

## THE MANTEL-HAENSZEL PROCEDURE IN EPIDEMIOLOGICAL STUDIES: AN INTRODUCTION

Parodi Stefano<sup>1</sup>, Bottarelli Ezio<sup>2</sup>

### KEYWORDS

Mantel-Haenszel, case control study, cohort study, confounders, survival curves.

### PAROLE CHIAVE

Mantel-Haenszel, studi caso controllo, studi di coorte, confondenti, curve di sopravvivenza.

### SUMMARY

Many epidemiological studies aim to evaluate the existence of associations between outcomes (i.e., disease incidence/prevalence, effect of treatment administration, etc.) and presumptive causes (i.e., exposures, risk factors, genetic markers, etc.). Results of these studies can be biased by the presence of external variable(s), associated both with the factor(s) and the outcome(s). These nuisance variables are called confounders. Several methods are available in order to control the effect of confounders.

In this paper, the Mantel-Haenszel (MH) statistical method for confounders control is illustrated. It represents a simple and useful tool to obtain estimates of association, adjusted for the effect of one or more confounders. It is very easy in computational form, it does not require specific software, and the interpretation of the results is friendly. Moreover, an application of the MH method to the two major epidemiological observational studies (case-control and cohort studies) is briefly illustrated by the aid of examples. Finally, the use of MH method in the comparison of two survival curves (log-rank test) is instanced.

### RIASSUNTO

Molti studi epidemiologici hanno lo scopo di verificare l'esistenza di associazione fra determinati eventi (es. incidenza o prevalenza di una malattia, esito di un trattamento, ecc.) e possibili cause (es. fattori di rischio, presenza di markers genetici). I risultati di questi studi possono essere distorti dalla presenza di variabili esterne associate sia all'effetto che alla presunta causa. Tali variabili prendono il nome di "confondenti" (confounders) ed il loro effetto viene detto "confondimento" (confounding).

Esistono numerosi metodi utilizzabili al fine di controllare le variabili di

---

1 Epidemiology and Biostatistics Section, Scientific Directorate, G. Gaslini Children's Hospital, Largo G. Gaslini, 5, 16147 Genoa (Italy); e-mail: stefanoparodi@ospedale-gaslini.ge.it

2 Università degli Studi di Parma, Dipartimento di Salute Animale. Via del Taglio 10, 43100 Parma (Italy). e-mail: ezio.bottarelli@unipr.it

confondimento. In questo lavoro viene illustrato il metodo di Mantel-Haenszel, che rappresenta uno strumento utile e semplice per ottenere stime di associazione corrette per l'effetto di uno o più confondenti. Le modalità di calcolo estremamente semplici (non è richiesto alcun software statistico) e la facile interpretazione dei risultati rendono questa metodica di impiego particolarmente amichevole. Viene anche mostrato un esempio di applicazione del test di Mantel-Haenszel nell'ambito dei due principali studi epidemiologici osservazionali (studio caso-controllo e studio di coorte). Infine, viene illustrata l'applicazione del metodo per il confronto di due curve di sopravvivenza (log-rank test).

## INTRODUCTION

Many epidemiological investigations are aimed at evaluating the association between a specific factor and one or more outcomes. Factors under studies typically include risk factors (e.g., exposures to toxic compounds) or genetic markers and, in Clinical Epidemiology, treatments administration. The most common outcomes of interest are the incidence or prevalence of a specific disease, and mortality for specific causes of death (6). Many estimates of association may be computed in relation with the study design and the type of available data (6, 10). However, some nuisance (i.e., external) variables, associated both with the factor (e.g., some exposure) and the outcome (e.g., incidence of a specific disease), may bias such association estimates. The phenomenon is known as confounding, and such external factors are accordingly named "confounders". A review of the main method to control the confounding effect in epidemiological investigations and, in particular, in case-control studies have been illustrated elsewhere (11).

In this paper, the Mantel-Haenszel (MH) method is illustrated, which represents a simple and useful tool to obtain estimates of association, adjusted for the effect of one or more confounders. MH method was introduced at the end of 1950s (8) and it has been largely applied to many different study designs. Recently, the development of Generalized Linear Models, implemented in many statistical packages, have quite reduced the scope of application of MH method. However, its very easy computational form, which does not require statistical software, and the friendly interpretation of the results allows the MH method to be still largely applied in simple epidemiological studies. Moreover, in the survival analysis framework, a variant of MH method (the log-rank test, also known as the Mantel-Cox test) is still probably the most largely applied tool for the comparison of survival curves, especially in Clinical Epidemiology.

