

Everything You Wanted to Know but Could Never Find from the Cochran-Mantel-Haenszel Test

José Moral de la Rubia¹, Adrián Valle de la O²

¹School of Psychology, Universidad Autónoma de Nuevo León, Nuevo León, Monterrey, Mexico

²School of Medicine and Health Science, Tecnológico de Monterrey, Monterrey, Mexico

Email: jose.morald@uanl.edu.mx, jose_moral@hotmail.com, adrian.valle@tec.mx, adrianvalle@usa.net

How to cite this paper: Moral, J. and Valle, A. (2023) Everything You Wanted to Know but Could Never Find from the Cochran-Mantel-Haenszel Test. *Journal of Data Analysis and Information Processing*, 11, 310-339.

<https://doi.org/10.4236/jdaip.2023.113016>

Received: May 15, 2023

Accepted: August 26, 2023

Published: August 29, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Cochran-Mantel-Haenszel (CMH) test, developed in the 1950s, is a classic in health research, especially in epidemiology and other fields in which dichotomous and polytomous variables are frequent. This nonparametric test makes it possible to measure and check the effect of an antecedent variable X on a health outcome Y, statistically controlling the effect of a third variable Z that acts as a confounding variable in the relationship between X and Y. Both X and Y are measured on a dichotomous qualitative scale and Z on a polytomous-qualitative or ordinal scale. It is assumed that the effect of X on Y is homogeneous between the k strata of Z, which is usually tested by the Breslow-Day test with the Tarone's correction or the Woolf's test. The main statistical programs have the CMH test together with a test to verify the assumption of a homogeneous effect across the strata, so that it is easy to apply. However, its fundamentals and details of calculations are a mystery to most researchers, and even difficult to find or understand. The aim of this article is to present these details in a clear and concise way, including the assumptions and alternatives to non-compliance. This technical knowledge is applied to a simulated realistic example of the area of epidemiology in health and, finally, an interpretive synthesis of the analyses is given. In addition, some suggestions for the test report are made.

Keywords

Odds Ratio, Effect Size, Statistical Control, Qualitative Variables, Nonparametric Statistics

1. Introduction

The Cochran-Mantel-Haenszel (CMH) test [1] [2] is a statistical test used to

examine the association between two categorical variables while controlling for the effects of a third variable. It is particularly useful when analyzing data with a 2×2 contingency table and a potential confounding variable. By calculating a test statistic that follows a chi-squared distribution, the CMH test determines whether the association between the variables of interest is statistically significant after accounting for the stratifying variable. This test, named after its developers, Cochran, Mantel, and Haenszel, is commonly employed in epidemiology and biostatistics to understand relationships in observational studies and clinical trials while adjusting for potential confounders.

This manuscript, focused on statistical aspects of data analysis, aims to present the technical details of the Cochran [1] and Mantel and Haenszel [2] test in order to achieve an adequate understanding of it. First, the hypotheses and assumptions of the test are stated. Next, the point estimation and confidence interval of the common odds ratio and the check of its significance with a Z test are shown; afterward, the computation of the Cochran's Q statistic and its sampling distribution is examined, as well as the statistical decision making. The issue of effect size for this nonparametric test is also addressed. The verification of the assumption of a homogeneous effect of X on Y in each stratum is deepened through two tests, namely: the Woolf's test [3] and the Breslow-Day test [4] with Tarone's correction [5]. The theoretical part ends by indicating how to proceed in case of unfulfillment of the aforementioned homogeneity assumption and what other, more generalized, alternative exists to the CMH test. Finally, an example of application of this test is presented. The effect of tobacco smoke, smoked at home, on bronchial asthma in pediatric minors is analyzed, statistically controlling the effect of air quality in the place of residence. For this purpose, a simulated sample composed of 2000 participants, which is composed of k sub-samples randomly collected from the general population of k countries, was generated. This example is a realistic simulation aimed at showing technical and interpretive aspects of the CMH test.

Due to the current emphasis on practical applications of statistical software in textbooks and teaching in empirical sciences, the mathematical foundations of statistical tests are often overlooked. However, curiosity often arises, especially among educators and researchers, regarding how these seemingly magical results generated by software are obtained, and there is an underlying question of whether these quick and easy results are indeed valid. The motivation of this article is to address this need by providing a comprehensible and practical approach to the CMH test.

With the CMH test, as well as the verification of its assumptions and alternatives, all calculations can be performed using statistical software. The article does not cover the need to know how to conduct the calculations semi-automatically, but focuses on unraveling the mystery of the mathematical foundation of the calculations, a mystery because the information is lost when it is systematically overlooked.

2. Statistical Hypothesis of the CMH Test

This non-parametric test, initially devised by Cochran [1] and deepened by Mantel and Haenszel [2], was developed to test the relationship between two dichotomous qualitative variables, X and Y, while statistically controlling the effect of a third variable Z with k qualitative or ordinal categories; these categories receive the name of strata. Variable X is a factor that affects health; it is evaluated using two categories: 0 = absent or not exposed to X, and 1 = present or exposed to X. On the other hand, variable Y is the clinical outcome or diagnostic status (case), and it is also evaluated using two categories: 0 = non-case and 1 = case [1] [2]. The information of these three variables is arranged in $k \times 2 \times 2$ tables of a $k \times 4$ table. The null hypothesis holds that variable X has no effect on variable Y, which implies that the odds ratio (OR) of exposure between cases and non-cases is unitary. The alternative hypothesis states that X has an effect on Y, so that the odds ratio is not unitary. If OR is lower than 1, then X acts as a protective factor for being a case; on the other hand, if OR is higher than 1, then X acts as a risk factor for being a case [6].

The odds of exposure for an event A are the ratio between the probability of being exposed and the probability of not being exposed: $Odds(A) = P(A)/[1 - P(A)]$. These odds are calculated both between cases and non-cases. The odds ratio (OR) is the quotient between the odds of exposure in cases (numerator) and the odds of exposure in non-cases (denominator):

$OR = Odds(A|case)/Odds(A|non-case)$. An OR higher than 1 informs how many times the presence of the risk factor is more likely than its absence among cases compared to non-cases. On the other hand, an OR lower than 1 informs how many times is more likely the presence of the protective factor than its absence among non-cases compared to cases. A unitary OR shows that the exposure factor has no effect on being a case or not.

Null hypothesis (H_0): $OR = 1$

Alternative hypothesis (H_1): $OR \neq 1$

3. CMH Test Assumptions

The test requires two dichotomous qualitative variables X and Y. The variable X is supposed to be a risk factor or a protection factor for health, and is measured or evaluated through two categories: exposed and unexposed. Variable Y is a health outcome (be it a disorder, syndrome, or disease), also evaluated through two categories (case vs non-case) assumed to be affected by or dependent on X. There is also a polytomous variable Z, with k qualitative or ordinal categories, that acts as a confounding variable in the relationship between X and Y. Thus, k independent random samples of paired data of X and Y, drawn from the strata or population groups defined by the variable Z, are required. Furthermore, assuming homogeneity across the OR s of the k strata, the effect of X on Y is estimated by the common odds ratio across the k strata.

In order to apply the CMH test, the raw OR of each table, the common OR

(the adjusted *OR* for the stratification variable *Z*), and the Cochran's Q test statistic, which is a generalization of the McNemar's test statistic, are calculated for $k \times 2$ tables [2]. Symbology is expressed in **Table 1**. Finally, it is necessary to verify the homogeneity of effect between the k strata or equivalence of the *OR* of the $k \times 2$ tables. The Woolf's test and the Breslow-Day test with the Tarone's correction are the most commonly used statistics to verify this assumption [6] [7].

4. Odds Ratio across Strata

From **Table 1**, the odds ratio between *X* and *Y* in stratum *i* of the confounding variable *Z*, OR_i , is calculated using the following probability quotient:

$$\widehat{OR}_i = \frac{Odds(X=1|Y=1)}{Odds(X=1|Y=0)} = \frac{P(X=1|Y=1)}{1-P(X=1|Y=1)} \cdot \frac{c_i}{m_{1i}} \bigg/ \frac{P(X=1|Y=0)}{1-P(X=1|Y=0)} \cdot \frac{d_i}{m_{0i}} = \frac{a_i/c_i}{b_i/d_i} = \frac{a_i d_i}{b_i c_i} \quad (1)$$

If the random sample drawn from a population stratum is large, the significance of each OR_i can be tested and a confidence interval can be estimated from its distributional convergence to a standard normal distribution. Nevertheless, in order to facilitate this convergence, a natural logarithmic transformation is applied [8].

