



Statistics

Efficiency of bridging between related dose finding studies

Efficacité de l'association croisée entre études corrélées de dosages

John O'Quigley

Laboratoire de statistique théorique et appliquée, université Pierre-et-Marie-Curie (Paris-6), 4, place Jussieu, 75252 Paris cedex 05, France

ARTICLE INFO

Article history:

Received 20 January 2012

Accepted after revision 31 January 2013

Available online 18 June 2013

Presented by Paul Deheuevels

ABSTRACT

The statistical problem of bridging is closely associated with the problem of heterogeneity in dose-finding studies. There are some distinctive features in the case of bridging which need to be considered if efficient estimation of the maximum tolerated dose (MTD) is to be accomplished. The case of two distinct populations is considered. Extensions to several populations are, at least in principle, straightforward although, in practice, likely to be awkward and infrequently encountered. The goal is to make efficient use of information gained in one study in the context of a second study. Since working models are typically misspecified it is not possible to just add a further parameter to deal with an added source of variability.

© 2013 Published by Elsevier Masson SAS on behalf of Académie des sciences.

R É S U M É

Le problème statistique de l'association croisée est intimement lié au problème de l'hétérogénéité dans les études de dosages. Quelques aspects spécifiques dans le cas de l'association croisée sont à considérer si une estimation efficace de la dose maximum tolérée (DMT) doit être obtenue. Le cas de deux populations distinctes est envisagé. Les extensions au cas de plusieurs populations sont, au moins en principe, directes, mais en pratique vraisemblablement gauches et peu fréquentes. Le but est d'utiliser de façon efficace l'information tirée d'une première étude dans le contexte d'une seconde. Comme les modèles sont en général mal spécifiés, il n'est pas possible de se contenter d'ajouter un paramètre pour traiter une nouvelle source d'aléa.

© 2013 Published by Elsevier Masson SAS on behalf of Académie des sciences.

1. Dose finding for heterogeneous groups

Assume that we have available k ordered doses; $d_1 < d_2 < \dots < d_k$. The probability of toxicity, $R(d_i)$, is such that $R(d_i) < R(d_j)$ whenever $i < j$. The maximum tolerated dose (MTD) is that dose $d_0 \in \{d_1, \dots, d_k\}$ such that, for some given $\theta \in (0, 1)$,

$$|R(d_0) - \theta| < |R(d_\ell) - \theta|, \quad \ell = 1, \dots, k; \quad d_\ell \neq d_0. \quad (1)$$

Y_j takes the value 1 in the case of a toxic response for the j th entered subject ($j = 1, \dots, n$) and 0 otherwise. The dose for the j th entered subject, X_j is viewed as random taking values $x_j \in \{d_1, \dots, d_k\}$; $j = 1, \dots, n$. Thus $\Pr(Y_j = 1 | X_j = x_j) = R(x_j)$. We suppose that

$$R(x_j) = \Pr(Y_j = 1 | X_j = x_j) = E(Y_j | x_j) = \psi(x_j, a) \quad (2)$$

E-mail address: john.oquigley@upmc.fr.

for some one parameter model $\psi(x_j, a)$ and a defined on the set \mathcal{A} . For every a , $\psi(x, a)$ should be monotone increasing in x and, for any x , $\psi(x, a)$ should be monotone in a . For every d_i there exists some $a_i \in \mathcal{A}$ such that $R(d_i) = \psi(d_i, a_i)$. Here, we take $\psi(d_i, a) = \alpha_i^{\exp(a)}$, where i runs from 1 to k , and where $0 < \alpha_1 < \dots < \alpha_k < 1$ and $-\infty < a < \infty$.

The true mechanism generating the observations can be quite removed from our working model overall, but, close to our target, the true situation and our working model coincide. For the six levels studied in the simulations by O'Quigley, Pepe and Fisher [3] the working model had $\alpha_1 = 0.05$, $\alpha_2 = 0.10$, $\alpha_3 = 0.20$, $\alpha_4 = 0.30$, $\alpha_5 = 0.50$ and $\alpha_6 = 0.70$. Once a model has been chosen and we have data in the form of the set $\Omega_j = \{y_1, x_1, \dots, y_j, x_j\}$, the outcomes of the first j experiments we obtain estimates $\hat{R}(d_i)$ ($i = 1, \dots, k$) of the true unknown probabilities $R(d_i)$ ($i = 1, \dots, k$) at the k dose levels. Denote the set of outcomes of the first j subjects as Ω_j . Given this information, a posterior distribution for a is denoted by $f(a, \Omega_j)$ from which we induce a posterior distribution for $\psi(d_i, a)$, $i = 1, \dots, k$, leading to summary estimates of the toxicity probabilities at each level. Specifically;

$$\hat{R}(d_i) = \int_{a \in \mathcal{A}} \psi(d_i, a) f(a, \Omega_j) da, \quad i = 1, \dots, k. \tag{3}$$