The general principle of the MH procedure relies on the Score test, which is a statistical method based on the likelihood theory (3, 4). In particular, in epidemiological studies, to control the effect of one (or few) confounders, the Score test is applied to data stratified on the basis of the levels (or strata) of such variable (e.g., age classes) (4, 11). Within each stratum, the outcome (e.g., number of observed events) is measured in the two or (rarely) more groups of the factor of interest (e.g., the exposure). A statistical distribution is assumed for the observed outcome, on the

basis of the method of sampling and the study design. For example, for counting, as the number of events, Binomial and Hypergeometric functions are often employed. For random variables belonging to the exponential family (which includes the Binomial and the Hypergeometric functions) the Score test takes the following generic formula (4):

(1)

$$t_s = \frac{\left[ \sum_j x_j - \hat{E} \left( \sum_j x_j \right) \right]^2}{\sum_j \hat{Var}(x_j)}$$

where  $x_j$  denotes the observed events in one group of exposure, within the  $j$  levels of the confounder, and  $E$  indicates the expected value under one specific hypothesis. The test is performed estimating the number of expected events under the hypothesis of no association between the factor of interest and the outcome.  $t_s$  follows asymptotically a chi squared distribution with 1 degree of freedom.

The MH procedure consists in calculating an estimate of a common effect of the exposure across the confounder strata using a weighted mean of an appropriate measure of association. In most epidemiological studies, such a measure represents an estimate of a relative risk between two different groups (exposed and unexposed subjects, treated and untreated patients, etc.), even if an absolute effect (e.g., mean difference) may also be estimated in some context (4). Under the null hypothesis of no association between the exposure and the outcome, the MH estimator of relative risk will tend to 1, while measures of absolute effect will tend to 0. In both such contexts, when MH approaches its expected value under the null hypothesis,  $t_s$  will tend to 0. Finally, another property of the MH estimators is the consistency, i.e., even in the presence of sparse data with 0 counts in some strata, a real number for MH estimator is obtained.

In this paper, an application of the MH method to the two major epidemiological observational studies (i.e., case-control and cohort studies) is briefly illustrated. Moreover, the last paragraph illustrates the MH method in the comparison of two survival curves. A more complete illustration of the MH procedure may be found in Kuritz et al (7). Italian readers may also refer to Grassi (4).

## APPLICATION OF THE MH METHOD IN CASE-CONTROL STUDIES

In case-control studies, the main measure of association between previous exposure and risk of developing a disease is the Odds Ratio (OR), which, for rare outcomes, provide an unbiased estimate of relative risk (RR) (2, 3, 6, 10).

In a typical (simplified) case-control study without matching, i.e., with an independent selection of cases and controls, data may be arranged as in Table 1. For

each level of the confounder an estimate of the OR may be obtained by the following formula:

(2)

$$\hat{OR}_j = \frac{a_j \cdot d_j}{b_j \cdot c_j}$$

The MH method allows to obtain a common estimate of OR ( $OR_{MH}$ ) across the strata of the confounder, by the following equation (8):

(3)

$$\hat{OR}_{MH} = \frac{\sum_{j=1}^K \frac{a_j d_j}{n_j}}{\sum_{j=1}^K \frac{b_j c_j}{n_j}}$$

It is easy to verify that equation 3 allows to obtain a consistent estimate of OR, i.e., even in the presence of sparse data with few or zero counts in some cells, a real number for  $OR_{MH}$  is obtained.

Please note that equation 3 is equivalent to equation 2 when  $j=1$ , i.e., in the absence of confounders.

As an estimator of RR,  $OR_{MH}=1$  under the null hypothesis of equal risks in exposed and unexposed subjects. It will take values higher than 1 if the exposure is positively associated with the disease (e.g., if it causes the disease), while it will range between 0 and 1 if the exposure plays some protective role (2, 10). In the case of control study without matching, the MH test for the null hypothesis coincides with the score test based on the conditional assumption of a hypergeometric distribution for the counts in each cell  $a_j$ :

(4)

$$X^2_{MH} = \frac{\left[ \sum_{j=1}^K a_j - \hat{E} \left( \sum_{j=1}^K a_j \right) \right]^2}{\sum_{j=1}^K \hat{Var}(a_j)}$$

where:

$$\hat{E} \left( \sum_{j=1}^K a_j \right) = \sum_{j=1}^K \frac{m_j n_j}{n_j}$$

and

$$\sum_{j=1}^K \widehat{Var}(a_j) = \sum_{j=1}^K \frac{m_{0j} m_{1j} n_{0j} n_{1j}}{n_j^2 (n_j - 1)}$$

