one wants to base *c* values on quality assessments, future studies need to address the quantification of this relationship and the relation with the interpretation of meta-analysis results.

Acknowledgments

This commentary was written after a discussion of the paper involved in a meeting of the Journal Club Mesch: http://www.hag.unimaas.nl/jcm/. Discussants: Ilja Arts, Luc Budé, Sjoerd Obe Hobma, Daniel Kotz, and Rixt Zijlstra.

References

[1] LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997;337:536−42.

[2] Lawlor DA, Davey Smith G, Ebrahim S. Commentary: the hormone replacement-coronary heart disease conundrum: is this the death of observational epidemiology? Int J Epidemiol 2004;33:464−7.

[3] Salanti G, Ioannidis JPA. Synthesis of observational studies should consider credibility ceilings. J Clin Epidemiol 2009;62:115−22.

[4] Budtz-Jorgensen E, Grandjean P, Weihe P. Separation of risks and benefits of seafood intake. Environ Health Perspect 2007;115:323−7.

[5] Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. BMJ 1998;316:140−4.

[6] Jha P, Flather M, Lonn E, Farkouh M, Yusuf S. The antioxidant vitamins and cardiovascular disease. A critical review of epidemiologic and clinical trial data. Ann Intern Med 1995;123:860−72.

[7] The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. N Engl J Med 1994;330:1029−35.

[8] IARC. Overall evaluation of carcinogenicity: an updating of IARC monographs volumes 1 to 42. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans: supplement 7. Lyon: International Agency for Research on Cancer; 1987.

[9] Bollen CW, Hoekstra MO, Arets HG. Pooling of studies in meta-analysis of observational research leads to precise but spurious results. Pediatrics 2006;117:261−2.