Semi-Bayes and empirical Bayes adjustment methods for multiple comparisons

Metodi di correzione semi-bayesiani e bayesiani empirici per confronti multipli

Marine Corbin,[1,2] Milena Maule,[1] Lorenzo Richiardi,[1] Lorenzo Simonato,[3] Franco Merletti,[1] Neil Pearce[4]

[1] Cancer Epidemiology Unit, CeRMS and CPO, University of Turin, Italy
[2] ENSAI (École Nationale de la Statistique et de l'Analyse de l'Information), Bruz, France
[3] Department of Environmental Medicine and Public Health, University of Padua, Italy
[4] Centre for Public Health Research, Massey University Wellington Campus, Private Box 756, Wellington, New Zealand

***Corrispondenza:** Milena Maule, Epidemiologia dei tumori, via Santena 7, 10126 Torino; tel. 011 6334628; fax 011 6334664; e-mail: milena.maule@cpo.it*

Riassunto

Spesso gli studi epidemiologici richiedono confronti multipli che possono generare un alto numero di risultati falsamente positivi semplicemente perché sono stati condotti molti test statistici. I metodi tradizionali per trattare il problema dei confronti multipli, come per esempio il metodo di Bonferroni, comportano il rischio di trascurare alcuni risultati potenzialmente rilevanti, in quanto non tengono conto del fatto che, sulla base delle evidenze disponibili, alcune esposizioni sono a priori più rilevanti di altre. Infatti, il metodo di Bonferroni non considera l'evidenza a priori, ma semplicemente corregge la significatività statistica (valori p) di ciascun risultato per il numero totale di confronti effettuati. Inoltre, questo metodo non agisce sulle stime di effetto (per esempio odds ratio e relativi intervalli di confidenza). I metodi Bayesiani empirici e semi-Bayesiani per correggere per confronti multipli permettono sia di ridurre il numero di associazioni falsamente positive sia di ottenere stime di effetto in media più valide. Nell'ambito di uno studio caso-controllo su fattori occupazionali e rischio di tumore del polmone abbiamo applicato questi metodi e valutato la loro performance.
(*Epidemiol Prev* 2008; 32(2): 108-10)

Parole chiave: metodi Bayesiani, confronti multipli, analisi esplorativa

Abstract

Epidemiological studies often involve multiple comparisons, and may therefore report many «false positive» statistically significant findings simply because of the large number of statistical tests involved. Traditional methods of adjustment for multiple comparisons, such as the Bonferroni method, may induce investigators to ignore potentially important findings, because they do not take account of the fact that some variables are of greater a priori interest than others. The Bonferroni method involves «adjusting» all of the findings to take account of the number of comparisons involved, even though the a priori evidence may be very strong for some exposures, but may be much weaker (or non-existent) for the other exposures being considered. Furthermore, the Bonferroni method only «adjusts» for estimates of statistical significance (p-values) and does not «adjust» the effect estimates themselves (e.g. odds ratios and 95% CI). Empirical Bayes and semi-Bayes methods can enable the avoidance of numerous false positive associations, and can produce effect estimates that are, on the average, more valid. In this paper, we report on a research in which we applied these methods to a case-control study of occupational risk factors for lung cancer and tested their performance.
(*Epidemiol Prev* 2008; 32(2): 108-10)

Key words: *Bayesian methods, multiple comparisons, exploratory analysis*

Introduction

Epidemiological studies often involve multiple comparisons, and may therefore report many «false positive» statistically significant findings simply because of the large number of statistical tests involved. Examples include occupational and environmental studies that typically investigate the effects of diverse exposures on several diseases, and studies of gene-environment interactions that entail the assessment of many environmental exposures in conjunction with multiple genetic polymorphisms.[1]
In these studies, the investigators usually have few or no *a priori* hypotheses about which associations might be expected (the so-called «fishing expeditions»), and although they are aware that a (computable) fraction of the resulting associations may be due to chance, they do not know which ones are due to chance (false positives) and which ones are true positive findings.
Thus, in a given study where multiple associations are tested, as the number of tests increases it becomes increasingly likely that there will be statistically significant outcomes (false positives) due to random variability, even if no real effects exist.[2] For one test, the probability of having a type I error is usually set at given number α, which is called the significance level. For n tests, the probability of having at least one type I error will be $1-(1-\alpha)^n$, and hence will increase as n increases. For instance, if we set α at 0.05, the probability of having at least one type I error for 10 tests will be 0.4, and for 100 tests will be 0.99.
Traditional methods of adjustment for multiple comparisons, such as the Bonferroni method may induce investigators to ignore potentially important findings, because they do not take account of the fact that some variables are of greater *a priori* interest than others. Thus, by decreasing the probability of type I error, they increase the probability of type II error.[3] For example, consider a case-control study of asbestos exposure and lung cancer, in which a number of other exposures are also considered. The Bonferroni method involves «adjusting» all of the find-

ings to take account of the number of comparisons involved, even though the *a priori* evidence is very strong for asbestos but may be much weaker (or non-existent) for the other exposures being considered. Furthermore, the Bonferroni method only «adjusts» for estimates of statistical significance (p-values) and does not «adjust» the effect estimates themselves (e.g. odds ratios and 95% CI) even though some of these may be biased away from the null due to random error. In epidemiological studies, Empirical Bayes (EB) and Semi-Bayes (SB) adjustment methods have been shown to be more valid approaches to the problem of multiple comparisons, particularly when the parameters to be estimated can be divided into groups within which they are similar or "exchangeable" on the basis of *a priori* knowledge.[1] Thus, Empirical Bayes and Semi-Bayes methods can enable the avoidance of numerous false positive associations, and can produce effect estimates that are, on the average, more valid.