It is easy to verify that when  $OR_{MH}$  tends to 1 (i.e., when there is no association between the exposure and the risk of developing the disease under study) equation 4 will tend to zero. In fact the denominator of equation 4 is:

$$\begin{aligned} \chi_{MH}^2 &= \left[ \sum_{j=1}^K a_j - \widehat{E} \left( \sum_{j=1}^K a_j \right) \right]^2 = \left[ \sum_{j=1}^K a_j - \sum_{j=1}^K \frac{m_{1j} n_{1j}}{n_j} \right]^2 = \left[ \sum_{j=1}^K a_j - \sum_{j=1}^K \frac{(a_j + c_j)(a_j + b_j)}{a_j + b_j + c_j + d_j} \right]^2 = \\ &= \left[ \sum_{j=1}^K \frac{a_j \cdot (a_j + b_j + c_j + d_j) - (a_j + c_j)(a_j + b_j)}{a_j + b_j + c_j + d_j} \right]^2 = \left[ \sum_{j=1}^K \frac{a_j d_j - b_j c_j}{n_j} \right]^2 \end{aligned}$$

If there is no effect of the exposure, the expected value of  $OR_j$  will tend to 1 in each stratum and, as a consequence, expected values of  $a_j d_j$  will be equal to  $b_j c_j$ ,  $\chi_{MH}^2$  will tend to 0, and  $OR_{MH}$  will tend to 1.

Many equations have been proposed to estimate the variance of  $OR_{MH}$  (2, 3 12). A consistent and unbiased method was illustrated by Robins et al (12) (see also Silcocks (14) for a formal demonstration):

$$(5) \quad \widehat{Var} \left[ \ln(\widehat{OR}_{MH}) \right] \cong \frac{\sum_{j=1}^K \frac{a_j d_j}{n_j} \cdot \frac{a_j + d_j}{n_j}}{2 \left( \sum_{j=1}^K \frac{a_j d_j}{n_j} \right)^2} + \frac{\sum_{j=1}^K \left( \frac{b_j c_j}{n_j} \cdot \frac{a_j + d_j}{n_j} + \frac{b_j + c_j}{n_j} \cdot \frac{a_j d_j}{n_j} \right)}{2 \left( \sum_{j=1}^K \frac{a_j d_j}{n_j} \right) \left( \sum_{j=1}^K \frac{b_j c_j}{n_j} \right)} + \frac{\sum_{j=1}^K \frac{b_j c_j}{n_j} \cdot \frac{b_j + c_j}{n_j}}{2 \left( \sum_{j=1}^K \frac{b_j c_j}{n_j} \right)^2}$$

Because  $OR_{MH}$  is a RR estimator, a log normal distribution may be assumed under the null hypothesis (10). Accordingly, confidence intervals of  $OR_{MH}$  at a selected  $\alpha$  value may be obtained from the following equation:

$$(6) \quad \widehat{OR}_{MH} \exp \left( \pm z_{\alpha/2} \sqrt{\widehat{Var} \left[ \log(\widehat{OR}_{MH}) \right]} \right)$$

More details and a numerical example have been provided elsewhere (11).

**Table 1** – Hypothetical data from a case-control study, stratified according to the levels of one confounder (e.g., age classes).

Stratum (confounder levels)	Exposure	Cases	Controls	Total
1	Exposed	a <sub>1</sub>	b <sub>1</sub>	n <sub>11</sub>
	Unexposed	c <sub>1</sub>	d <sub>1</sub>	n <sub>01</sub>
	Total	m <sub>11</sub>	m <sub>01</sub>	n <sub>1</sub>
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
j	Exposed	a <sub>j</sub>	b <sub>j</sub>	n <sub>1j</sub>
	Unexposed	c <sub>j</sub>	d <sub>j</sub>	n <sub>0j</sub>
	Total	m <sub>1j</sub>	m <sub>0j</sub>	n <sub>j</sub>
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
K	Exposed	a <sub>K</sub>	b <sub>K</sub>	n <sub>1K</sub>
	Unexposed	c <sub>K</sub>	d <sub>K</sub>	n <sub>0K</sub>
	Total	m <sub>1K</sub>	m <sub>0K</sub>	n <sub>K</sub>

In many case-control studies, confounding may be controlled by selecting control subjects on the basis of the main characteristics (i.e., the distribution of main confounders) of each case. Such a method is known as “matching” (2, 11). In a matched case-control study with a ratio 1:1 between cases and controls, results may be resumed as shown in Table 2.

