Monografías del Semin. Matem. García de Galdeano. **27:** 19–26, (2003).

Heterogeneity analysis on multi-centre trials: application to early breast cancer

Virtudes Alba[1], M. Dolores Jiménez[2], Ana Muñoz[2]

Dpt. Estadística e Investigación Operativa, Universidad de Jaén[1] y de Sevilla[2]

e-mail: mvalba@ujaen.es, dolores@cica.es, anamun@cica.es

Abstract

Meta-analysis provides a framework for systematic and explicit processes of review and summation. It is widely used in clinical medicine since it allows to extract more powerful conclusions than that obtained from individual studies. An important issue in meta-analysis is which metric is appropriated to measure the effect of treatments or interventions. Another topic is how to analyze the heterogeneity of results. In this paper we analyze the heterogeneity of randomized trials within the breast cancer research for two set of trials that involve treatments with tamoxifen and polychemotherapy. In order to determine under which conditions the treatments produce an important reduction on mortality rates, we model the results for the risk difference as a summary-measure.

Keywords: Meta-analysis, risk difference, breast cancer

AMS Classification: 62F03, 62F10

1 Introduction

Meta-analysis is a group of techniques for the quantitative synthesis of primary study results. It is widely used in lots areas of research and specially in medicine and health care. An important point in meta-analysis is to choose a measure of outcome for each study before combining them.

Here, we will focus in methods for meta-analysis within medical research and using binary outcomes. In this case, there are several effect measures, or "metrics", directly available. These measures are obtained by comparing the probabilities of events in the two groups of patients. Point estimates of these measures are obtained by substituting the probabilities for the observed rate of events.

Hence, if k represents the number of available trials and p_{ie} and p_{ic} the proportions of subjects with the event in the treatment and control groups for the trial i, respectively. The main used measures are the following [3]:

- The *absolute risk reduction* or *risk difference* is the difference in the probabilities of an event in the control and treatment groups and it is estimated as the corresponding difference in the event rates,

$$RD_i = p_{ie} - p_{ic}, i = 1, ..., k.$$

- The *relative risk* is defined as the probability of an event in the treatment group divided by the probability of an event in the control group and it is estimated as the ratio of the corresponding event rates,

$$RR_i = \frac{p_{ie}}{p_{ic}}, i = 1, ..., k.$$

- The *odds ratio* is defined as the odds of an event in the treatment group divided by the odds of an event in the control group and it is estimated as follows:

$$OR_i = \frac{p_{ie}\,(1 - p_{ic})}{p_{ic}\,(1 - p_{ie})}, i = 1, ..., k.$$

- The "*number needed to treat*" (NNT) is defined as the inverse of the risk difference,

$$NNT_i = \frac{1}{RD_i}, i = 1, ..., k.$$

2 What measure we have to use?

Each measure has its own interpretation and statistical properties that make it suitable for some applications but perhaps not for others. Historically, the *odds ratio* was the most common measure to report for randomised clinical trials and the case control studies, while *relative risks* are often used for reporting cohort studies. However, *the number needed to treat* is a measure that is being used increasingly because for clinical decision making, it is more meaningful than the other measures. The measure NNT can be interpreted as the number of patients that need to be treated using the experimental/new treatment, rather than the placebo/old treatment in order to prevent one additional adverse outcome (Cook, [1]).

To examine empirically how assessment of treatment effect and heterogeneity may differ when different methods are utilized, Lau et al. [2] have studied 125 meta-analysis that can be considered representative of those performed by clinical investigators. Some of their conclusions were:

- There was no meta-analysis in which the risk difference and odds ratio were discrepant to extent that one indicated significant benefit while the other indicated significant harm.
- For most meta-analysis, the odds ratio and risk difference agreed in statistical significance, leading to similar conclusions about whether treatments affected outcome.
- The risk difference metric gives large weight to trials with small proportions p_{ie} and p_{ic}, whereas the odds ratio gives large weight to trials with p_{ie} and p_{ic} near to 0.5.

After these comments, we will take as a measure of outcome the risk difference for two reasons: it has a reasonably easier interpretation maintaining good analytic properties.

3 Study heterogeneity analysis

When we dispose of a set of studies, the point estimates of the effect of treatment may differ between them. This variability can be modeled under two type of assumptions. The first suppose that the differences are dued to sampling error and then the effect estimates are considered to be homogeneous. This situation can be modeled by a *fixed effect model*.

However, often the variability in effect estimates exceeds that expected from sampling error alone, that is, there are "real" and "systematic" differences between studies. This situation is modeled by a *random effect model*.

So, given a set of results after deciding which modelization is more appropriated, we can extract some general conclusions about the evidence of the effect of an intervention or treatment by estimating the treatment effect of all combined trials.