Materials and methods

The aim of Bayesian estimation is to reduce the expected squared error of an estimator.[4] The method consists in giving a prior expectation to the estimated parameter. Then a posterior estimate is calculated as a weighted mean of the standard estimate and the prior expectation. Consequently, if the prior expectation is not too far from the true parameter, the expected squared error and the probability of type I error are reduced. EB and SB adjustment methods assume that the observed variation of the estimated parameters (e.g. odds ratios) around their global mean is larger than the variation of the true parameters.[1] The EB method aims at estimating the variation of the true parameters directly from the data, whereas the SB method specifies an *a priori* value for the variation of the true parameters so that they have a reasonable range of variation (e.g. a Var_{true} of 0.25 implies that 95% of the true relative risks are within a 7-fold range of each other[1]). The «adjustment» then consists of shrinking outlying estimates towards the geometric mean of the estimates' distribution. The larger the individual variance of the estimates, the stronger is the shrinkage, so that the shrinkage is stronger for less reliable estimates based on small numbers. The effect of shrinkage is to reduce the overall variance of the estimates.

The EB method estimates the variance of the true log odds ratios (Var_{true}) as:

$$Var_{true} = Var_{obs} - Var_{mean} \qquad [1]$$

where Var_{obs} is the observed sample variance of the log odds ratios estimates, and Var_{mean} is the mean of the estimated variances of each log odds ratio estimate. Since Var_{true} must be a positive value, Var_{obs} must be greater than Var_{mean}. If the estimated variances do not satisfy this inequality, the SB method, in which Var_{true} is set by the investigator, should be used instead of the EB method.

EB and SB adjustments can be validly used under specific conditions. Firstly, the distribution of the estimates to be adjusted must be well approximated by a log normal distribution. Secondly, if the exposures are quantitative, they must be rescaled so that the log odds ratios for a one-unit increment of exposure must be comparable. Finally, if there are prior associations between the odds ratios, these must be taken into account.[1] We have considered the simplest case in which there are no such associations or they can be neglected. We applied the EB and SB adjustment methods to a case-control study of occupational risk factors for lung cancer.[5] This study was carried out between 1990 and 1992 in two areas of Italy: Turin and the Eastern part of Veneto region. Cases (956 men and 176 women) were all individuals with incident primary lung cancer, aged less than 75 and resident in the study areas. Controls (1,253 men and 300 women) were randomly selected from the local population registries and frequency matched with cases by gender, study area and five-year age groups. Information on basic demographic details, active and passive smoking, and a lifetime occupational history was collected. In particular, the dates of beginning and ending work, as well as the job title and branch of industry, were recorded for each occupational period that lasted at least 6 months. Job titles and branches of industry were coded blind to case-control status according to the International Standard Classification of Occupations (ISCO)[6] and the International Standard Industrial Classification (ISIC),[7] respectively. These classifications, based on a maximum number of 5 and 4 digits, respectively, increase the specificity of each occupation/industry with increasing number of digits (e.g., ISCO code 93: painters, and ISCO code 93190: structural steel and ship painters). A logistic regression model was built for each job and for each industry, for men and women separately. Covariates included in the models were age, study area, and cigarette smoking status. Odds ratios of lung cancer with corre-

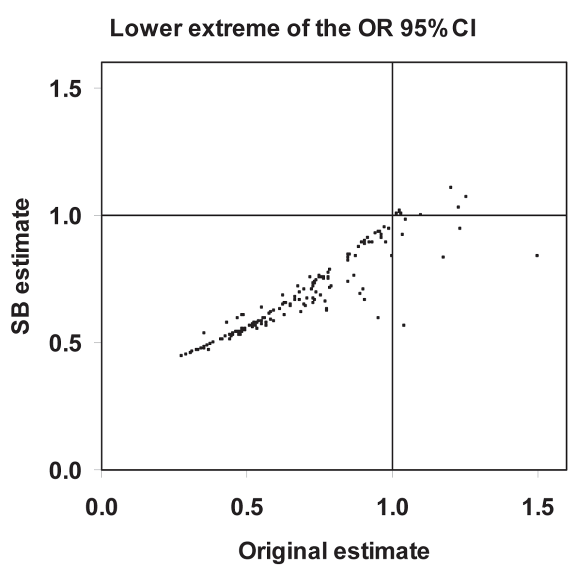

Figure 1. Scatter plot of the lower bound of the Semi-Bayes (SB) adjusted 95% confidence intervals (CI) against the lower bound of the standard 95% CI for increased odds ratios (OR) of lung cancer for different job titles, defined on the basis of 2, 3, 4 and 5 ISCO digits.[6] Men.

sponding 95% confidence intervals (CI) were estimated for all jobs and industries and SB and EB adjustments were applied to the obtained estimates. The computation of EB estimates was often impossible because the estimate of Var_{true} was negative (see equation 1 above). Even when EB adjustment was possible, the estimated Var_{true} was very small, resulting in too few statistically significant ORs after EB adjustment.[1] We therefore used SB adjustment in the analyses. In this paper, we show the effects of SB adjustment on the estimates of ORs of lung cancer for job titles among men. Analyses were conducted using SAS and R software. The codes for the SB adjustments are available upon request.

