KM 26(1987)
pag 109-120

EMPIRICAL BAYES METHODS IN CLINICAL TRIALS META-ANALYSIS

Th. Stijnen and J.C. van Houwelingen *)

Abstract

In this paper we discuss empirical Bayes methods for combining evidence from a series of clinical experiments comparing two treatments. The approach is based on a random effects model for the treatment effects. We describe two nonparametric empirical Bayes procedures for estimating the effect sizes in the individual trials. The empirical Bayes estimates should have smaller mean squared error than the observed effect sizes. Moreover, the empirical Bayes estimates exhibit less variability and are more comparable to each other than the original estimates.

---

*)
Department of Medical Statistics
University of Leiden
P.O. Box 9512
2300 RA Leiden
The Netherlands
tel. 071-277036/277025

1. Introduction

Meta-analysis is defined here as the statistical analysis with the purpose of combining and integrating the findings of several research studies. Such analyses become increasingly popular in medical research where information on the efficacy of a treatment is available for a number of comparative clinical trials. Most meta-analyses in the medical field concentrate on estimating and testing the common treatment effect, which is then assumed to be equal for all trials, and little attention is paid to possible inhomogeneity in effect sizes between trials. DerSimonian and Laird (1986) discuss a random effects model that adequately incorporates inhomogeneity of treatment effects. In this paper we adopt the same model and consider the use of empirical Bayes methods in order to construct improved estimators of the effect sizes in the individual trials. Several authors, among others Champney (1983) and Hedges and Olkin (1985), have proposed the use of empirical Bayes methods in meta-analyses. However, they assume a functional form for the distribution of the effect sizes. In this paper we describe two empirical Bayes methods which can be used if this distribution is left completely unspecified.

Suppose we want to combine information from a series of k comparative clinical trials in which an experimental treatment is compared with a control treatment. Let the true difference in efficacy between the two treatments in the i^{th} $(i=1,\ldots,k)$ trial be characterized by some unknown parameter θ_i. For example, when the outcome of the treatments is measured by some continuous variable X, the true effect size in the i^{th} trial could be defined as $\theta_i=\mu_t-\mu_c$, with μ_t and μ_c the population means of X for the experimental and control treatment, respectively. If the outcome is measured by a dichotomous variable, say dead or alive, common choices for θ_i are the risk difference, the relative risk or the odds ratio.

In each trial the true effect size θ_i is estimated by the observed effect size y_i, which has standard error σ_i. For instance, y_i is the observed difference $\bar{x}_t - \bar{x}_c$ between the sample means of the outcome variable X, and σ_i^2 is estimated by the sum of the

squares of the estimated standard errors of the means. If it is assumed that the true treatment effects θ_i all are equal, the common effect size can be estimated by some weighted average of the y_i's in the usual way. However, differences in for instance the support of clinical care and minor differences in the treatments, such as differences in dose or duration, can lead to varying effect sizes between trials. In this paper we consider empirical Bayes (e.B.) methods in order to make improved estimates of the effect sizes of the individual trials. These e.B. estimates will show considerably less dispersion than the original observed effect sizes.

We concentrate in this paper on empirical Bayes estimators that are non-parametric, i.e. the distribution of the effect sizes is completely unspecified. In the next section we specify the statistical model and introduce the empirical Bayes approach. In section 3 we describe two non-parametric e.B. estimators. The use of these methods is illustrated in section 4 with an examples of a series of clinical trials reported in the medical literature. We finish the paper with a short discussion in section 5.

2. The empirical Bayes approach

We adopt the following model.

(a) $\theta_1,\ldots,\theta_k$ constitute a random sample of treatment effects from a population of possible treatment evaluations. The distribution of effect sizes is denoted by G, which has mean μ and standard deviation σ_θ.

(b) Given θ_i, y_i has a normal distribution with mean θ_i and standard deviation σ_i.