Statistical hypotheses:

$$H_0 : OR_i = 1$$

$$H_1 : OR_i \neq 1$$

Test statistic and its asymptotic sampling distribution (**Table 1**):

$$\ln(\widehat{OR}_i) = \ln\left(\frac{a_i d_i}{b_i c_i}\right) \quad (2)$$

$$\hat{\sigma}_{\ln(OR_i)} = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}} \quad (3)$$

$$Z = \frac{\ln(\widehat{OR}_i)}{\hat{\sigma}_{\ln(OR_i)}} = z \sim N(0,1) \quad (4)$$

Decision based on the critical value ($z_{1-\alpha/2}$) or the critical level (*p*-value) in a two-tailed test:

$$\text{If } z \leq z_{1-\frac{\alpha}{2}} \text{ or } p\text{-value} = 2 \times [1 - P(Z \leq |z|)] \geq \alpha, H_0 \text{ is accepted}$$

$$\text{If } z > z_{1-\frac{\alpha}{2}} \text{ or } p\text{-value} = 2 \times [1 - P(Z \leq |z|)] < \alpha, H_0 \text{ is rejected}$$

$z_{1-\alpha/2}$ = critical value or quantile of order $1 - \alpha/2$ of a standard normal distribution $N(0, 1)$. If $\alpha = 0.05$ (conventional significance level), $z_{0.975} = 1.96$.

$P(Z \leq z) = F_Z(z)$ = value of the cumulative distribution function or cumulative probability of a standard normal distribution.

Table 1. Joint and marginal frequencies of X and Y in group or stratum i of Z.

X Independent variable	Y = Dependent or predicted variable		Σ_{rows}
	1 = case	0 = non-case	
1 = exposed	$n_{11i} = a_i$	$n_{10i} = b_i$	$n_{1\cdot i} = n_{1i}$
0 = no exposed	$n_{01i} = c_i$	$n_{00i} = d_i$	$n_{0\cdot i} = n_{0i}$
Σ_{columns}	$n_{\cdot 1i} = m_{1i}$	$n_{\cdot 0i} = m_{0i}$	$n_{\cdot \cdot i} = T_i$

Note. X = {0 = not exposed, 1 = exposed} = factor that supposedly generates an effect on health, Y = {0 = no case, 1 = case} = health status or outcome, Z = {1, 2, ..., k} = confounding variable in the relationship between X and Y, which defines the k strata. Joint absolute frequencies in stratum i : $n_{11i} = a_i$ = exposed cases, $n_{10i} = b_i$ = exposed non-cases, $n_{01i} = c_i$ = non-exposed cases, and $n_{00i} = d_i$ = non-exposed non-cases. Absolute marginal frequencies in stratum i : $n_{1\cdot i} = n_{1i}$ = exposed, $n_{0\cdot i} = n_{0i}$ = non-exposed, $n_{\cdot 1i} = m_{1i}$ = cases, and $n_{\cdot 0i} = m_{0i}$ = non-cases. Total stratum of i : $n_{\cdot \cdot i} = T_i$.

Confidence interval at $(1 - \alpha) \times 100$ of OR_i :

$$P\left(\ln(\widehat{OR}_i) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)} \leq \ln(OR_i) \leq \ln(\widehat{OR}_i) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}\right) = 1 - \alpha \quad (5)$$

$$P\left(e^{\ln(\widehat{OR}_i) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}} \leq OR_i \leq e^{\ln(\widehat{OR}_i) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}}\right) = 1 - \alpha \quad (6)$$

$$1 \in \left[e^{\ln(\widehat{OR}_i) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}}, e^{\ln(\widehat{OR}_i) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}} \right] \Rightarrow H_0 : OR_i = 1$$

$$1 \notin \left[e^{\ln(\widehat{OR}_i) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}}, e^{\ln(\widehat{OR}_i) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_i)}} \right] \Rightarrow H_1 : OR_i \neq 1$$

5. Odds Ratio across Strata

The Mantel-Haenszel common odds ratio [2] can be calculated through the following ratio of sums of proportions (Table 1):

$$\widehat{OR}_{MH} = \frac{\sum_{i=1}^k \frac{a_i d_i}{T_i}}{\sum_{i=1}^k \frac{b_i c_i}{T_i}} \quad (7)$$

The significance of the Mantel-Haenszel common odds ratio can be tested using a Z test, as the SPSS program does [9].

Statistical hypotheses. The null hypothesis holds that population OR_{MH} is equal to 1, that is, variable X has no effect on variable Y across the homogenous k strata. The alternative hypothesis states that population OR_{MH} is not equal to 1 and, consequently, variable X has effect on variable Y.

$$H_0 : OR_{MH} = 1$$

$$H_1 : OR_{MH} \neq 1$$

Assumptions: 1) X is a dichotomous qualitative variable that plays the role of predictor or independent variable; 2) Y, also a dichotomous qualitative variable, serves the role of dependent variable; 3) there are k independent strata defined by the qualitative or ordinal categories of a confounding variable Z; 4) there exists a random sample of paired data from X and Y in each stratum; and 5) the odds ratios are homogeneous across the k strata.

The Z-test statistic and its asymptotic sampling distribution: A natural logarithmic transformation is applied to the Mantel-Haenszel common odds ratio.

$$\ln(\widehat{OR}_{MH}) \tag{8}$$

The asymptotic standard error of the log-transformed Mantel-Haenszel common odds ratio is estimated through the formula given by Robins, Breslow, and Greenland [10].

$$\hat{\sigma}_{\ln(OR_{MH})} = \sqrt{\frac{1}{2} \left(\frac{\sum_{i=1}^k s_i u_i}{u^2} + \frac{\sum_{i=1}^k (s_i v_i + t_i u_i)}{uv} + \frac{\sum_{i=1}^k t_i v_i}{v^2} \right)} \tag{9}$$

$$s_i = \frac{a_i + d_i}{T_i} \quad t_i = \frac{b_i + c_i}{T_i} \quad u_i = \frac{a_i \times d_i}{T_i} \quad v_i = \frac{b_i \times c_i}{T_i}$$

$$\sum_{i=1}^k s_i u_i = \sum_{i=1}^k \left(\frac{a_i + d_i}{T_i} \times \frac{a_i \times d_i}{T_i} \right)$$

$$\sum_{i=1}^k t_i v_i = \sum_{i=1}^k \left(\frac{b_i + c_i}{T_i} \times \frac{b_i \times c_i}{T_i} \right)$$

$$\sum_{i=1}^k (s_i v_i + t_i u_i) = \sum_{i=1}^k \left(\frac{a_i + d_i}{T_i} \times \frac{b_i \times c_i}{T_i} + \frac{b_i + c_i}{T_i} \times \frac{a_i \times d_i}{T_i} \right)$$

$$u = \sum_{i=1}^k u_i = \sum_{i=1}^k \frac{a_i \times d_i}{T_i}$$

$$v = \sum_{i=1}^k v_i = \sum_{i=1}^k \frac{b_i \times c_i}{T_i}$$

$$\widehat{OR}_{MH} = u/v$$

The quotient between the log-transformed Mantel-Haenszel common odds ratio (numerator) and its standard deviation or error (denominator) constitutes the Z test statistic that converges, in distribution, to a standard normal distribution.

$$Z = \frac{\ln(\widehat{OR}_{MH})}{\hat{\sigma}_{\ln(OR_{MH})}} = z \sim N(0,1) \tag{10}$$

Decision and confidence interval. The decision on H_0 can be taken based on the critical value ($z_{1-\alpha/2}$) or the critical level (p-value) in a two-tailed test.

If $z \leq z_{1-\frac{\alpha}{2}}$ or $p\text{-value} = 2 \times [1 - P(Z \leq |z|)] \geq \alpha$, H_0 is accepted

If $z > z_{1-\frac{\alpha}{2}}$ or $p\text{-value} = 2 \times [1 - P(Z \leq |z|)] < \alpha$, H_0 is rejected

For calculating the confidence interval at $(1 - \alpha) \times 100$ for the Mantel-Haenszel common odds ratio, first, the lower and upper limits of the log-transformed OR_{MH} are obtained, using the Wald's method [11]; next, the transformation is undone in order to arrive to the Mantel-Haenszel common odds ratio limits.

$$P\left(\ln(\widehat{OR}_{MH}) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})} \leq \ln(OR_{MH}) \leq \ln(\widehat{OR}_{MH}) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}\right) = 1 - \alpha \quad (11)$$

$$P\left(e^{\ln(\widehat{OR}_{MH}) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}} \leq OR_{MH} \leq e^{\ln(\widehat{OR}_{MH}) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}}\right) = 1 - \alpha \quad (12)$$

$$1 \in \left[e^{\ln(\widehat{OR}_{MH}) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}}, e^{\ln(\widehat{OR}_{MH}) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}} \right] \Rightarrow OR_{HM} = 1$$

$$1 \notin \left[e^{\ln(\widehat{OR}_{MH}) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}}, e^{\ln(\widehat{OR}_{MH}) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}} \right] \Rightarrow OR_{HM} \neq 1$$