In a brief way, the fixed effect model suppose that treatment effects are the same in all the primary studies. For the risk difference, this idea is represented by the null hypothesis $H_0 : RD_1 = RD_2 = ... = RD_k$ where k is the number of available studies.

The fixed effect risk difference summary estimate is a weighted average given by

$$\overline{y}_w = \frac{\sum_{i=1}^{k} w_i RD_{iRD}}{\sum_{i=1}^{k} w_i},$$

where the weights are given by

$$w_i = \left(\frac{p_{ie}\,(1 - p_{ie})}{n_{ie}} + \frac{p_{ic}\,(1 - p_{ic})}{n_{ic}} \right)^{-1},$$

being n_{ie} and n_{ic} the number of subjects randomised in the experimental and control group to the trial i, respectively.

These inverse-variance weights minimize the variance of the summary estimate and then, the variance of the fixed effect risk difference is

$$Var(\overline{y}_w) = \frac{1}{\sum_{i=1}^{k} w_i}.$$

So, to test if the treatment effect estimates can be considered homogeneous or not, that is, if it is appropriated to suppose a fixed effect model for the risk difference, we consider the test statistic

$$Q = \sum_{i=1}^{k} w_i \left(RD_i - \overline{y}_w\right)^2. \tag{1}$$

Under H_0, Q is approximately distributed as a χ^2 distribution with $k-1$ degrees of freedom. Hence, we regret H_0 if the value of Q exceeds by the $1-\alpha$ percentile of the corresponding χ^2 distribution.

On the other hand, the random effect model supposes that the measures RD_i, $i = 1, ..., k$ are observations of a random variable with unknown mean μ and variance σ^2. The point estimate for the mean treatment effect of all trials is another weighted average given by

$$\overline{y}_w^* = \frac{\sum_{i=1}^{k} w_i^* RD_i}{\sum_{i=1}^{k} w_i^*},$$

where now the weights w_i^* are given by

$$w_i^* = \frac{1}{\left(\widehat{\sigma}^2 + \frac{1}{w_i}\right)},$$

being

$$\widehat{\sigma}^2 = \begin{cases} 0 & \text{if } Q \leq k-1 \\ \dfrac{(Q-(k-1))\sum_{i=1}^{k} w_i}{\left(\sum_{i=1}^{k} w_i\right)^2 - \sum_{i=1}^{k} w_i^2} & \text{if } Q > k-1 \end{cases}$$

and Q is the statistic defined in (1).

Note that these weights combine the between and within study variance and now, the variance of the weighted average is the following

$$Var\left(\overline{y}_w^*\right) = \frac{1}{\sum_{i=1}^{k} w_i^*}$$

Hence, to test if the random effect model is appropriated to data, we state the null hypothesis $H_0' : \sigma^2 = 0$ that we can see it is equivalent to test H_0.

4 Application

Now, we are going to analyze the heterogeneity of results obtained from multi-centre trials in order to extract conclusions more general than those we can obtain with a single study. In fact, we will apply this modelization to results of randomised trials within breast cancer research taking as a measure of the treatment effect the risk difference.

We dispose of two sets of data of randomised trials carried out since 1970 in different countries that involve treatments with tamoxifen and polychemotherapy in early breast cancer. Our aim is to analyze the possible heterogeneity of the results and determine under which conditions the treatments produce an important reduction on mortality rates.

We will analyze each set of data separately.

4.1 Treatments with tamoxifen

We have the results related to 55 trials that involve treatments with tamoxifen for early breast cancer (EBCTCG, [5]). The principal events were the proportions of mortality in the experimental and control groups that we identify by p_e and p_c, respectively. We have others additional variables like, "the adjuvant tamoxifen schedule" (D), classified in 20, 30 and 40 mg/day, "duration of treatment" measured in yerars (TD) and "type of treatment" (TT). The values of the variables TD and TT are the followings:

$$TD = \begin{cases} 1 & \text{treatment for 1 year} \\ 2 & \text{treatment about 2 years} \\ 3 & \text{treatment for more than 2 years} \end{cases}$$

$$TT = \begin{cases} 1 & \text{tamoxifen plus chemotherapy} \\ 0 & \text{chemotherapy alone} \end{cases}$$

Let us consider $RD_i, i = 1, ..., 55$ the risk differences measured on each study. Firstly, we analyze if the results of the 55 trials can be considered homogeneous or not, that is, if the absolute benefits of these treatments are equals or not for the 55 combined studies.

Evaluating the expression of the statistic Q, we obtain that $Q = 458.698$ and the corresponding p-value for a ji-square with 54 degree of freedom is 0. Therefore, we can say that the absolute risk reduction produced by the treatments with tamoxifen are obviously heterogeneous.

Now, we investigate some possible sources of the heterogeneity taking into account certain attributes and conditioning the results to these conditions.