Before introducing empirical Bayes estimators it is instructive to consider for a moment the situation in which the distribution G and the σ_i's are known. In that case one could estimate θ_i with the expectation of the posterior distribution of θ_i given the observation y_i:

$$d_{G,i}(y_i) = E\{\theta_i \| y_i\} = \frac{\int \theta_i \sigma_i^{-1} h((y_i-\theta_i)/\sigma_i)\, dG(\theta_i)}{\int \sigma_i^{-1} h((y_i-\theta_i)/\sigma_i)\, dG(\theta_i)}, \qquad (2.1)$$

where h denotes the standard normal density. The estimator $d_{G,i}$ is the usual (non-empirical) Bayes estimator with respect to G under squared error loss. $d_{G,i}$ is optimal in the sense that it minimizes $E\{d_{G,i}(Y_i)-\theta_i\}^2$. Suppose for instance that G is a normal distribution with known mean μ and standard deviation σ_θ. Then (2.1) reduces to (cf. DeGroot (1970), theorem 9.5.1)

$$d_{G,i}(Y_i) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_i^2} Y_i + \frac{\sigma_i^2}{\sigma_\theta^2 + \sigma_i^2} \mu \ . \qquad (2.2)$$

So in this case the Bayes estimator of the effect size θ_i is a weighted mean of the observed effect size y_i and the mean μ of the prior distribution of effect sizes.

Now let us turn back to the case that G is unknown. In the empirical Bayes approach the observations $y_1,\ldots,y_n$ are used to construct an estimator d_i of $d_{G,i}$. If some functional form for the distribution G is assumed, the construction of e.B. estimators is very straightforward. For instance, suppose that G is normal with unknown mean μ and variance σ_θ^2. Then there are several methods to construct estimators $\hat{\mu}$ and $\hat{\sigma}_\theta^2$ for μ and σ_θ^2, respectively. One of the methods used by DerSimonian and Laird (1986) is as follows. Estimate μ by the weighted average

$$\hat{\mu} = \sum_{i=1}^{n} w_i y_i \ / \ \sum_{i=1}^{n} w_i \ , \qquad (2.3)$$

where the weights w_i are equal to $\hat{\sigma}_i^{-2}$, and estimate σ_θ^2 by

$$\hat{\sigma}_\theta^2 = \max\{\ 0, \frac{\Sigma\ (\ w_i(y_i - \hat{\mu})^2) - (n-1)}{\Sigma\ w_i - (\ \Sigma\ w_i^2 \ / \ \Sigma\ w_i)}\ \} \ . \qquad (2.4)$$

Substituting these estimates in (2.2) we obtain the following e.B. estimator for the effect size θ_i in the i^{th} trial.

$$d_{G,i}(Y_i) = \frac{\hat{\sigma}_\theta^2}{\hat{\sigma}_\theta^2 + \hat{\sigma}_i^2} Y_i + \frac{\hat{\sigma}_i^2}{\hat{\sigma}_\theta^2 + \hat{\sigma}_i^2} \hat{\mu} \ . \qquad (2.5)$$

Much work has been done in constructing and investigating asymptotic properties of nonparametric e.B. estimators for the

case that all σ_i's are known and equal (see Singh (1979), e.g.). Very little attention has been paid to the case of unequal σ_i's. In the next section we describe two non-parametric e.B. estimators which are applicable in that case.

3. Two nonparametric empirical Bayes estimators

Method I

The first e.B. estimator is based on nonparametric estimation of the distribution of effect sizes G. Once G is estimated by $\hat{G}$, an e.B. estimator can be obtained by taking the Bayes estimator with respect to $\hat{G}$, i.e. substituting $\hat{G}$ in (2.1). In order to estimate G we use a general method due to Laird (1978) for nonparametric maximum likelihood estimation of a mixing distribution. She considered the following model. Let $\theta_1,\ldots,\theta_n$ be a random sample from an unknown distribution G. The θ_i's are not observed. Let further $y_1,\ldots,y_n$ be independent observations such that, conditionally on θ_i, y_i is distributed with density $h_i(y_i\|\theta_i)$. Hence, marginally y_i has density