6. CMH Q Statistic

The Q statistic is a quotient. Its numerator is the square of the difference between the sum of exposed cases in each stratum and the sum of the expected values under the assumption of independence for these frequencies. Its denominator is the sum of the variances of the frequencies in each stratum under the assumption of independence. The higher the numerator is with respect to the denominator, the higher the relationship between the two dichotomous variables [2] [6].

$$Q_{CMH} = \frac{\left(\sum_{i=1}^k a_i - \sum_{i=1}^k e_i\right)^2}{\sum_{i=1}^k s_i^2} = \frac{\left(\sum_{i=1}^k a_i - \sum_{i=1}^k \frac{n_i m_i}{T_i}\right)^2}{\sum_{i=1}^k \frac{n_{0i} n_{1i} m_{0i} m_{1i}}{T_i^3 - T_i^2}} \quad (13)$$

The Q test statistic converges to a chi-square distribution with 1 degree of freedom; guarantee of an adequate convergence includes: frequencies expected under the null hypothesis of independence ($E_{ij} = n_i \times m_j / T_i$) higher than or equal to 5, absence of any null observed joint frequency, and totals (T_i) higher than 25 in each stratum. In order to improve the convergence to a chi-square distribution with one degree of freedom, it is recommended to apply the Yates' correction; the reason for implementing this controversial correction is the need to smooth the transition from a discrete binomial distribution to a continuous chi-square distribution [12].

$$Q'_{CMH} = \frac{\left(\left|\sum_{i=1}^k a_i - \sum_{i=1}^k \frac{n_i m_i}{T_i}\right| - 0.5\right)^2}{\sum_{i=1}^k \frac{n_{0i} n_{1i} m_{0i} m_{1i}}{T_i^3 - T_i^2}} \quad (14)$$

$$Q_{CMH} \text{ and } Q'_{CMH} \sim \chi_1^2$$

Decision based on the critical value or critical level in a right-tailed test:

If $Q_{CMH} \leq {}_{1-\alpha}\chi_1^2$ or $P(\chi_1^2 \geq Q_{CMH}) \geq \alpha$, $H_0 : OR = 1$ is accepted

If $Q_{CMH} > {}_{1-\alpha}\chi_1^2$ or $P(\chi_1^2 \geq Q_{CMH}) < \alpha$, $H_0 : OR = 1$ is rejected

${}_{1-\alpha/2}\chi_1^2$ = critical value or quantile of order $1 - (\alpha/2)$ of a chi-square distribution with one degree of freedom.

$P(\chi_1^2 \geq Q_{CMH})$ = probability value to the right tail of a chi-square distribution with one degree of freedom.

It should be noted that Mantel and Haenszel [2], had previously presented a formula to define the confidence interval of the common odds ratio, but it is less efficient and precise, especially with small samples [10].

$$P \left[e^{\left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{Q_{CMH}}}\right) \ln(\widehat{OR}_{MH})} \leq OR_{MH} \leq e^{\left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{Q_{CMH}}}\right) \ln(\widehat{OR}_{MH})} \right] = 1 - \alpha \quad (15)$$

7. Effect Size

The magnitude of the effect size for the CMH test is the common *OR* that estimates the effect of X on Y and, for interpretation, it can be transformed into a Cohen's *d* with Chinn's formula [13]. Values of *d* lower than 0.2 reflect an effect size that is trivial, from 0.2 to 0.49 is small, from 0.5 to 0.79 is large, and higher than or equal to 0.8 is very large (Cohen, 1988).

$$d = \frac{\ln(\widehat{OR}_{MH})}{\pi/\sqrt{3}} \quad (16)$$

8. Tests of Homogeneity of Odds Ratios

In order to avoid overextending this manuscript, only the two most widely used tests for testing the assumption of homogeneity between *ORs* will be presented, namely: Woolf's (1957) test and Breslow-Day test (1980) with Tarone's (1985) correction. Nevertheless, it should be noted that there exist more tests, such as those developed by Bliss; Zelen; Yusuf, Peto, Lewis, Collins, and Sleight; Liang and Self; and Kulinskaya and Dollinger [14] [15] [16]. These last two are especially recommended for small samples [17].

8.1. Woolf's Test

Woolf's test is one of the first tests developed for assessing homogeneity of *ORs*. Woolf's method is based on a statistic first introduced by Cochran [1] and it aims to test the assumption of homogeneity of *ORs* for *k* independent strata from 2×2 tables (effect of a dichotomous variable X on a dichotomous variable Y). The calculation of Woolf's statistic is simple and its small sample performance is better than that of other tests [7]; it is available in programs such as Real Statistics Resource Pack for Excel [18] and R software package for statistical

analysis [19].

$$H_0 : OR_1 = OR_2 = \dots = OR_k$$

$$H_1 : \exists OR_i \neq OR_j ; i \neq j = 1, 2, \dots, k$$

Assumptions: 1) a random sample composed of, at least, 25 participants in each stratum; 2) two dichotomous qualitative variables: one is the independent or antecedent variable X and the other is the dependent or consequent variable Y; and 3) a polytomous variable Z that generates the k independent strata in the 2×2 tables and that acts as a confounding variable in the relationship between X and Y [20].

Test statistic:

$$Q_{Woolf} = \sum_{i=1}^k \frac{(\ln \widehat{OR}_i - \ln \widehat{OR}_{comun})^2}{\hat{\sigma}_{\ln(OR_i)}^2} \tag{17}$$

$$\ln(\widehat{OR}_i) = \ln\left(\frac{a_i d_i}{b_i c_i}\right) \tag{18}$$

$$\hat{\sigma}_{\ln(OR_i)}^2 = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i} \tag{19}$$

$$\widehat{OR}_{common} = e^{\frac{\sum_{i=1}^k (\ln(OR_i) / \hat{\sigma}_{\ln(OR_i)}^2)}{\sum_{i=1}^k (1 / \hat{\sigma}_{\ln(OR_i)}^2)}} \tag{20}$$

Sampling distribution and decision based on the critical value or critical level in a right-tailed test:

$$Q_{Woolf} \sim \chi_{k-1}^2$$

If $Q_{Woolf} \leq \chi_{k-1}^2(1-\alpha)$ or $P(\chi_{k-1}^2 \geq Q) \geq \alpha$, H_0 is accepted

If $Q_{Woolf} > \chi_{k-1}^2(1-\alpha)$ or $P(\chi_{k-1}^2 \geq Q) < \alpha$, H_0 is rejected

8.2. Breslow-Day Test Using Tarone's Bias Correction

The Breslow-Day test requires large samples in each stratum, and is not adequate to use it with small sample sizes ($n < 20$). Its test statistic is consistent, but it presents a slight bias, and that is why Tarone [5] developed a correction. Precisely, Breslow [21] recommends the use of this correction, although he points out that, when the sample size is large, the resulting change in the estimate is minimal and only perceptible beyond the third or fourth decimal place of the probability value. The Breslow-Day test using Tarone's correction is the most powerful when comparing various OR homogeneity tests, all of which are not very powerful with small strata sizes [17]. It is worth noting that the Breslow-Day test using Tarone's correction is the most frequently OR homogeneity test included in several statistical software packages available in the market and, consequently, it is the most reported OR homogeneity test in research articles [7].

The omnibus formulation of hypotheses and assumptions of the Breslow-Day

test, with or without the Tarone's correction, are the same as in the Woolf's test; nevertheless, its asymptotic approximation to the chi-square distribution with $k - 1$ degrees of freedom is better when Tarone's correction is used.

Breslow-Day test statistic is a goodness-of-fit test. The expected absolute frequencies are calculated under the assumption of a common OR estimated by the formula of Mantel and Haenszel [2].

$$Q_{BD} = \sum_{i=1}^k \frac{(a_i - E(a_i | OR_{MH}))^2}{E(a_i | OR_{MH})} + \frac{(b_i - E(b_i | OR_{MH}))^2}{E(b_i | OR_{MH})} + \frac{(c_i - E(c_i | OR_{MH}))^2}{E(c_i | OR_{MH})} + \frac{(d_i - E(d_i | OR_{MH}))^2}{E(d_i | OR_{MH})} \tag{21}$$

$a_j = n_{11j}$ = absolute (joint) frequency of exposed cases in stratum j .

$b_j = n_{10j}$ = absolute (joint) frequency of exposed non-cases in stratum j .

$c_j = n_{01j}$ = absolute (joint) frequency of non-exposed cases in stratum j .

$d_j = n_{00j}$ = absolute (joint) frequency of non-exposed non-cases in stratum j .