We can take $\int_{\mathcal{A}} \psi(d_i, u) g(u) du = \theta$, an integral equation providing the starting dose. Given the set Ω_j and the log-likelihood $\mathcal{L}_j(a)$ the posterior density for a is:

$$f(a|\Omega_j) = H_j^{-1} \exp\{\mathcal{L}_j(a)\} g(a); \quad H_j = \int_0^\infty \exp\{\mathcal{L}_j(u)\} g(u) du. \tag{4}$$

The dose $x_{j+1} \in \{d_1, \dots, d_k\}$ assigned to the $(j + 1)$ th included patient is the dose minimizing the Euclidean distance between θ and $\int \psi\{x_{j+1}, u\} f(u|\Omega_j) du$. We can extend the single model to a class of models $\psi_m(x_j, a)$ for $m = 1, \dots, M$ where there are a total of M possible models. In particular, we might consider [2]

$$\psi_m(d_i, a) = \alpha_{mi}^{\exp(a)} \quad (i = 1, \dots, k; m = 1, \dots, M) \tag{5}$$

where $0 < \alpha_{m1} < \dots < \alpha_{mk} < 1$ and $-\infty < a < \infty$, as an immediate generalization of the single model. Any prior information concerning the plausibility of each model is denoted $\pi(m)$, $m = 1, \dots, M$, where $\pi(m) \geq 0$ and where $\sum_m \pi(m) = 1$. In the simplest case where each model is weighted equally, we would take $\pi(m) = 1/m$. The logarithm of the likelihood is, up to a constant term:

$$\mathcal{L}_{mj}(a) = \sum_{\ell=1}^j y_\ell \log \psi_m(x_\ell, a) + \sum_{\ell=1}^j (1 - y_\ell) \log(1 - \psi_m(x_\ell, a)). \tag{6}$$

Under model m , we obtain a summary value of the parameter a , in particular the maximum of the posterior mode and we refer to this as \hat{a}_{mj} . Given the value of \hat{a}_{mj} under model m we have an estimate of the probability of toxicity at each dose level d_i via: $\hat{R}(d_i) = \psi_m(d_i, \hat{a}_{mj})$ ($i = 1, \dots, k$). On the basis of this formula, and having taken some value for m , the dose to be given to the $(j + 1)$ th patient, x_{j+1} is determined. Thus, we need some value for m and we make use of the posterior probabilities of the models given Ω_j . Denoting these posterior probabilities by $\pi(m|\Omega_j)$, then:

$$\pi(m|\Omega_j) = \frac{\pi(m) \int_{-\infty}^\infty \exp\{\mathcal{L}_{mj}(u)\} g(u) du}{\sum_{m=1}^M \pi(m) \int_{-\infty}^\infty \exp\{\mathcal{L}_{mj}(u)\} g(u) du}. \tag{7}$$

One useful class of models, studied in [1] and [6] in the case that $M = 2$, takes the model for one reference group as $\psi(d_i, a)$ and, for the second group, considers a shift in the index i so that i' for the second group can be expressed via $i' = i + t$. When there are no group effects then $t = 0$ and a value $t = 1$ for instance means that the second group shares the same probabilities of toxicity as the first but “shifted” by one level. For this reason the model is sometimes called the shift model. More generally, we write:

$$\log \psi(d_i, a) = \exp(a) \log \alpha_{\phi(i)}; \quad \phi(i) = i + zh_i(t) \tag{8}$$

where

$$h_i(t) = tI(1 \leq i + t \leq k) + kI(i + t > k) + I(i + t < 1), \quad t = 0, 1, 2, \dots \tag{9}$$

In the case that $M = 2$ a straightforward extension of condition 7 of [1], applied to the groups separately, allows us to claim consistency of the model in terms of correct identification of the MTD for large samples.