$$f_i(y_i\|G) = \int h_i(y_i\|\theta)\, dG(\theta) \,. \tag{3.1}$$

The parametric form of $h_i(\cdot\|\cdot)$ is assumed to be known. Laird showed that, under some weak regularity conditions, the nonparametric maximum likelihood estimator $\hat{G}$ of G is a discrete distribution with a finite number of atoms. Once it is known that $\hat{G}$ is discrete with k steps, $\hat{G}$ is easily computed with the EM algorithm (Dempster e.a., 1977), e.g.. When k is unknown, one approach is to start with a large (close to n) value of k, say k_0, and take k to be the number of distinct steps with positive probability. (Choosing k_0 to be greater than k causes no problems if the EM algorithm is used, since that is capable to converge to a point on a ridge.) We apply this method in order to estimate the distribution G of effect sizes, taking $h_i(y_i\|\theta_i)$ to be equal to $\hat{\sigma}_i^{-1}h((y_i-\theta_i)/\hat{\sigma}_i)$, with h the standard normal density. (In order to simplify notation, we shall write σ_i in stead of $\hat{\sigma}_i$ in the sequel.) The EM algorithm used to compute $\hat{G}$ is very simple programmable as follows. Let $\theta_1^{(s)},\ldots,\theta_k^{(s)}$ be the atoms of $\hat{G}$ at

the s^{th} iteration, and let $\pi_1^{(s)},\ldots,\pi_k^{(s)}$ be the corresponding probabilities. Then the $(\theta_j^{(s+1)}, \pi_j^{(s+1)})$, $j=1,\ldots,k$ at the next iteration are determined as follows. First compute z_{ij} $(i=1,\ldots,n;\ j=1,\ldots k)$ by

$$z_{ij} = \pi_j^{(s)}\sigma_i^{-1}h((y_i-\theta_j^{(s)})/\sigma_i) \;/\; \sum_{j=1}^{k}\{\pi_j^{(s)}\sigma_i^{-1}h((y_i-\theta_j^{(s)})/\sigma_i)\}\ .$$

Then the atoms and probabilities of $\hat{G}$ at the $(s+1)^{th}$ iteration are determined by

$$\pi_j^{(s+1)} = \sum_{i=1}^{n} z_{ij}/n \quad \text{and} \quad \theta_j^{(s+1)} = \frac{\sum_{i=1}^{n} z_{ij}y_i/\sigma_i^2}{\sum_{i=1}^{n} z_{ij}/\sigma_i^2}\ .$$

A good set of starting values is a uniform distribution on an equally spaced grid between the minimum and the maximum of the y_i's.

Let (θ_j^*,π_j^*), $j=1,\ldots k$, be the estimate $\hat{G}$ of G obtained after a sufficient number of iterations. Then the following e.B. estimator can be obtained by substituting $\hat{G}$ in (2.1).

$$d_i(y) = \frac{\sum_{j=1}^{k} \theta_j^*\,\pi_j^*\,h((y-\theta_j^*)/\sigma_i)}{\sum_{j=1}^{k} \pi_j^*\,h((y-\theta_j^*)/\sigma_i)}\ . \tag{3.2}$$

Method II

The second e.B. estimator we want to introduce is based on direct estimation of $d_{G,i}$. This method has the special feature that, in order to estimate the effect size in the i^{th} trial, it only makes use of the observations y_j of trials with $\sigma_j < \sigma_i$. The idea is as follows. From (2.1) one can easily derive that $d_{G,i}$ can be written as

$$d_{G,i}(y) = y + \sigma_i \frac{f_i'(y\|G)}{f_i(y\|G)}\ , \tag{3.3}$$

where $f_i(\cdot\|G)$ is the marginal density of y_i given by (3.1), and $f_i'(\cdot\|G)$ is its derivative. Let n_i denote the number of σ_j's with $\sigma_j < \sigma_i$, then $f_i(\cdot\|G)$ can be estimated by

$$\hat{f}_i(y) = \frac{1}{n_i} \sum_{j:\sigma_j<\sigma_i} (\sigma_i^2-\sigma_j^2)^{-\frac{1}{2}} \; h\left[\frac{y-y_j}{(\sigma_i^2-\sigma_j^2)^{\frac{1}{2}}} \right], \tag{3.4}$$