OR_{MH} = common odds ratio of k strata estimated through Mantel-Haenszel method

$E(a_j | OR_{MH})$ = expected value for the number of exposed cases in stratum j conditioned on OR_{MH} . This value is obtained upon solving the following quadratic equation:

$$\begin{aligned} & (1 - \widehat{OR}_{MH}) \times E^2(a_i | OR_{MH}) + [n_{0i} - m_{1i} + \widehat{OR}_{MH} (n_{1i} + m_{1i})] \\ & \times E(a_i | OR_{MH}) - m_{1i} n_{1i} \widehat{OR}_{MH} = 0 \end{aligned} \tag{22}$$

$$\alpha = 1 - \widehat{OR}_{MH}$$

$$\beta = n_{0i} - m_{1i} + \widehat{OR}_{MH} (n_{1i} + m_{1i})$$

$$\gamma = -m_{1i} n_{1i} \widehat{OR}_{MH}$$

$$E(a_i | OR_{MH}) = \frac{-\beta \mp \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha} \tag{23}$$

Every quadratic equation has two solutions. In this case, the solution in the set of real numbers with a value higher than 0 and lower than or equal to the minimum of m_{1i} and n_{1i} is chosen. The expected values are not rounded to integers, but are handled with their decimal places.

$$0 < E(a_i | OR_{MH}) \leq \min(m_{1i}, n_{1i})$$

Once the expected frequency of the exposed cases, $E(a_i | OR_{MH})$, is estimated, the other expected frequencies can be obtained under the condition of a common odds ratio (OR_{MH}).

$$E(b_i | OR_{MH}) = n_{1i} - E(a_i | OR_{MH}) \tag{24}$$

$$E(c_i | OR_{MH}) = m_{1i} - E(a_i | OR_{MH}) \tag{25}$$

$$E(d_i | OR_{MH}) = T_i - m_{1i} - n_{1i} + E(a_i | OR_{MH}) \tag{26}$$

$$T_i = E(a_i | OR_{MH}) + E(b_i | OR_{MH}) + E(c_i | OR_{MH}) + E(d_i | OR_{MH})$$

Next, taking into account the expected values $E(a_i | OR_{MH})$ of each stratum, the Mantel-Haenszel common odds ratio (1959) can be calculated, and this allows checking that there are no errors in the calculation of the expected frequencies, as the OR_{MH} value is already known.

$$\widehat{OR}_{MH} = \frac{E(a_i | OR_{MH})(n_{0i} - m_{1i} + E(a_i | OR_{MH}))}{(m_{1i} - E(a_i | OR_{MH}))(n_{1i} - E(a_i | OR_{MH}))} \tag{27}$$

The squares of the differences between the observed and expected frequency within each stratum are equal.

$$\begin{aligned} (a_i - E(a_i | OR_{MH}))^2 &= (b_i - E(b_i | OR_{MH}))^2 = (c_i - E(c_i | OR_{MH}))^2 \\ &= (d_i - E(d_i | OR_{MH}))^2 \end{aligned}$$

Likewise, the variances of the frequencies conditioned to a common odds ratio OR_{MH} coincide within each stratum, and they are obtained through the following formula:

$$\begin{aligned} s^2(a_i | OR_{MH}) &= \frac{1}{\frac{1}{E(a_i | OR_{MH})} + \frac{1}{E(b_i | OR_{MH})} + \frac{1}{E(c_i | OR_{MH})} + \frac{1}{E(d_i | OR_{MH})}} \\ &= \frac{1}{\frac{1}{E(a_i | OR_{MH})} + \frac{1}{n_{1i} - E(a_i | OR_{MH})} + \frac{1}{m_{1i} - E(a_i | OR_{MH})} + \frac{1}{T_i - m_{1i} - n_{1i} + E(a_i | OR_{MH})}} \end{aligned} \tag{28}$$

Thus, the Breslow-Day statistic can be simplified to the following formula with k addends:

$$Q_{BD} = \sum_{i=1}^k \frac{(a_i - E(a_i | OR_{MH}))^2}{s^2(a_i | OR_{MH})} \tag{29}$$

Tarone's correction for the Breslow-Day statistic involves subtracting from the Q_{BD} a quotient whose numerator is the square of the difference between the sum of the exposed cases and the sum of their expected values conditioned to the Mantel-Haenszel common odds ratio, whereas its denominator is the sum of the variances of the exposed cases conditioned to the Mantel-Haenszel common odds ratio; hence, once the data for the Q_{BD} calculation are available, its correction is very simple.

$$\begin{aligned} Q_{BDT} &= Q_{BD} - \frac{(a - E(a | OR_{MH}))^2}{s^2(a | OR_{MH})} \\ &= \sum_{i=1}^k \frac{(a_i - E(a_i | OR_{MH}))^2}{s^2(a_i | OR_{MH})} - \frac{(a - E(a | OR_{MH}))^2}{s^2(a | OR_{MH})} \end{aligned} \tag{30}$$

$$a = \sum_{i=1}^k a_i$$

$$E(a | OR_{MH}) = \sum_{i=1}^k E(a_i | OR_{MH}) \tag{31}$$

$$s^2(a | OR_{MH}) = \sum_{i=1}^k s^2(a_i | OR_{MH}) \tag{32}$$

Sampling distribution and decision based on the critical value or critical level in a right-tailed test:

$$Q_{BDT} \sim \chi_{k-1}^2$$

If $Q_{BDT} \leq_{1-\alpha} \chi_{k-1}^2$ or $P(\chi_{k-1}^2 \geq Q) \geq \alpha$, H_0 is accepted

If $Q_{BDT} >_{1-\alpha} \chi_{k-1}^2$ or $P(\chi_{k-1}^2 \geq Q) < \alpha$, H_0 is rejected

9. Unfulfillment of the Assumption of Homogeneous Effect in the k Strata

In the event that the null hypothesis of equivalence of the odds ratio between the k strata is rejected, it is necessary to identify which strata are homogeneous and which strata are heterogeneous in order that the CMH test can be applied to the homogenous ones, and the effect size can be reported either by strata groups or individual strata.

Pairwise comparisons can be made with a Z test [7] with the Holm-Bonferroni or Holm-Sidak correction in order to control the increase in type-1 error or with the Benjamin-Hochberg correction in order to control the false discovery rate [22].

Statistical hypotheses:

$$H_0 : OR_i = OR_j; i \neq j = 1, 2, \dots, k$$

$$H_1 : OR_i \neq OR_j$$

Test statistic:

$$Z = \frac{\ln(\widehat{OR}_i) - \ln(\widehat{OR}_j)}{\sqrt{\hat{\sigma}_{\ln(OR_i)}^2 + \hat{\sigma}_{\ln(OR_j)}^2}} = \frac{\ln\left(\frac{a_i d_i}{b_i c_i}\right) - \ln\left(\frac{a_j d_j}{b_j c_j}\right)}{\sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}} \times \sqrt{\frac{1}{a_j} + \frac{1}{b_j} + \frac{1}{c_j} + \frac{1}{d_j}}} \sim N(0,1) \tag{33}$$

Decision based on the critical value or critical level in a two-tailed test:

If $z \leq z_{1-\frac{\alpha}{2}}$ or $2 \times (1 - P(Z \leq |z|)) \geq \alpha$, H_0 is accepted

If $z > z_{1-\frac{\alpha}{2}}$ or $2 \times (1 - P(Z \leq |z|)) < \alpha$, H_0 is rejected

Given k independent groups, the number of pairwise comparisons corresponds to combinatorics without repetition: k of 2, $C(k, 2)$, which means $m = k(k - 1)/2$ comparisons. To make the Holm-Bonferroni correction, the probabilities are ordered ascendingly: an order value (j) equal to 1 is assigned to the smallest probability, an order value (j) equal to 2 is given to the second smallest probability, and so on until assigning a value equal to m to the highest probability; in the event that two or more probability values are repeated, an average order value is as-

signed (sum of the corresponding orders divided by the number of added orders). Next, the significance level α (usually 0.05) is divided by $m + 1 - i$, thus obtaining the corrected significance level: $\alpha_{HB} = \alpha / (m + 1 - i)$. In the Holm-Sidak correction, the corrected significance level is: $\alpha_{HS} = 1 - (1 - \alpha)^{1 / (m + 1 - i)}$. The Benjamini-Hochberg corrected significance level is: $\alpha_{BH} = (i / m) \times \alpha$ [22]. See an example with four independent groups (Table 2).