An estimate of OR (Maximum Likelihood Estimate) may be obtained by the following equation (2):

(7)

$$\hat{OR} = \frac{B}{C}$$

**Table 2** – Hypothetical data from a matched case-control study, with a matching ratio 1:1.

	Exposed Controls	Unexposed Controls	Total
Exposed Cases	A	B	A+B
Unexposed Cases	C	D	C+D
Total	A+C	B+D	A+B+C+D

It is easy to show that equation (7), which represents the Maximum Likelihood Estimate of OR in a matched study, is equivalent to OR<sub>MH</sub>. In fact, data in Table 2 may also be summarized in a table stratified by each case (Table 3).

Let consider that a<sub>j</sub>d<sub>j</sub>=0 for all strata in Table 3, except for the n=B strata like that corresponding to ID=2. Moreover, let consider that b<sub>j</sub>c<sub>j</sub>=0 for all strata except for the n=C strata like that corresponding to ID=3. Applying equation (3) to the data in Table 3, the following estimate of OR<sub>MH</sub> is obtained, which corresponds to the MLE estimate reported in equation (7):

$$\hat{OR}_{MH} = \frac{\sum_{j=1}^K \frac{a_j d_j}{n_j}}{\sum_{j=1}^K \frac{b_j c_j}{n_j}} = \frac{B}{C}$$

Finally, the null hypothesis:  $OR_{MH}=1$  may be tested by the Mc Nemar chi squared test ( $\chi^2_{MN}$ ) (2)

$\chi^2_{MN}$  is a Score test and then it may be considered as a MH test.

$$\chi^2_{MN} = \frac{(B - C)^2}{B + C}$$

**Table 3** – Hypothetical data from a matched case-control study, with a matching ratio 1:1, stratified by each case. ID identifies each case in the data set.

ID	Type of case	Exposure	N. of Cases	N. of Controls
1	<i>Exposed case matched with an exposed control</i>	<i>Exposed</i>	1	1
		<i>Unexposed</i>	0	0
2	<i>Exposed case matched with an unexposed control</i>	<i>Exposed</i>	1	0
		<i>Unexposed</i>	0	1
3	<i>Unexposed case matched with an exposed control</i>	<i>Exposed</i>	0	1
		<i>Unexposed</i>	1	0
4	<i>Unexposed case matched with an unexposed control</i>	<i>Exposed</i>	0	0
		<i>Unexposed</i>	1	1
...	...	<i>Exposed</i>	...	...
		<i>Unexposed</i>	...	...
j	...	<i>Exposed</i>	a <sub>j</sub>	b <sub>j</sub>
		<i>Unexposed</i>	c <sub>j</sub>	d <sub>j</sub>
...	...	<i>Exposed</i>	...	...
		<i>Unexposed</i>	...	...

**APPLICATION OF THE MH METHOD IN THE COHORT STUDY**

Cohort (or follow-up) study is considered as the most important investigation in observational Epidemiology (6, 10). In many cases, a group of healthy people (the “cohort”) is identified and split into (at least) two categories or sub-cohorts, on

the basis of the presence of one or more specific exposures. The two sub-cohorts are followed up for a time period and the occurrence of the outcome of interest (e.g., the incidence of one or more disease) is observed. A measure of the impact of the disease in each sub-cohort may be obtained estimating the corresponding rate  $\lambda$  (3, 6, 10):

$$\hat{\lambda} = \frac{n}{m}$$

where  $n$  represents the number of observed events and  $m$  is the sum of the follow up time, often expressed in years, for each subject in the sub cohort (person-years at risk). Considering the shorter life of livestock or pet animals, in veterinary epidemiologic studies the follow up time is often expressed on a narrower scale, i.e., animal-months at risk, etc. The association between the exposure under study and the outcome may be estimated by the ratio between the rates in the exposed and unexposed sub-cohorts. For rare diseases, such a measure (rate ratio) is an unbiased estimate of RR (6, 10).

In a cohort study, in the presence of one or more confounders, data may be resumed as in Table 4.

The association between the exposure and the incidence of the disease under study, adjusted for the effect of the confounder, may be obtained by the Mantel-Haenszel rate ratio ( $RR_{MH}$ ), which is estimated by the following equation (13):

(8)

$$\hat{RR}_{MH} = \frac{\sum_{j=1}^K \frac{n_{E,j} \cdot m_{NE,j}}{m_j}}{\sum_{j=1}^K \frac{n_{NE,j} \cdot m_{E,j}}{m_j}}$$

The null hypothesis:  $RR_{MH}=1$  may be tested by the following chi squared test, which is obtained by the same method (Score test) used for equation (4), under the assumption of a conditional binomial distribution for the events  $n_{E,j}$  in Table 4:

(9)

$$\chi^2_{MH} = \frac{\sum_{j=1}^K n_{E,j} - \hat{E}\left(\sum_{j=1}^K n_{E,j}\right)}{\sum_{j=1}^K \hat{V}\hat{a}r(n_{E,j})}$$

where:

$$\sum_{j=1}^K \hat{E}(n_{E,j}) = \sum_{j=1}^K \frac{m_{E,j}}{m_j} n_j$$

and:

$$\sum_{j=1}^K \hat{V}ar(n_{E,j}) = \sum_{j=1}^K \frac{m_{E,j} \cdot m_{NE,j}}{m_j^2} n_j$$

**Table 4** – Hypothetical data from a cohort study, stratified according to the levels of one confounder (e.g., age classes). PY=Person years at risk.