1. If we form groups depending on the duration of the applied treatment, we obtain tree subgroups and we can interested in to study the behaviour of the results within subgroups.

Description of the subgroups	k	Q	p-value	Modelling
D_1 for $TD = 1$	14	19.8144	0.0999	Fixed effect $\overline{y}_w = -0.0311$
D_2 for $TD = 2$	32	414.518	0	Random effect $\begin{cases} \overline{y}^*_w = -0.04876 \\ \hat{\sigma}^2 = 0.01413 \end{cases}$
D_3 for $TD = 3$	9	19.4522	0.0126	Random effect $\begin{cases} \overline{y}^*_w = -0.05371 \\ \hat{\sigma}^2 = 0.00128 \end{cases}$

Table 1: Heterogeneity analysis of mortality depending on the duration of treatment.

The table 1 shows the number of trials on each subgroup, the values of the Q-Statistic, the associated p-value and the effect model more appropriated for each case.

From these values, we can conclude that for patients with a treatment for one year (subpopulation represented by D_1), the reduction of the mortality is about 3.11% whereas for treatments longer, this reduction increases until the 5.37%.

2. In a similar way, we can analyze if there are difference on the results depending on other variables. For instance, if we consider the variables "type of treatment" and "duration of treatment" as attribute and we perform a similar analysis, we obtain the results showed on table 2.

   From these values we point out two relevant results:

   - For treatments with 30 mg/day of adjuvant tamoxifen (subgroup represented by D_{30}) we obtain an important reduction on mortality rates (6.63%).
   - For treatments that combine tamoxifen and chemotherapy the reduction on the mortality is more than twice those with tamoxifen alone, 7.19% against 3%.

4.2 Treatments with polychemotherapy

Now, we are going to consider data from trials of polychemotherapy for early breast cancer (EBCTCG, [4]). These data refer to randomised trials that began before 1990 in different countries and involve treatment groups that differ only with respect to the adjuvant chemotherapy regiment.

To analyze the effect of treatments with polychemotherapy on mortality in the control and experimental groups, we dispose of the proportions of mortality, p_e and p_c, respectively, and the type of applied treatment denoted by T with the following values:

Attribute	Description of the subgroups	k	Q	p-value	Modelling
TT	P_1 for $TT=1$	23	398.373	0	Random effect $\begin{cases} \overline{y}_w^* = -0.0719 \\ \hat{\sigma}^2 = 0.01881 \end{cases}$
	P_0 for $TT=0$	32	40.5954	0.1161	Fixed effect $\overline{y}_w = -0.03004$
D	D_{20} for $D=20$	28	35.1237	0.135759	Fixed effect $\overline{y}_w = -0.0321$
	D_{30} for $D=30$	17	387.825	0	Random effect $\begin{cases} \overline{y}_w^* = -0.0663 \\ \hat{\sigma}^2 = 0.02792 \end{cases}$
	D_{40} for $D=40$	9	6.63707	0.57247	Fixed effect $\overline{y}_w = -0.0288$

Table 2: Heterogeneity analysis of mortality depending on the type of treatment and the adjuvant tamoxifen schedule.

Identification of subgroups	k	Q	p-value	Modelling
S_1 for $T=1$	19	34.6749	0.0104	Random Effect $\begin{cases} \overline{y}_w^* = -0.0433 \\ \hat{\sigma}^2 = 0.002 \end{cases}$
S_2 for $T=2$	9	6.14524	0.6309	Fixed Effect $\overline{y}_w = -0.05149$
S_3 for $T=3$	19	8.0898	0.9772	Fixed Effect $\overline{y}_w = -0.04048$

Table 3: Heterogeneity analysis of mortality depending on the type of treatment.

$$T = \begin{cases} 1 & \text{treatment with } CMF \\ 2 & \text{treatment with } CMF \text{ with extra drugs} \\ 3 & \text{Other polychemotherapy} \end{cases}$$

where C represents *cyclosphosphamide*, M *methotrexate* and F *fluorouracil*, and CMF represents the treatment with the combination of the three composites.

The table 3 contains the results of the modelization more appropriate for each case. These values shows that the patients treated with CMF plus extra drugs (subgroup represented by S_2), the reduction on the mortality rates is one point higher than the other two cases.

References

[1] Cook, R. J. and Sackett, D. L. (1995) The number needed to treat: a clinically useful measure of treatment effect. British Medical Journal, 310, 452-454.

[2] Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I. and Lau J. (2000) Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Statistics in Medicine, 19, 1707-1728.

[3] Sutton, A. J., Abrams, K. R. et al. (2000) Methods for Meta-analysis in Medical Research. John Willey and Sons.

[4] Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *The Lancet,* 1998; 352: 930-42.Howard, R.A.

[5] Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. The Lancet, 1451-1467. (1998).