$$m = \binom{k}{2} = \binom{4}{2} = \frac{k(k-1)}{2} = \frac{4 \times 3}{2} = \frac{12}{2} = 6$$

$$p_1 < p_2 < p_3 < p_4 < p_5 < p_6$$

If $p_i \geq \alpha_c$, H_0 is accepted

If $p_i < \alpha_c$, H_0 is rejected

A more general test than the CMH test is conditional logistic regression. This multivariate technique can include more than one antecedent or predictor variable, whether qualitative (dichotomous, polytomous, ordinal) or quantitative (discrete or continuous). When conditional logistic regression handles a single dichotomous predictor, apart from the confounding variable that defines k strata, the score test statistic corresponds to the CMH Q statistic [20]. Being θ the estimated regression parameter or weight, the score test statistic, denoted by $S(\theta)$, is the quotient between the derivative in θ of the natural logarithm of the likelihood function $\ln(\theta|x)$ and the Fisher information about the parameter θ . The statistic, in this one-parameter situation, follows a chi-square distribution with one degree of freedom, as the CMH Q statistic does.

$$S(\theta) = \frac{\frac{\delta}{\delta\theta} \ln \mathcal{L}(\theta|x)}{I(\theta)} = \frac{\frac{\delta}{\delta\theta} \ln(\mathcal{L}(\theta|x))}{-E \left[\frac{\delta^2}{\delta\theta^2} \ln(f_x(x|\theta)) \right]} = Q_{CMH} \sim \chi_1^2$$

Conditional logistic regression was developed in 1978 by Breslow, Day, Halvorsen, Prentice, and Sabai [23]. When data for the predicted dichotomous variable and predictors are recorded by stratum, this technique has a clear

Table 2. Holm-Bonferroni, Holm-Sidak, and Benjamin-Hochberg corrections for a significance level α at 0.05 and four compared groups.

p_i	i	Holm-Bonferroni $\alpha_{HB} = \alpha / (m + 1 - i)$	Holm-Sidak $\alpha_{HS} = 1 - (1 - \alpha)^{1 / (m + 1 - i)}$	Benjamini-Hochberg $\alpha_{BH} = (i / m) \times \alpha$
p_1	1	$0.05 / 6 = 0.0083$	$1 - (0.95)^{1/6} = 0.0085$	$(1/6) \times 0.05 = 0.0083$
p_2	2	$0.05 / 5 = 0.01$	$1 - (0.95)^{1/5} = 0.0102$	$(2/6) \times 0.05 = 0.0167$
p_3	3	$0.05 / 4 = 0.0125$	$1 - (0.95)^{1/4} = 0.0127$	$(3/6) \times 0.05 = 0.025$
p_4	4	$0.06 / 3 = 0.0167$	$1 - (0.95)^{1/3} = 0.0170$	$(4/6) \times 0.05 = 0.0333$
p_5	5	$0.05 / 2 = 0.025$	$1 - (0.95)^{1/2} = 0.0253$	$(5/6) \times 0.05 = 0.0467$
p_6	6	$0.05 / 1 = 0.05$	$1 - (0.95)^1 = 0.05$	$(6/6) \times 0.05 = 0.05$

advantage over unconditional logistic regression, especially when there are many small strata sizes, such as the extreme case of paired data or $n/2$ strata of size 2. Logistic regression includes the confounding variable that generates the k strata as one more predictor, and estimate location parameters for each of its k strata. Nevertheless, conditional logistic regression does not require calculating these parameters and achieves more efficiency in its estimates and power, which becomes more evident as the number of strata increases [4].

Conditional logistic regression is available in the R statistical software using the “clogit” function in the survival package. It is located in this package because the log-likelihood of a conditional logistic model is the same as the log-likelihood of a Cox model with a paired or stratified data structure [24].

10. Simulated Example of the Application of the CMH Test

In order to see a more realistic example, a feasible study in epidemiology is presented. A sample composed of 2000 Mexican males from 0 to 14 years old (pediatric population) was collected and classified into two groups: 1 = case of bronchial asthma and 0 = no case. The diagnosis of unspecified uncomplicated bronchial asthma (CA23.32) was made by pediatricians from public health centers using the criteria of the eleventh edition of the International Classification of Diseases (ICD-11) of the World Health Organization [25]. The exposure factor was established based on indoor smoking of tobacco at home by parents and other adults who reside with the minor: 1 = daily exposure to tobacco smoke at home and 0 = no exposure to household tobacco smoke. Tobacco is understood as any product processed from the leaves of the plant known as *Nicotianatabacum*, such as cigarettes, cigars, or pipe tobacco. Of the 2000 participants, 595 lived in an urban location with good air quality most days of the year, 1007 in a location with usually acceptable air quality, and 398 with usually poor air quality, according to the official Mexican standard: NOM-172-SEMARNAT-2019 on Guidelines to Obtain and Communicate the Air Quality and Health Risks Index [26]. The objective of the study is to verify the relation between daily exposure to tobacco smoke at home and the child’s bronchial asthma by controlling the confounding variable (air quality in the urban area of residence) in males aged 0 to 14 years using the CMH test [2] [6] with a significance level at 5%. In case of positive association, the effect size with the Mantel-Haenszel common odds ratio [2] is estimated. Also, the assumption of homogeneous odds ratio across the three strata is tested by the Woolf’s test [3] and Breslow-Day test with the Tarone’s correction [5] with a significance level at 5%.

Table 3 shows the three 2×2 contingency tables stratified by the confounding variable (air quality in the urban area of residence) of the pediatric participants (0 to 14 years of age). The three strata are good, fair, and poor air quality most days of the year. The two crossed dichotomous variables are: X = daily exposure to tobacco smoke at home = {0 = no, 1 = yes} and Y = bronchial asthma = {0 = no case, 1 = case}.

Table 3. Joint absolute and marginal frequencies of daily exposure to tobacco smoke at home and whether or not the participant is a case of bronchial asthma, stratifying by the confounding variable (air quality in the urban area of residence) of the participants.

Daily exposure	Air quality in the urban area of residence								
	Good			Fair			Poor		
	Case	No C	Σ_{row}	Case	No C	Σ_{row}	Case	No C	Σ_{row}
1 = Yes	80	381	461	145	642	787	62	210	272
0 = No	13	121	134	26	194	220	24	102	126
Σ_{column}	93	502	595	171	836	1007	86	312	398

Note. X = daily exposure to tobacco smoke at home = {0 = no, 1 = yes}, Y = bronchial asthma = {0 = no case, 1 = case}, Z = air quality in the urban area of residence = {1 = good, 2 = fair, 3 = poor}, Σ_{column} = sum per column in each stratum, Σ_{row} = sum per row in each stratum.

The first step is to calculate the odds ratio of each stratum both in its point estimate, using Equation (1), and 95% interval confidences, using Equations(2), (3), (5), and (6). In order to facilitate these calculations, **Table 4** has been created. The significance of each *OR* is calculated with the Z-test (Equation 4), and the effect size in each stratum with the *OR* transformed into a Cohen’s *d* with Chinn’s formula [13] that appears in the Equation (16).

Let Z be the confounding variable, that is, air quality in the urban area of residence of the participant: Z = {1 = good, 2 = fair, 3 poor}.

The *OR* for strata 1 (good air quality) and 2 (fair quality) are significant with a significance level at 5%, since they do not include the value 1 when they are estimated with a confidence level at 95%, and the two-tailed probability-values of their Z-test statistics are lower than the α significance level of 0.05. They reflect a small effect size ($0.2 \leq d < 0.5$) and a detrimental or health risk effect ($OR > 1$). The *OR* of stratum 3 (poor air quality) is not significant at significance level equal to 0.05 and shows a trivial effect size ($d < 0.2$). Nevertheless, the 95% confidence intervals of the three *OR* overlap, so that they seem to be equivalent from this first analysis, and the significant, small, and detrimental effect that daily exposure to tobacco smoke at home has on bronchial asthma in pediatric male patients remains unclear.

$$\widehat{OR}_1 = \frac{a_1 d_1}{b_1 c_1} = \frac{80 \times 121}{381 \times 13} = \frac{9680}{4953} = 1.9544$$

$$\ln(\widehat{OR}_1) = \ln(1.9544) = 0.6701$$

$$\hat{\sigma}_{\ln(OR_1)} = \sqrt{\frac{1}{a_1} + \frac{1}{b_1} + \frac{1}{c_1} + \frac{1}{d_1}} = \sqrt{\frac{1}{80} + \frac{1}{381} + \frac{1}{13} + \frac{1}{121}} = 0.3167$$

$$Z = \frac{\ln(\widehat{OR}_1)}{\hat{\sigma}_{\ln(OR_1)}} = \frac{\ln(1.9544)}{0.3167} = \frac{0.6701}{0.3167} = 2.1156$$

$$Sig = 2 \times (1 - P(Z \leq 2.1156)) = 0.0344 < \alpha = 0.05$$

Table 4. Joint frequencies between X and Y, and total for each air quality stratum.