Stratum (confounder levels)	Exposure	Cases	PY
1	Exposed	$n_{E,1}$	$m_{E,1}$
	Unexposed	$n_{NE,1}$	$m_{NE,1}$
	Total	$n_1$	$m_1$
...	...	...	...
...	...	...	...
...	...	...	...
j	Exposed	$n_{E,j}$	$m_{E,j}$
	Unexposed	$n_{NE,j}$	$m_{NE,j}$
	Total	$n_j$	$m_j$
...	...	...	...
...	...	...	...
...	...	...	...
K	Exposed	$n_{E,K}$	$m_{E,K}$
	Unexposed	$n_{NE,K}$	$m_{NE,K}$
	Total	$n_K$	$m_K$

As in a case-control study, also for  $RR_{MH}$  a log-normal distribution may be assumed near the null hypothesis. Accordingly, the following equation, similar to equation (6), provides an estimate of the confidence interval of  $RR_{MH}$  at a selected  $\alpha$  level:

(10)

$$\hat{R}R_{MH} \exp\left(\pm Z_{\alpha/2} \sqrt{\hat{V}ar\left[\log\left(\hat{R}R_{MH}\right)\right]}\right)$$

A consistent estimate of the variance of  $\log(RR_{MH})$  may be obtained by the following formula (5):

(11)

$$\hat{V}ar\left[\log\left(\hat{R}R_{MH}\right)\right] \cong \frac{\sum_{j=1}^K \frac{m_{E,j} \cdot m_{NE,j}}{m_j^2} n_j}{\left(\sum_{j=1}^K \frac{m_{NE,j} \cdot n_{E,j}}{m_j}\right) \left(\sum_{j=1}^K \frac{m_{E,j} \cdot n_{NE,j}}{m_j}\right)}$$

Table 5 shows an example of a hypothetical cohort study with a putative confounder at two levels (e.g., gender).

**Table 5.** Example of a cohort study with a two-level putative confounder.

	a) Whole cohort		b) Stratum 1 (e.g., Males)		c) Stratum 2 (e.g., Females)			
	n	PY	n	PY	n	PY		
Exposed	108	448,701	Exposed	30	32,178	Exposed	78	416,523
Unexposed	51	210,626	Unexposed	44	117,021	Unexposed	7	93,605

The RR estimate for the whole cohort is:

$$\hat{RR}_w = \frac{108}{448701} / \frac{51}{210626} = 0.99$$

The RR estimates for the two strata are, respectively:

$$\hat{RR}_1 = \frac{30}{32178} / \frac{44}{117021} = 2.48$$

and

$$\hat{RR}_2 = \frac{78}{416523} / \frac{7}{93605} = 2.50$$

The RR estimates for the two strata are very similar, but they clearly differ from the RR estimate for the whole cohort, pointing out the presence of a strong confounding effect.

A common estimate of RR may be obtained by applying equation (8):

$$\hat{RR}_{MH} = \frac{\frac{30 \cdot 117021}{149199} + \frac{78 \cdot 93605}{510128}}{\frac{44 \cdot 32178}{149199} + \frac{7 \cdot 416523}{510128}} = 2.49$$

Its statistical significance, under the null hypothesis of no association between exposure and risk of developing the disease under study (i.e.,  $H_0: RR_{MH}=1$ ), may be obtained by applying equation (9):

$$X_{MH}^2 = \frac{\left( 30 + 78 - \frac{32178}{149199} \cdot 74 - \frac{416523}{510128} \cdot 85 \right)^2}{\frac{117021 \cdot 32178}{149199^2} \cdot 74 + \frac{93605 \cdot 416523}{510128^2} \cdot 85} = 20.29$$

Such a value exceeds the conventional critical value for  $\alpha=0.05$  ( $\chi^2_c=3.84$ ), then the null hypothesis is rejected and the association may be considered as statisti-

cal significant.