Number of digits of ISCO codes	Number of OR[a] estimates	Number of statistically significantly increased ORs[a]			
		expected*	observed		
			no adjustment	Bonferroni adjustment	semi-Bayes adjustment
1	10	0.25	1	0	1
2	61	1.53	4	0	4
3	122	3.05	6	0	2
4-5	154	3.85	5	0	2

[a] OR, odds ratio *(based on $\alpha = 0.05$)"

Table 1. Frequencies of statistically significant increased risks of lung cancer for job titles (defined on the basis of 1 to 5 digit ISCO codes[6]) before and after Bonferroni and Semi-Bayes adjustments. Men.

Results

Figure 1 shows the lower bound of the confidence intervals of the SB-adjusted ORs of lung cancer against the lower bound of the confidence intervals of the standard (not SB-adjusted) ORs for job titles associated with an OR above 1. The figure thus depicts the changes in the statistical significance of risk estimates produced by SB adjustment. In the lower left quadrant are increased risk estimates that did not reach statistical significance with either the standard unadjusted estimation nor after SB adjustment (probable true negatives). In the lower right quadrant are risk estimates that did reach statistical significance in the standard unadjusted estimation, but lost it after SB adjustment (probable false positives). The upper right quadrant contains risk estimates that were statistically significant before and after SB adjustment (probable true positives). As expected, the upper left quadrant is empty, i.e. there were no increased risk estimates that gained statistical significance through SB adjustment (probable false negatives). Both the probable elimination of false positives and the shrinkage of estimates produced by the SB adjustment can be observed in Figure 1. The probable elimination of false positives is shown by the fact that only some of the statistically significant findings remained so after SB adjustment. Shrinkage is shown by the fact that SB-adjusted estimates depart from a 45-degree line to be pulled towards the centre of the distribution. Table 1 shows the numbers of statistically significant estimates expected by chance, observed without adjustment for multiple comparisons and observed after Bonferroni correction and SB adjustment. Bonferroni adjustment is strongly penalising since it removes all statistically significant estimates. SB adjustment, on the other hand, can be used to identify the most robust findings for further investigation. For 3-digit ISCO codes, for example, if a decision rule was based on an elevated OR with $p < 0.05$, in our study we would further investigate 6 occupations among men using unadjusted results, versus 2 using SB adjustment.

Discussion

SB and EB shrinkage methods are valid and robust methods for addressing the problem of multiple comparisons in occupational studies. Our findings show that the SB method, by reducing the variability of the estimates, may decrease the number of false positive findings, while not eliminating all positive findings. A potentially interesting application of SB method in occupational epidemiology is the analysis of the numerous situations of exposure identified by the combination of ISCO and ISIC codes. Rothman[8] pointed out that deciding to not making Bonferroni adjustments for multiple comparisons may be preferable because it leads to fewer errors of interpretation, given that the observed data are real observations on a natural phenomenon and not random numbers. Moreover, when dealing with inferential investigations, statistical significance is just one of the several components of causal inference. However, Bayesian adjustments for multiple comparisons avoid the problems associated with the Bonferroni method, while producing effect estimates that are, on the average, more valid than the unadjusted estimates. Furthermore, they can be easily carried out using a standard statistical software. Thus, their use is recommended in exploratory analyses.

Acknowledgements. The study was conducted within the framework of projects partially supported by the Italian Association for Cancer Research (AIRC), the Compagnia San Paolo/FIRMS and the Regione Piemonte. Neil Pearce's work on this paper was supported by the Progetto Lagrange, Fondazione CRT / ISI.

Conflitti di interesse: nessuno

References

1. Steenland K, Bray I, Greenland S, Boffetta P. Empirical Bayes adjustments for multiple results in hypothesis-generating or surveillance studies. *Cancer Epidemiol Biomarkers Prev* 2000; 9: 895-903.
2. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* 2002; 2: 8.
3. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ* 1998; 316: 1236-38.
4. Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000; 29: 158-167.
5. Richiardi L, Boffetta P, Simonato L, et al. Occupational risk factors for lung cancer in men and women: a population-based case-control study in Italy. *Cancer Causes Control* 2004; 15: 285-94.
6. International Standard Classification of Occupation. Geneva: International Labour Office, 1968.
7. International Standard Industrial Classification. New York: United Nations Publications (ST/STAT/M.4/Rev. 2/Add. 1, Sales No.: E.71.XVII.8), Publishing Service United Nations, 1971.
8. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; 1: 43-6.