Z = air quality	a_i	b_i	c_i	d_i	T_i
1 = good	80	381	13	121	595
2 = fair	145	642	26	194	1007
3 = poor	62	210	24	102	398
Σ_i	287	1233	63	417	2000

Note. $a_i = n_{11i}$ = exposed cases in each stratum i , $b_i = n_{10i}$ = exposed non-cases in each stratum i , $c_i = n_{01i}$ = non-exposed cases in each stratum i , $d_i = n_{00i}$ = non-exposed non-cases in each stratum i , $T_i = n_{..i}$ = total participants in each stratum or sum per row (i from 1 to 3), Σ_i = sum by column i from 1 to 3.

$$P\left(e^{\frac{\ln(\widehat{OR}_1) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_1)}}{1}} \leq OR_1 \leq e^{\frac{\ln(\widehat{OR}_1) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_1)}}{1}} \right) = 1 - \alpha$$

$$P\left(e^{\ln(1.9544) - 1.96 \times 0.3167} \leq OR_1 \leq e^{\ln(1.9544) + 1.96 \times 0.3167} \right) = 0.95$$

$$P(OR_1 \in [1.0505, 3.6358]) = 0.95$$

$$1 \notin [1.0505, 3.6358] \Rightarrow OR_1 \neq 1, OR_1 > 1$$

$$d_1 = \frac{\ln(\widehat{OR}_1)}{\pi/\sqrt{3}} = \frac{\ln(1.9544)}{\pi/\sqrt{3}} = 0.3694$$

$$\widehat{OR}_2 = \frac{a_2 d_2}{b_2 c_2} = \frac{145 \times 194}{642 \times 26} = \frac{28130}{16692} = 1.6852$$

$$\ln(\widehat{OR}_2) = \ln(1.6852) = 0.5219$$

$$\hat{\sigma}_{\ln(OR_2)} = \sqrt{\frac{1}{a_2} + \frac{1}{b_2} + \frac{1}{c_2} + \frac{1}{d_2}} = \sqrt{\frac{1}{145} + \frac{1}{642} + \frac{1}{26} + \frac{1}{194}} = 0.2282$$

$$Z = \frac{\ln(\widehat{OR}_2)}{\hat{\sigma}_{\ln(OR_2)}} = \frac{\ln(1.6852)}{0.2282} = \frac{0.5219}{0.2282} = 2.2872$$

$$Sig = 2 \times (1 - P(Z \leq 2.2872)) = 0.0222 < \alpha = 0.05$$

$$P\left(e^{\frac{\ln(\widehat{OR}_2) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_2)}}{1}} \leq OR_2 \leq e^{\frac{\ln(\widehat{OR}_2) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_2)}}{1}} \right) = 1 - \alpha$$

$$P\left(e^{\ln(1.6852) - 1.96 \times 0.2282} \leq OR_2 \leq e^{\ln(1.6852) + 1.96 \times 0.2282} \right) = 0.95$$

$$P(OR_2 \in [1.0775, 2.6357]) = 0.95$$

$$1 \notin [1.0775, 2.6357] \Rightarrow OR_2 \neq 1, OR_2 > 1$$

$$d_2 = \frac{\ln(\widehat{OR}_2)}{\pi/\sqrt{3}} = \frac{\ln(1.6852)}{\pi/\sqrt{3}} = 0.2877$$

$$\widehat{OR}_3 = \frac{a_3 d_3}{b_3 c_3} = \frac{62 \times 102}{210 \times 24} = \frac{6324}{5040} = 1.2548$$

$$\ln(\widehat{OR}_3) = \ln(1.2548) = 0.2269$$

$$\hat{\sigma}_{\ln(OR_3)} = \sqrt{\frac{1}{a_3} + \frac{1}{b_3} + \frac{1}{c_3} + \frac{1}{d_3}} = \sqrt{\frac{1}{62} + \frac{1}{210} + \frac{1}{24} + \frac{1}{102}} = 0.2690$$

$$Z = \frac{\ln(\widehat{OR}_3)}{\hat{\sigma}_{\ln(OR_3)}} = \frac{\ln(1.2548)}{0.2690} = \frac{0.2269}{0.2690} = 0.8437$$

$$Sig = 2 \times (1 - P(Z \leq 0.8437)) = 0.3989 > \alpha = 0.05$$

$$P\left(e^{\ln(\widehat{OR}_3) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_3)}} \leq OR_3 \leq e^{\ln(\widehat{OR}_3) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_3)}} \right) = 1 - \alpha$$

$$P\left(e^{\ln(1.2548) - 1.96 \times 0.2690} \leq OR_3 \leq e^{\ln(1.2548) + 1.96 \times 0.2690} \right) = 0.95$$

$$P(OR_3 \in [0.7406, 2.1259]) = 0.95$$

$$1 \in [0.7406, 2.1259] \Rightarrow OR_3 = 1$$

$$d_3 = \frac{\ln(\widehat{OR}_3)}{\pi/\sqrt{3}} = \frac{\ln(1.2548)}{\pi/\sqrt{3}} = 0.1251$$

The second step is to calculate the common odds ratio for the three strata using the Mantel-Haenszel formula [2]. Firstly, Mantel-Haenszel point estimate is calculated using Equation (7).

$$\widehat{OR}_{MH} = \frac{\sum_{i=1}^k \frac{a_i d_i}{T_i}}{\sum_{i=1}^k \frac{b_i c_i}{T_i}} = \frac{\frac{80 \times 121}{595} + \frac{145 \times 194}{1007} + \frac{62 \times 102}{398}}{\frac{381 \times 13}{595} + \frac{642 \times 26}{1007} + \frac{210 \times 24}{398}} = \frac{60.0928}{37.5637} = 1.5998$$

Secondly, the 95% confidence interval for the Mantel-Haenszel common OR (OR_{MH}) is calculated by applying a logarithmic transformation upon OR_{MH} as shown in the Equation (8); next, the asymptotic standard deviation of the natural logarithm of OR_{MH} is computed [10], as in Equation (9); then, the lower and upper limits for the log-transformed OR_{MH} are obtained using the Wald's method [11], as in Equation (11); and lastly, the logarithmic transformation is undone in order to obtain the limits for the OR_{MH} as in the Equation (12). The effect size, which turns out to be small, is calculated with Equation (16), which transforms an OR into a Cohen's d .

$$\ln(OR_{MH}) = \ln(1.5998) = 0.4699$$

$$u = \sum_{i=1}^k u_i = \sum_{i=1}^k \frac{a_i \times d_i}{T_i} = \frac{80 \times 121}{595} + \frac{145 \times 194}{1007} + \frac{62 \times 102}{398} = 60.0928$$

$$\sum_{i=1}^k s_i u_i = \sum_{i=1}^k \left(\frac{a_i + d_i}{T_i} \times \frac{a_i \times d_i}{T_i} \right)$$

$$= \frac{80 + 121}{595} \times \frac{80 \times 121}{595} + \frac{145 + 194}{1007} \times \frac{145 \times 194}{1007} + \frac{62 + 102}{398} \times \frac{62 \times 102}{398}$$

$$= 21.4472$$

$$\begin{aligned}
 v &= \sum_{i=1}^k v_i = \sum_{i=1}^k \frac{b_i \times c_i}{T_i} = \frac{381 \times 13}{595} + \frac{642 \times 26}{1007} + \frac{210 \times 24}{398} = 37.5637 \\
 \sum_{i=1}^k t_i v_i &= \sum_{i=1}^k \left(\frac{b_i + c_i}{T_i} \times \frac{b_i \times c_i}{T_i} \right) \\
 &= \frac{381+13}{595} \times \frac{381 \times 13}{595} + \frac{642+26}{1007} \times \frac{642 \times 26}{1007} + \frac{210+24}{398} \times \frac{210 \times 24}{398} \\
 &= 23.9533 \\
 \sum_{i=1}^k (s_i v_i + t_i u_i) &= \sum_{i=1}^k s_i v_i + \sum_{i=1}^k t_i u_i \\
 &= \frac{80+121}{595} \times \frac{381 \times 13}{595} + \frac{145+194}{1007} \times \frac{642 \times 26}{1007} + \frac{62+102}{398} \times \frac{210 \times 42}{398} \\
 &\quad + \frac{381+13}{595} \times \frac{80 \times 121}{595} + \frac{642+26}{1007} \times \frac{145 \times 194}{1007} + \frac{210+24}{398} \times \frac{62 \times 102}{398} \\
 &= 13.6103 + 38.6456 = 52.2559 \\
 \hat{\sigma}_{\ln(OR_{MH})} &= \sqrt{\frac{1}{2} \left(\frac{\sum_{i=1}^k s_i u_i}{u^2} + \frac{\sum_{i=1}^k (s_i u_i + t_i v_i)}{uv} + \frac{\sum_{i=1}^k t_i v_i}{v^2} \right)} \\
 &= \sqrt{\frac{1}{2} \left(\frac{21.4472}{60.0928^2} + \frac{52.2559}{60.0928 \times 37.5637} + \frac{23.9533}{37.5637^2} \right)} \\
 &= \sqrt{\frac{0.0461}{2}} = 0.1518 \\
 P \left(e^{\frac{\ln(\widehat{OR}_{MH}) - z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}}{\hat{\sigma}_{\ln(OR_{MH})}} \leq OR_{MH} \leq e^{\frac{\ln(\widehat{OR}_{MH}) + z_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\ln(OR_{MH})}}{\hat{\sigma}_{\ln(OR_{MH})}}} \right) &= 1 - \alpha \\
 P \left(e^{\ln(1.5998) - 1.96 \times 0.1518} \leq OR_{MH} \leq e^{\ln(1.5998) + 1.96 \times 0.1518} \right) &= 0.95 \\
 P(OR_{MH} \in [1.1882, 2.1540]) &= 0.95 \\
 1 \notin [1.1882, 2.1540] &\Rightarrow OR_{MH} \neq 1, OR_{MH} > 1 \\
 d = \frac{\ln(\widehat{OR}_{MH})}{\pi/\sqrt{3}} = \frac{\ln(1.5998)}{\pi/\sqrt{3}} &= 0.2590
 \end{aligned}$$