Finally, the confidence interval of  $RR_{MH}$  may be obtained according to equation (10) and equation (11):

$$Var[\log(\hat{RR}_{MH})] \cong \frac{\frac{117021 \cdot 32178}{149199^2} \cdot 74 + \frac{93605 \cdot 416523}{510128^2} \cdot 85}{\left(\frac{117021}{149199} \cdot 30 + \frac{93605}{510128} \cdot 78\right) \left(\frac{32178}{149199} \cdot 44 + \frac{416523}{510128} \cdot 7\right)} = 0.0439$$

95%CI ( $RR_{MH}$ ):

$$2.49 \cdot e^{(\pm 1.96 \cdot \sqrt{0.0439})} = [1.65; 3.75]$$

### APPLICATION OF THE MH METHOD IN SURVIVAL ANALYSIS

The MH method has been largely applied for the comparison of survival curves, where it is commonly known as the log-rank test or the Mantel-Cox test.

Results from a simple survival analysis may be resumed as in Table 6, where the observed events (i.e., the number of deaths) are stratified by the follow up time intervals and by the presence of a hypothetical treatment.

**Table 6** – Hypothetical data from a survival analysis comparing two groups of subjects (e.g., treated and untreated). Events are stratified according to the observed intervals of the follow up time.

Follow-up time (t)	Groups under studies	At risk (live at t)	Dead at t	Total subjects at t
1	Treated	$m_{1,1}$	$n_{1,1}$	$Tot_{1,1}$
	Untreated	$m_{2,1}$	$n_{2,1}$	$Tot_{2,1}$
	Total	$m_1$	$n_1$	$Tot_1$
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
j	Treated	$m_{1,j}$	$n_{1,j}$	$Tot_{1,j}$
	Untreated	$m_{2,j}$	$n_{2,j}$	$Tot_{2,j}$
	Total	$m_j$	$n_j$	$Tot_j$
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
K	Treated	$m_{1,K}$	$n_{1,K}$	$Tot_{1,K}$
	Untreated	$m_{2,K}$	$n_{2,K}$	$Tot_{2,K}$
	Total	$m_K$	$n_K$	$Tot_K$

Please note the analogy with Table 4, where data were stratified accordingly to the levels of a confounder and a hypothetical exposure. Under the hypothesis of a conditional hypergeometric distribution for the counts  $n_{j,E}$ , a MH test may be obtained as follows (Mantel, 1966):

(12)

$$\chi^2_{MH} = \frac{\left[ \sum_{j=1}^K n_{1,j} - \hat{E} \left( \sum_{j=1}^K n_{1,j} \right) \right]^2}{\sum_{j=1}^K \hat{V}ar(n_{1,j})}$$

where:

$$\sum_{j=1}^K \hat{E}(n_{1,j}) = \sum_{j=1}^K \frac{Tot_{1,j}}{Tot_j} n_j$$

and:

$$\sum_{j=1}^K \hat{V}ar(n_{1,j}) = \sum_{j=1}^K \frac{Tot_{1,j} Tot_{2,j} n_j m_j}{Tot_j^2 (Tot_j - 1)}$$

which, in the case of only 1 observed event for each not censored time interval (e.g., in the Kaplan-Meier survival tables), is equal to the variance of a binomial variable (Grassi, 1994):

$$\sum_{j=1}^K \hat{V}ar(n_{1,j}) = \sum_{j=1}^K n_j \frac{Tot_{1,j} Tot_{2,j}}{Tot_j^2}$$

In this context, the MH test is equivalent to the Cox score test to compare two survival distributions under the assumption of proportional risks (Cox, 1972).

In epidemiological literature, some alternative equations exist to obtain the comparison between survival curves in a univariable analysis. For example, a log-rank test may be obtained by a “Pearson’s chi squared-like” procedure, i.e., estimating the variance of each observed event by its expected value (see, for example, Bland and Altman, 2004) (1). Probably this is the reason why some authors consider the log-rank test and the MH test as two different procedures.

An estimate of relative risk between treated and untreated subjects may be obtained via the Hazard Ratio, which may be still estimated by the MH procedure (Grassi, 1994):

(13)

$$\hat{HR}_{MH} = \frac{\sum_{j=1}^K n_{1,j} \cdot m_{2,j} / Tot_j}{\sum_{j=1}^K n_{2,j} \cdot m_{1,j} / Tot_j}$$

When only 1 event is observed for each follow up time  $j$ , like in the survival analysis by the Kaplan Meier method (1), equation (13) may be replaced by the following simple formula (4):

(14)

$$\hat{HR}_{MH} = \frac{\sum_{(1)} E_{2,j}}{\sum_{(2)} E_{1,j}}$$

where  $\sum_{(1)}$  indicates the sum of expected deaths in the group 2 corresponding to the observed deaths in the group 1, while  $\sum_{(2)}$  indicates the sum of expected deaths in the group 1 corresponding to the observed deaths in the group 2.