2. Efficiency gains

Alternatively, in place of Δ , we include a regression coefficient b . We have n_1 observations in the first group, n_2 in the second, so that $n = n_1 + n_2$. The pair $(\hat{a}_{1n}, \hat{b}_{1n})$ then maximizes:

$$\prod_{i=1}^{n_1} \psi_1(x_i, a)^{y_i} \{1 - \psi_1(x_i, a)\}^{1-y_i} \times \prod_{i=n_1+1}^n \psi_2(x_i, a, b)^{y_i} \{1 - \psi_2(x_i, a, b)\}^{1-y_i}.$$

Next, the pair $(\hat{a}_{2n}, \hat{b}_{2n})$ maximizes separately:

$$\prod_{i=1}^{n_1} \psi_1(x_i, a)^{y_i} \{1 - \psi_1(x_i, a)\}^{1-y_i} \quad \text{and} \quad \prod_{i=n_1+1}^n \psi_1(x_i, b)^{y_i} \{1 - \psi_1(x_i, b)\}^{1-y_i}.$$

We need $\psi_1(\cdot)$ in the calculation of both \hat{a}_{2n} and \hat{b}_{2n} . Introduce:

$$s_{11}(x, y, a) = y \log\{\psi_1(x, a)\} + (1 - y) \log\{1 - \psi_1(x, a)\};$$

and

$$s_{22}(x, y, a, b) = y \log\{\psi_2(x, a, b)\} + (1 - y) \log\{1 - \psi_2(x, a, b)\}.$$

Write their first-order partial derivatives as:

$$s_{11}(x, y, a) = \frac{\partial s_1}{\partial a}, \quad s_{21}(x, y, a, b) = \frac{\partial s_2}{\partial a} \quad \text{and} \quad s_{22}(x, y, a, b) = \frac{\partial s_2}{\partial b}.$$

The pair $(\hat{a}_{1n}, \hat{b}_{1n})$ satisfies the following estimating equations:

$$\begin{cases} \sum_{i=1}^{n_1} s_{11}(x_i, y_i, \hat{a}_{1n}) + \sum_{i=n_1+1}^n s_{21}(x_i, y_i, \hat{a}_{1n}, \hat{b}_{1n}) = 0, \\ \sum_{i=n_1+1}^n s_{22}(x_i, y_i, \hat{a}_{1n}, \hat{b}_{1n}) = 0, \end{cases} \tag{10}$$

whereas $(\hat{a}_{2n}, \hat{b}_{2n})$ satisfies:

$$\begin{cases} \sum_{i=1}^{n_1} s_{11}(x_i, y_i, \hat{a}_{2n}) = 0, \\ \sum_{i=n_1+1}^n s_{11}(x_i, y_i, \hat{b}_{2n}) = 0. \end{cases} \tag{11}$$

The variances of \hat{a}_{1n} and \hat{a}_{2n} indicate relative design efficiency. Under the model restrictions given in [5], there exists a pair (a_0, b_0) such that:

$$\begin{cases} R_1(x) = \psi_1(x, a_0), \\ R_2(x) = \psi_2(x, a_0, b_0), \end{cases} \quad \text{for all } x. \tag{12}$$

Assuming the shift model to be correct then the dose-toxicity curve for the second group is given by $\psi_1(x, a_0)$ whereas the actual dose-toxicity curve for the group is $\psi_2(x, a_0, b_0)$. However, our conclusion will not be altered, since we are comparing the performance of \hat{a}_{1n} and \hat{a}_{2n} only and the outcome for group 2 has no effect on the outcome for group 1 in this case. Under weak conditions we have that $(\hat{a}_{1n}, \hat{b}_{1n})$ converges almost surely to (a_0, b_0) and $(\hat{a}_{2n}, \hat{b}_{2n})$ almost surely to (a_0, b'_0) . The technical details follow those in [4]. For brevity we drop the arguments (x_i, y_i, a_0) or (x_i, y_i, a_0, b_0) wherever it is possible.

Theorem 1. Assume condition 7 of [1] is satisfied. When both n_1 and n_2 become large, the asymptotic variances of \hat{a}_{1n} and \hat{b}_{1n} can then be approximated by:

$$V(\hat{a}_{1n}) = \frac{\sum_{i=n_1+1}^n s_{22}^2}{\sum_{i=n_1+1}^n s_{22}^2 (\sum_{i=1}^{n_1} s_{11}^2 + \sum_{i=n_1+1}^n s_{21}^2) - (\sum_{i=n_1+1}^n s_{21} s_{22})^2} \tag{13}$$

and

$$V(\hat{b}_{1n}) = \frac{\sum_{i=1}^{n_1} s_{11}^2 + \sum_{i=n_1+1}^n s_{21}^2}{\sum_{i=n_1+1}^n s_{22}^2 (\sum_{i=1}^{n_1} s_{11}^2 + \sum_{i=n_1+1}^n s_{21}^2) - (\sum_{i=n_1+1}^n s_{21} s_{22})^2}. \tag{14}$$

The above formulas are evaluated at (a_0, b_0) .