Using the logarithmic transformation of the OR_{MH} and its standard deviation, the significance of the effect of X on Y can be tested with a Z test, just as it was done with the odds ratio in each one of the three strata [10]. See Equation (10). In a two-tailed test with a significance level of 5%, the Mantel-Haenszel common OR is different from 1. Thus, exposure acts as a risk factor for being a case.

$$\begin{aligned}
 Z &= \frac{\ln(\widehat{OR}_{MH})}{\hat{\sigma}_{MH}} = \frac{\ln(1.5998)}{0.1518} = \frac{0.4699}{0.1518} = 3.0959 > z_{0.975} = 1.96 \\
 2 \times (1 - P(Z \leq 3.0959)) &= 0.0020 < \alpha = 0.05
 \end{aligned}$$

Thirdly, the CMH Q statistic is computed using Equation (13). **Table 5** shows the frequencies of the exposed cases (a_i), the marginal frequencies

Table 5. Frequency of exposed cases, their expected values and their variances assuming independence between X and Y.

Air quality	a_i	n_{1i}	n_{0i}	m_{1i}	m_{0i}	T_i	e_i	s_i^2
Good	80	461	134	93	502	595	72.0555	13.7142
Fair	145	787	220	171	836	1007	133.6415	24.2629
Poor	62	272	126	86	312	398	58.7739	14.6230
Σ_i	287	1520	480	350	1650	2000	264.4708	52.6001

Note. Strata or categories of the confounding variable Z = air quality in the area of residence = {1 = good, 2 = fair, 3 = bad}, X = antecedent variable: daily exposure to tobacco smoke at home = {0 = no, 1 = yes}, Y = outcome variable: case of bronchial asthma = {0 = no, 1 = yes}, $a_i = n_{1i}$ = exposed cases in stratum i , (cases of bronchial asthma that were exposed to tobacco smoke at home in stratum i), $n_{1i} = n_{1\cdot}$ = total of participants (with and without bronchial asthma) exposed to tobacco smoke at home in stratum i , $n_{0i} = n_{0\cdot}$ = total of participants (with and without bronchial asthma) non-exposed to tobacco in stratum i , $m_{1i} = m_{1\cdot}$ = total cases of bronchial asthma in stratum i , $m_{0i} = m_{0\cdot}$ = total of non-cases (participants without bronchial asthma) in stratum i , $T_i = n_{\cdot i}$ = total number of participants in stratum i , $e_i = (n_{1i} \times m_{1i}) / T_i$ = expected number of a_i under the assumption of independence between X and Y, s_i^2 = variance of a_i under the assumption of independence between X and Y.

(n_{1i} , n_{0i} , m_{1i} , m_{0i}) and the total (T_i) in each stratum. With these data, the expected value, e_i , and the variance of the exposed cases in each stratum, s_i^2 , are calculated under an assumption of independence between X and Y.

The null hypothesis regarding of unitary OR is rejected with the Q statistic without the Yates' continuity correction. The latter statistic is calculated using Equation (14).

$$\begin{aligned}
 Q_{CMH} &= \frac{\left(\sum_{i=1}^k a_i - \sum_{i=1}^k e_i\right)^2}{\sum_{i=1}^k s_i^2} = \frac{\left(\sum_{i=1}^k a_i - \sum_{i=1}^k \frac{n_{1i}m_{1i}}{T_i}\right)^2}{\sum_{i=1}^k \frac{n_{0i}n_{1i}m_{0i}m_{1i}}{T_i^3 - T_i^2}} \\
 &= \frac{(287 - 264.4708)^2}{52.6001} = \frac{507.5630}{52.6001} = 9.6495 >_{0.95} \chi_1^2 = 3.8415 \\
 P(\chi_1^2 \geq 9.6495) &= 0.0019 < \alpha = 0.05
 \end{aligned}$$

Owing to the use of an asymptotic approximation of a discrete distribution to a chi-square distribution with one degree of freedom, it is advisable to use the Yates' correction. Nevertheless, considering that the sample size is large, the difference between the statistics without or with continuity correction is minimal, $Q_{CMH} = 9.6495$, $P(\chi_1^2 \geq 9.6495) = 0.0019$ and $Q'_{CMH} = 9.2259$, $P(\chi_1^2 \geq 9.6495) = 0.0024$, respectively.

$$\begin{aligned}
 Q'_{CMH} &= \frac{\left(\left|\sum_{i=1}^k a_i - \sum_{i=1}^k \frac{n_{1i}m_{1i}}{T_i}\right| - 0.5\right)^2}{\sum_{i=1}^k \frac{n_{0i}n_{1i}m_{0i}m_{1i}}{T_i^3 - T_i^2}} \\
 &= \frac{(|287 - 264.4708| - 0.5)^2}{52.6001} = \frac{485.2838}{52.6001} = 9.2259 >_{0.95} \chi_1^2 = 3.8415
 \end{aligned}$$

$$P(\chi_1^2 \geq 9.2259) = 0.0024 < \alpha = 0.05$$

Likewise, because the sample size is large, when the confidence interval of the common odds ratio is calculated with the Mantel-Haenszel formula [2], as shown in the Equation (15), 95% CI [1.1814, 2.1663], the result is very similar to that obtained with the formula given by Robins *et al.* [10], as shown in the Equation (12), 95% CI [1.1882, 2.1540]. The latter is more efficient and accurate; however, this difference becomes evident especially when the sample is small.

$$P \left(e^{\left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{\hat{\theta}_{CMH}}} \right) \ln(\widehat{OR}_{MH})} \leq OR_{MH} \leq e^{\left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{\hat{\theta}_{CMH}}} \right) \ln(\widehat{OR}_{MH})} \right) = 1 - \alpha$$

$$P \left(e^{\left(1 - \frac{1.96}{\sqrt{9.2259}} \right) \ln(1.5998)} \leq OR_{MH} \leq e^{\left(1 + \frac{1.96}{\sqrt{9.2259}} \right) \ln(1.5998)} \right) = 0.95$$

$$P(1.1814 \leq OR_{MH} \leq 2.1663) = 0.95$$

Next, the assumption of homogeneity of the *OR* across the three strata is tested using the Woolf's test [3], as shown in Equation (17). The required calculations for this test are presented in **Table 6**, which includes the odds ratios (*OR_i*), the logarithmic transformations of these odds ratios ($\ln(OR_i)$) using Equation (18), and the variances of the logarithms of the odds ratio for each stratum ($s^2[\ln(OR_i)]$) using Equation (19). The natural base exponential of the quotient between the sum of the three quantities in the fifth column of **Table 6** and the sum of the three quantities in the sixth column of the same table yields the common odds ratio, calculated with Equation (20).

$$\widehat{OR}_{common} = e^{\frac{\sum_{i=1}^k (\ln(OR_i) / \sigma_{\ln(OR_i)}^2)}{\sum_{i=1}^k (1 / \sigma_{\ln(OR_i)}^2)}} = e^{\frac{19.8392}{42.9932}} = e^{0.4615} = 1.5864$$

Table 6. Odds ratio by stratum and calculations for Woolf's test.