An approximated estimate of the variance of  $\log(HR_{MH})$  may be obtained from the following formula, which is valid for Kaplan-Meier survival tables (4):

(15)

$$Var \left[ \log \left( \hat{HR}_{MH} \right) \right] \cong \frac{\sum E_{1,j} \cdot E_{2,j}}{\left( \sum_{(1)} E_{2,j} \right) \cdot \left( \sum_{(2)} E_{1,j} \right)}$$

Confidence interval for  $HR_{MH}$  may be estimated applying the asymptotic normal assumption for the Hazard Ratio under the null hypothesis of similar risks between the two groups under study, similarly to the procedure previously illustrated for case-control (equation 6) and cohort (equation 10) studies:

(16)

$$\hat{HR}_{MH} \exp \left( \pm z_{\alpha/2} \sqrt{Var \left[ \log \left( \hat{HR}_{MH} \right) \right]} \right)$$

When applied to an open cohort, i.e. in the presence of censored data, the MH procedure provides an unbiased estimate of the Hazard Ratio only under the assumption of a “not informative” censoring process, which means that the risk for the subjects lost to follow-up (corresponding to the censored times) is assumed to be similar to that of the remaining subjects in the same sub-cohort. Accordingly, the same assumption is also applied to the log-rank test.

Table 7a reports simulated data regarding the survival experience of two groups of 15 patients treated with a standard therapy (group 1) and with a new drug (group 2). Excluding censored follow up time, data may be arranged as in Table 7b, where expected values  $E_{2,j}$  and  $E_{1,j}$  are computed.

Applying equation (14),  $HR_{MH}$  is easily estimated from data in Table 7b:

$$\hat{HR}_{MH} = \frac{\sum_{(1)} E_{2,j}}{\sum_{(2)} E_{1,j}} = \frac{0.5 + 0.517 + 0.520 + 0.522 + 0.6 + 0.611 + 0.643 + 0.700}{0.444 + 0.458 + 0.313} = 3.8$$

(15): An estimate of the variance of  $\log(\text{HR}_{\text{MH}})$  is easily obtained from equation

$$\text{Var}[\log(\hat{H}R_{\text{MH}})] \cong \frac{0.5 \cdot 0.5 + 0.483 \cdot 0.517 + 0.444 \cdot 0.556 + \dots + 0.357 \cdot 0.643 + 0.3 \cdot 0.7}{(0.5 + 0.517 + \dots + 643 + 0.700) \cdot (0.444 + 0.458 + 0.313)} = 0.4687$$

and the 95% confidence interval for  $\text{HR}_{\text{MH}}$  may be obtained from equation 15:

$$\left[ 3.8 \cdot e^{(\pm 1.96 \sqrt{0.4687})} \right] \rightarrow [1.0; 14.5]$$

The log-rank test statistic, according to equation (12), is:

$$\chi^2_{\text{MH}} = \frac{[8 - 4.602]^2}{\frac{15 \cdot 15}{30^2} + \frac{14 \cdot 15}{29^2} + \frac{12 \cdot 15}{27^2} + \frac{12 \cdot 13}{25^2} + \frac{11 \cdot 13}{24^2} + \frac{11 \cdot 12}{23^2} + \frac{8 \cdot 12}{20^2} + \frac{7 \cdot 11}{18^2} + \frac{5 \cdot 11}{16^2} + \frac{5 \cdot 9}{14^2} + \frac{3 \cdot 7}{10^2} + \frac{0 \cdot 5}{5^2}} = 4.398$$

which exceeds the critical value of 3.84 for the  $\chi^2$  test with 1 degree of freedom at the conventional 0.05  $\alpha$  level, thus indicating that there is a moderate evidence of a difference between the survival probability in the two groups under study ( $p < 0.05$ ) and that the new treatment is more effective than the previous one.

**Table 7a.** Comparison of survival between two groups of patients (simulated data). Group 1 was treated with a standard pharmacologic treatment, while group 2 was treated with a new drug.