and $f_i'(\cdot\|G)$ is estimated by the derivative $\hat{f}_i'(\cdot)$ of $\hat{f}_i(\cdot)$. Now, an e.B. estimator is given by

$$d_i(y) = y + \sigma_i \frac{\hat{f}_i'(y)}{\hat{f}_i(y)} . \tag{3.5}$$

In order to show that (3.5) is a reasonable estimator of $d_{G,i}$, we let see that the estimators $\hat{f}_i(y)$ and $\hat{f}_i'(y\|G)$ are unbiased estimators of $f_i(y\|G)$ and $\hat{f}_i'(y\|G)$, respectively. Let $\beta_j=(\sigma_i^2-\sigma_j^2)^{\frac{1}{2}}$, then the expectation of the summand corresponding to σ_j in (3.4) can be written as

$$E\left[\frac{1}{\beta_j} h((y-y_j)/\beta_j) \right] = E_G\left[E\left[\frac{1}{\beta_j} h((y-y_j)/\beta_j)\|\theta_j \right] \right] =$$

$$E_G\left[\int \frac{1}{\beta_j} h((y-u)/\beta_j) \frac{1}{\sigma_j} h((u-\theta_j)/\sigma_j)\, du \right] = \tag{3.6}$$

The expression between square brackets in (3.6) is recognized as the convolution of a $N(0,\beta_j^2)$ and a $N(\theta_j,\sigma_j^2)$ distribution. Hence (3.6) is equal to (cf. (3.1))

$$E_G\left[\frac{1}{\sigma_i} h((y-\theta_j)/\sigma_i) \right] = \int \frac{1}{\sigma_i} h((y-\theta_j)/\sigma_i)\, dG(\theta_j) = f_i(y\|G)$$

So each summand in (3.4) is an unbiased estimator of $f_i(.\|G)$. Hence $\hat{f}_i$ is unbiased. The same follows analogously for $\hat{f}_i'$. Unfortunately, the e.B. estimator given by (3.5) has the unattractive property that it is not necessarily a nondecreasing function of y. It is known that the class of monotone estimators is essentially complete if the conditional distribution of y_i given θ_i has monotone likelihood ratio in θ (Berger (1980), theorem 8.7). Hence the e.B. estimator (3.5) is not admissible.

Therefore we smooth d_i using the following monotonisation procedure due to van Houwelingen (1973). Let $F_i^{-1}(\cdot\|\theta)$ denote the inverse of the cumulative distribution function of y_i given θ, and let $\alpha(\theta)$ be defined as

$$\alpha(\theta) = \int I[d_i(y)<\theta]\, f_i(y\|\theta)\, dy\,, \tag{3.7}$$

where I[A] denotes the indicator function of a set A. Then the e.B. estimator d_i^* given by

$$d_i^*(y) = \sup\{\theta\colon F_i^{-1}(\alpha(\theta)\|\theta) < y\} \tag{3.8}$$

is nondecreasing in y and has smaller risk than d_i, i.e. $E(d_i^*(y_i)-\theta_i)^2 < E(d_i(y_i)-\theta_i)^2$. For a proof the reader is referred to van Houwelingen (1973,1977) or Brown e.a. (1977).
In the next section we apply the above methods to the data of a series of clinical trials from the medical literature.

4. Example

Our example of the use of the e.B. methods from section 3 concerns a series of 25 randomized clinical trials reported by Collins and Langman (1985). In each trial a group of patients with acute upper gastrointestinal bleeding treated with a histamine H_2 antagonist was compared with a control group treated with a placebo. The data consist of the number of patients in the treatment and control group, n_T and n_C, and the number of patients with persistent or recurrent bleeding in each of the groups, d_T and d_C. As a measure of the treatment effect we choose the (natural) logarithm of the odds ratio $\theta = \log\{p_T(1-p_C)/(p_C(1-p_T))\}$, where p_T and p_C are the probabilities of persistent or recurrent bleeding in the treatment and control group, respectively. The effect size θ in each trial is estimated by (Fleiss (1981), p. 67)

$$y = \log\left[\frac{(d_T+.5)/(n_T-d_T+.5)}{(d_C+.5)/(n_C-d_C+.5)}\right]. \tag{4.1}$$