Corollary 1. Under the conditions of Theorem 1, the asymptotic variance of \hat{a}_{2n} can be approximated by $V(\hat{a}_{2n}) = (\sum_{i=1}^{n_1} s_{11}^2)^{-1}$, where s_{11} is calculated at a_0 .

To compare the performance of \hat{a}_{1n} and \hat{a}_{2n} we apply Cauchy–Schwarz's inequality:

$$\left(\sum_{i=n_1+1}^n s_{21}s_{22} \right)^2 \leq \sum_{i=n_1+1}^n s_{21}^2 \sum_{i=n_1+1}^n s_{22}^2.$$

Hence, it follows from (13) that $V(\hat{a}_{1n}) \leq V(\hat{a}_{2n})$. Moreover, a sufficient condition for their equality is that there is a scalar κ such that:

$$s_{21}(x, y, a_0, b_0) = \kappa \times s_{22}(x, y, a_0, b_0), \quad \text{for all } x, y. \quad (15)$$

Therefore the variance of \hat{a}_{1n} is smaller than that of \hat{a}_{2n} unless (15) is satisfied. We could constrain the difference to be strictly greater than zero, corresponding to the case where we know that, should any difference exist, it can only be in a given direction; a well-known example being heavily pre-treated and lightly pre-treated patients. The MTD for the heavily pre-treated patients will be no higher than that for the lightly pre-treated patients. This amounts to incorporating into the design known features of the experiment and the idea is easily extended. For example, we may not only know the only possible direction any difference may take but we may also be able to say something about the size of the difference. In the case of heavily versus lightly pre-treated patients, a frequently encountered situation, the definition of the two groups is not usually very sharp so that, if a difference should exist, it may not be that great. The possible differences could be, say, either zero, one level or no more than two levels. Expressed in this way it immediately opens the possibility of bringing prior information, or prior opinion, on board and that is readily done without impact on the efficiency findings discussed here.

The estimating equation providing the estimate \hat{a}_n , is derived from the likelihood for n patients, and can be expressed equivalently as a sum over the dose levels as

$$W_n(a) = \sum_{i=1}^k w_n(d_i) U_{i,n}(a) = 0 \quad (16)$$

where

$$U_{i,n}(a) = H\{n_i(n)\} \left[\frac{v_i(n)}{n_i(n)} \frac{\psi'}{\psi}(d_i, a) + \left\{ 1 - \frac{v_i(n)}{n_i(n)} \right\} \times \frac{-\psi'}{1-\psi}(d_i, a) \right], \quad i = 1, \dots, k$$

and $H(s) = I(s \neq 0)$, i.e., a function taking the value 1 when s is not equal to 0, and is zero otherwise and $0/0$ is defined as being equal to 1. We obtain \hat{a}_n as the zero of the above estimating equation and, at that value, the inverse of $-\partial U_{i,n}(a)/\partial a$ provides an estimate of the variance denoted $\text{Var}(\hat{a}_n)$. Finally, it is worth noting that bridging is not only a question of efficiency. Sometimes there can be so few patients in the second study, when compared to the first, that the only way to make progress is to, in some sense, piggyback the second study on to the first. We then use as much information as needed to create enough structure to make inferences feasible in the smaller study.

Acknowledgements

I would like to acknowledge the reviewers for picking up errors and the editor for suggestions in helping make a more sharply focused presentation.

References

- [1] J. O'Quigley, A theoretical study of the continual reassessment method, *J. Stat. Plan. Infer.* 136 (2006) 1765–1780.
- [2] J. O'Quigley, M. Conaway, Continual reassessment and related dose-finding designs, *Stat. Sci.* 25 (2010) 202–216.
- [3] J. O'Quigley, M. Pepe, L. Fisher, Continual Reassessment Method; a practical design for Phase 1 clinical studies in cancer, *Biometrics* 46 (1) (1990) 33–48.
- [4] J. O'Quigley, L.Z. Shen, Continual reassessment method: A likelihood approach, *Biometrics* 52 (1996) 163–174.
- [5] J. O'Quigley, L. Shen, A. Gamst, Two sample continual reassessment method, *J. Biopharm. Stat.* 9 (1999) 17–44.
- [6] J. Shu, Continual reassessment designs in the presence of population heterogeneity, PhD thesis, Dept. of Statistics, University of Virginia, USA, 2012.