Follow up time (years)	Group 1			Group 2		
	Lost	Death	At Risk	Lost	Death	At Risk
0.118	0	1	15	0	0	15
0.203	0	1	14	0	0	15
0.208	1	0	13	0	0	15
0.879	0	0	12	0	1	15
0.912	0	0	12	1	0	14
1.257	0	1	12	0	0	13
1.509	0	0	11	0	1	13
1.925	0	1	11	0	0	12
2.037	1	0	10	0	0	12
2.541	1	0	9	0	0	12
2.854	0	1	8	0	0	12
2.914	0	0	7	1	0	12
3.321	0	1	7	0	0	11
4.079	1	0	6	0	0	11
4.512	0	0	5	0	1	11
4.772	0	0	5	1	0	10
5.514	0	1	5	0	0	9
5.555	0	0	4	1	0	9
6.001	0	0	4	1	0	8
6.231	1	0	4	0	0	7

3 The p value obtained by a statistical software (Stata for Windows release 9.2) is 0.036.

6.348	0	1	3	0	0	7
6.398	0	0	2	1	0	7
6.404	0	0	2	1	0	6
6.787	1	0	2	0	0	5
6.798	1	0	1	0	0	5
6.821	0	0	0	0	1	5
6.940	0	0	0	1	0	4
7.251	0	0	0	1	0	3
7.310	0	0	0	1	0	2
7.521	0	0	0	1	0	1

**Table 7b.** Data of Table 7a, after exclusion of censored times. Expected values  $E_{1,j}$  and  $E_{2,j}$  are computed to obtain MH estimate and log-rank test.

Follow up time (years)	Death	Group 1		Death	Group 2		Total
		At Risk	$E_{1,j}$		At Risk	$E_{2,j}$	
0.118	1	15	0.500	0	15	<b>0.500</b>	30
0.203	1	14	0.483	0	15	<b>0.517</b>	29
0.879	0	12	<b>0.444</b>	1	15	0.556	27
1.257	1	12	0.480	0	13	<b>0.520</b>	25
1.509	0	11	<b>0.458</b>	1	13	0.542	24
1.925	1	11	0.478	0	12	<b>0.522</b>	23
2.854	1	8	0.400	0	12	<b>0.600</b>	20
3.321	1	7	0.389	0	11	<b>0.611</b>	18
4.512	0	5	<b>0.313</b>	1	11	0.688	16
5.514	1	5	0.357	0	9	<b>0.643</b>	14
6.348	1	3	0.300	0	7	<b>0.700</b>	10
6.821	0	0	<b>0.000</b>	1	5	1.000	5
Total	8	-	4.602	4	-	7.399	-

## ACKNOWLEDGEMENTS

This work was partly supported by a grant from the Italian Neuroblastoma Foundation (Fondazione Italiana Neuroblastoma).

## REFERENCES

- 1) Bland J.M., Altman DG (2004) The logrank test. *BMJ.* 1;328:1073.
- 2) Breslow N.E., Day N.E. (1980) *Statistical Methods in Cancer Research – Volume 1 – The analysis of case-control studies.* IARC Scientific Publications N. 32, Lyon.
- 3) Clayton D., Hills M. (1993). *Statistical models in Epidemiology.* Oxford University Press, Oxford (UK).
- 4) Grassi M. Combinazione di Tabelle 2x2. In Grassi M. (1994) *Statistica in Medicina – Un approccio basato sull verosimiglianza.* McGraw-Hill Libri Italia srl, Milano, Italy; p.415-455.

- 5) Greenland S., Robins J.M. (1985) Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, 41:55-68.
- 6) Kleinbaum D.G., Kupper L.L., Morgenstern H. (1982) *Epidemiologic research: principles and quantitative methods*. John Wiley & Sons, Inc., New York.
- 7) Kuritz S.J., Landis J.R., Koch G.G. (1988) A general overview of Mantel-Haenszel methods: applications and recent developments. *Ann Rev Public Health*, 9:123-60.
- 8) Mantel N., Haenszel W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl. Cancer Inst.*, 22: 719-748.
- 9) McCullagh P., Nelder J.A. (1989) *Generalized Linear Models*. Chapman and Hall, 2nd edition, New York.
- 10) Parodi S., Bottarelli E. (2004) Introduzione allo studio caso-controllo in Epidemiologia. *Ann. Fac. Med. Vet. Parma*, 24:209-236.
- 11) Parodi S., Bottarelli E. (2005) Controlling for confounding in case-control studies. *Ann. Fac. Med. Vet. Parma*, 25:19-46.
- 12) Robins J., Breslow N., Greenland S. (1986) Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*, 42:311-323.
- 13) Rothman K.J., Boice J.D. (1982) *Epidemiologic analysis with a programmable calculator*. Brookline, MA: Epidemiology Resources.
- 14) Silcocks P. (2005) An easy approach to the Robins-Breslow-Greenland variance estimator. *Epidemiologic Perspectives & Innovations*, 2:9.
- 15) Woolf B. (1955) On estimating the relationship between blood group and disease. *Ann. Human Genet.*, 19: 251-253.