Given θ, the observed log odds ratio y is approximately normally distributed with mean θ and estimated standard error (Fleiss (1981), p. 67)

$$\hat{\sigma} = \left[\frac{1}{d_T+.5} + \frac{1}{n_T-d_T+.5} + \frac{1}{d_C+.5} + \frac{1}{n_C-d_C+.5} \right]^{1/2} . \qquad (4.2)$$

Collins and Langman report that the usual chi square test for heterogeneity indicates that there is some heterogeneity ($X^2_{[24]}=38.1$, p=0.02) and they compute a 'typical' odds ratio of .89 with an approximate confidence interval of (.73 , 1.08).

In table 1 for each trial the data are given, together with the observed log odds ratio and the e.B. estimates obtained with the two methods described in the previous section. For illustration, also the parametric e.B. estimator given by (2.3)-(2.5), which is based on a normal distribution of treatment effects, is given. The estimates of μ and σ^2_θ were $\hat{\mu}$ = -.103 and $\hat{\sigma}^2_\theta$ = .100.

In table 1 the trials have been sorted by the size of the estimated standard error of the observed log odds ratio. Obviously all e.B. estimators exhibit considerably less variability than the observed log odds ratios. The Pearson correlation coefficient between the two series of nonparametric e.B. estimates is equal to r=.89. The correlation between the parametric and the nonparametric e.B. estimators is equal to .90 and .88 for method I and II, respectively. The correlation coefficients of the two nonparametric and the parametric e.B. estimates in table 1 with the original observed treatment effects y_i are .79, .88 and .83, respectively.

When considering the e.B. estimates produced with method II, one has to remember that for each trial the estimate is based only on the observed log odds ratios of the trials with smaller $\hat{\sigma}_i$. Therefore the method II estimates on the first few lines of table 1 have to be rather unreliable.

The maximum likelihood estimate of the prior distribution needed for method I has two mass points (θ^*_1, θ^*_2) = (-.98, .003) with probabilities (π^*_1, π^*_2) = (.17, .83). Given the estimated prior distribution, the e.B. estimates are easily computed with formula (3.2). The estimated distribution of the effect sizes suggest that in the majority of the trials there is no treatment effect

at all, but there seems to be a minority of trials in which there might be a substantial treatment effect.

n_T	d_T	n_C	d_C	observed log odds ratio y_i	estimated standard error σ_i	empirical Bayes estimates		
						nonparametric method I	nonparametric method II	parametric (normal prior)
259	50	260	51	-.02	.22	.00		-.05
153	44	132	30	.31	.27	.00	.11	.13
106	16	107	15	.08	.38	-.00	.17	-.03
78	14	80	21	-.47	.38	-.15	-.06	-.25
51	16	54	15	.17	.42	-.00	.02	-.00
56	11	53	12	-.18	.46	-.04	-.08	-.13
50	8	50	16	-.87	.48	-.51	-.30	-.34
40	9	48	11	-.02	.50	-.03	-.12	-.08
33	12	36	10	.38	.51	-.00	.08	.03
46	5	47	12	-.98	.56	-.48	-.38	-.31
31	6	29	12	-1.00	.57	-.50	-.54	-.32
33	6	39	7	.02	.60	-.04	-.04	-.07
36	11	26	5	.57	.60	-.01	.08	.04
21	10	19	8	.21	.62	-.03	.05	-.04
20	7	20	8	-.20	.64	-.09	-.19	-.12
45	9	43	3	1.10	.67	-.00	.20	.12
34	3	31	13	-1.90	.67	-.80	-.68	-.43
24	4	24	5	-.25	.71	-.11	-.09	-.13
14	2	15	6	-1.20	.86	-.35	-.44	-.24
15	2	14	3	-.50	.92	-.17	-.20	-.14
18	5	12	1	1.14	1.00	-.04	-.05	.01
10	1	9	4	-1.60	1.08	-.35	-.43	-.22
10	3	11	1	1.18	1.09	-.05	.18	-.00
18	0	19	3	-2.10	1.55	-.28	-.42	-.18
14	0	14	0	.00	2.03	-.15	-.26	-.10

TABLE 1. Observed log odds ratios and empirical Bayes estimates of the true log odds ratios for the series of clinical trials of Collins and Langman.

5. Discusssion

In this paper we described two nonparametric e.B. methods for estimating the treatment sizes in the individual trials of a series of clinical trials comparing the same experimental treatment with a control treatment. Both methods have the property of shrinking the observed effect sizes to some weighted average, and therefore considerably reduce the variability in the observed treatment effects. Based on nonparametric maximum likelihood estimation of the underlying distribution of the effect sizes, method I has the attractive property of giving an explicit

estimate of this distribution. However, this method seems to be rather rigid and might shrink to drastically. We applied this method to several other series of clinical trials reported in the medical literature. In all cases the estimated distribution had at most two mass points. Notice that (3.2) implies that the e.B. estimate always lies in between these two mass points. For instance, in the example of section 4, this method cannot produce estimates that are positive. The second method (II) does not have this disadvantage, but this method has the unattractive property of not using all the available data. Since only data of larger trials are used, this disadvantage is most serious for the large trials. However, since large trials can give a precise estimate of their treatment effect on their own, there is less need for improved estimates for large trials. Method II seems to behave somewhat more flexible than method I. It produces in our example estimates which are higher correlated with the original observed effect sizes than the estimates produced by method I.

Since no theoretical results are known, the nonparametric e.B. proposed in this paper have to be regarded as exploratory. Deriving analytic results for method I seems to be very difficult. For method II, at least for the non-monotonized version, asymptotic results concerning the variance of d_i and the rate of convergence of d_i to $d_{G,i}$, can be derived more easily. Much research has to be done, for instance to construct confidence intervals for the effect sizes of the individual trials, or to incorporate covariate information into the model.

References

Berger,J.O. (1980)
Statistical Decision Theory; Foundations, Concepts and Methods. Springer Series in Statistics, Springer Verlag, New York.

Brown,L.D., Cohen,A., Strawderman,W.E. (1976)
A complete class theorem for strict monotone likelihood ratio with applications. Annals of Statistics 4, p. 712-722.

Collins,R. and Langman,M. (1985)
Treatment with histamine H_2 antagonists in acute upper gastrointestinal hemorrhage. The New England Journal of Medicine, sept. 12, p.660-666.

Champney,T.F. (1983)
Adjustments for selection: Publication bias in quantitative research synthesis. Doctoral dissertation, The University of Chicago.

DeGroot,M.H. (1970)
Optimal Statistical Decisions, McGraw Hill Book Company, New York.

Dempster,A.P., Laird,N.M. and Rubin,D.B. (1977)
Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, ser. B, 39 p.1-38 (with discussion)

DerSimonian,R and Laird,N.M. (1986)
Meta-analysis in clinical trials. Controlled Clinical Trials 7, p. 177-188.

Fleiss,J.L. (1981)
Statistical Methods for Rates and Proportions. Second edition. John Wiley & Sons, New York.

Hedges,L.V. and Olkin,I. (1985)
Statistical Methods for Meta-analysis, Academic Press, Inc.

Houwelingen,J.C. van (1973)
On empirical Bayes rules for the continuous exponential family. Doctoral thesis, University of Utrecht.

Houwelingen,J.C. van (1977)
Monotonizing empirical Bayes estimators for a class of discrete distributions with monotone likelihood ratio. Statistica Neerlandica 31, p. 95-104.

Laird,N.M. (1978)
Nonparametric maximum likelihood estimation of a mixing distribution. JASA, vol. 73, p.805-811.

Singh,R.S. (1979)
Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. Annals of Statistics 7, p. 890-902.

Ontvangen: 11-06-1987
Geaccepteerd: 22-07-1987