AQ	<i>OR_i</i>	$\ln(OR_i)$	$s^2[\ln(OR_i)]$	$\ln(OR_i) / s^2$	$1 / s^2[\ln(OR_i)]$	$[\ln(OR_i) - \ln(OR)]^2 / s^2$
G	1.9544	0.6701	0.1003	6.6798	9.9689	0.4339
F	1.6852	0.5219	0.0521	10.0231	19.2048	0.0702
P	1.2548	0.2269	0.0724	3.1363	13.8195	0.7600
Σ_i				19.8392	42.9932	1.2640

Note. Strata or categories of the confounding variable $Z = AQ = \text{air quality} = \{G = \text{good}, F = \text{fair}, P = \text{poor}\}$, OR_i = odds ratio of stratum i , $\ln(OR_i)$ = natural logarithm of the odds ratio of stratum i , $s^2[\ln(OR_i)]$ = variance of the natural logarithm of the odds ratio of stratum i , $\ln(OR_i) / s^2[\ln(OR_i)]$ = quotient between the natural logarithm of the odds ratio of stratum i and its variance, $1 / s^2[\ln(OR_i)]$ = inverse of the variance of the natural logarithm of the odds ratio of stratum i , $[\ln(OR_i) - \ln(OR)]^2 / s^2[\ln(OR_i)]$ = quotient between the square of the difference between the natural logarithm of the odds ratio of stratum i and the natural logarithm of the common odds ratio (numerator) and the variance of the natural logarithm of the odds ratio of stratum i (denominator), Σ_i = sum per column (i from G to P).

The sum of the three quantities in the sixth column of **Table 6** yields the Woolf's Q test statistic. Each of these three quantities is the quotient between the square of the difference between the natural logarithm of the odds ratio of each stratum and the natural logarithm of the common odds ratio (numerator) and the variance of the natural logarithm of the odds ratio of each stratum (denominator), as shown in Equation (17).

$$Q_{Woolf} = \sum_{i=1}^k \frac{(\ln \widehat{OR}_i - \ln \widehat{OR}_{common})^2}{\hat{\sigma}_{\ln(OR_i)}^2} = 0.4339 + 0.0702 + 0.7600 = 1.2640$$

The value of the test statistic Q is small and lower than the quantile of order 0.95 of a chi-square distribution with two degrees of freedom ($k - 1 = 3 - 1 = 2$); thus, the null hypothesis of equality of effect across the three strata is hold.

$$Q_{Woolf} = 1.2640 < {}_{1-\alpha} \chi_{k-1}^2 = {}_{0.95} \chi_2^2 = 5.991$$

$$P(\chi_2^2 \geq Q_{Woolf} = 1.2640) = 0.5315 > \alpha = 0.05$$

Finally, the assumption of homogeneity of the effect across the three strata is tested by Breslow-Day test with Tarone's correction [5], as shown in Equation (21). As in the case of the Woolf's test, $Q_{Woolf} = 1.2640 < {}_{0.95} \chi_2^2 = 5.991$, $P(\chi_2^2 \geq 1.2640) = 0.5315 > \alpha = 0.05$, the null hypothesis of equivalent odds ratio across the three strata is fulfilled. Because the samples per stratum are large, the difference of the Breslow-Day test in regard to Woolf's test is very small, $Q_{BD} = 1.270705$, $P(\chi_2^2 \geq 1.270705) = 0.529749$, and it becomes minimal upon applying the Tarone's correction, $Q_{BDT} = 1.270382$, $P(\chi_2^2 \geq 1.270382) = 0.529834$.

The computation of the Breslow-Day test statistic starts with the calculation of the expected value, $E(a_1 | OR_{MH})$, using Equation (23), and the variance of the number of exposed cases in relation to the Mantel-Haenszel common OR in the first stratum, $s^2(a_1 | OR_{MH})$, using Equation (28). In order to obtain the expected value $E(a_1 | OR_{MH})$, it is necessary to pose a quadratic equation that should be solved, which is shown in Equation (22). The joint and marginal frequencies observed for stratum 1 (good air quality) are reproduced in **Table 7** in order to facilitate the follow-up of the calculations.

The expected value of a_1 (exposed cases from the first stratum) conditioned to the common OR estimated by the Mantel and Haenszel formula [2], upon having solved the quadratic equation and chosen a result within the range of 0 to 93 (minimum value for n_{11} and m_{11}), is equal to 77.8904. It can be verified that this expected frequency is close to the observed one, which is equal to 80.

$$\begin{aligned} & (1 - \widehat{OR}_{MH}) \times E^2(a_1 | OR_{MH}) + [n_{01} - m_{11} + \widehat{OR}_{MH} (n_{11} + m_{11})] \\ & \times E(a_1 | OR_{MH}) - m_{11} n_{11} \widehat{OR}_{MH} = 0 \\ & (1 - 1.5998) \times E^2(a_1 | OR_{MH}) + [134 - 93 + 1.5998 \times (461 + 93)] \\ & \times E(a_1 | OR_{MH}) - 93 \times 461 \times 1.5998 = 0 \end{aligned}$$

Table 7. Joint and marginal frequencies of stratum 1 (good air quality).

X_i	Y = 1 = case	Y = 0 = non case	Σ_j
X = 1 = exposed	$a_1 = 80$	$b_1 = 381$	$n_{11} = 461$
X = 0 = non exposed	$c_1 = 13$	$d_1 = 121$	$n_{01} = 134$
Σ_i	$m_{11} = 93$	$m_{01} = 502$	$T_1 = 595$

Note. X = exposure to indoor tobacco smoke at home, Y = bronchial asthma in pediatric minors. Σ_j = sum per rows ($j = 0$ and 1), Σ_i sum per columns ($i = 0$ and 1).

$$-0.5998 \times E^2(a_1 | OR_{MH}) + 927.2668 \times E(a_1 | OR_{MH}) - 68586.4895 = 0$$

$$\alpha = -0.5998$$

$$\beta = 927.2668$$

$$\gamma = -68586.4895$$

$$E(a_1 | OR_{MH}) = \frac{-\beta \mp \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

$$= \frac{-927.2668 \mp \sqrt{927.2668^2 - 4 \times (-0.5998) \times (-68586.4895)}}{2 \times (-0.5998)}$$

$$= \begin{cases} 1468.1739 \\ 77.8904 \end{cases}$$

$$0 < E(a_1 | OR_{MH}) = 77.8904 \leq \min(m_{11}, n_{11}) = \min(93, 461) = 93$$

With $E(a_1 | OR_{MH})$, the expected values for the other three cells of the 2×2 table of the first stratum are computed. The expected frequencies are close to the observed ones, namely: $b_1 = 381$, $c_1 = 13$, and $d_1 = 121$. These calculations correspond to Equations (24), (25), and (26).

$$E(b_1 | OR_{MH}) = n_{11} - E(a_1 | OR_{MH}) = 461 - 77.8904 = 383.1096$$

$$E(c_1 | OR_{MH}) = m_{11} - E(a_1 | OR_{MH}) = 93 - 77.8904 = 15.10960$$

$$E(d_1 | OR_{MH}) = T_1 - m_{11} - n_{11} + E(a_1 | OR_{MH})$$

$$= 595 - 461 - 93 - 77.8904 = 118.8904$$

The sum of the four expected frequencies is equal to the total of 595 participants in the first stratum, and it can be verified that the expected frequency for a_1 allows to correctly obtain the OR_{MH} . See Equation (27).

$$T_1 = E(a_1 | OR_{MH}) + E(b_1 | OR_{MH}) + E(c_1 | OR_{MH}) + E(d_1 | OR_{MH})$$

$$= 77.8904 + 383.1096 + 15.10960 + 118.8904 = 595$$

$$\widehat{OR}_{MH} = \frac{E(a_1 | OR_{MH})(n_{01} - m_{11} + E(a_1 | OR_{MH}))}{(m_{11} - E(a_1 | OR_{MH}))(n_{11} - E(a_1 | OR_{MH}))}$$

$$= \frac{77.8904 \times (134 - 94 + 77.8904)}{(94 - 77.8904)(461 - 77.8904)} = 1.5998$$

The variance of the exposed cases, conditioned to the common OR estimated by the formula of Mantel and Haenszel formula [2] in the first stratum,

is equal to 11.1058. See Equation (28).

$$s^2(a_1 | OR_{MH}) = \frac{1}{\frac{1}{E(a_1 | OR_{MH})} + \frac{1}{E(b_1 | OR_{MH})} + \frac{1}{E(c_1 | OR_{MH})} + \frac{1}{E(d_1 | OR_{MH})}}$$

$$= \frac{1}{\frac{1}{77.8904} + \frac{1}{383.1096} + \frac{1}{15.1096} + \frac{1}{118.8904}} = 11.1058$$

We proceed in the same way with the calculation of the expected value and the conditioned variance to the common OR of the exposed cases in the second stratum. The observed joint and marginal frequencies of stratum 2 (fair air quality) are shown in **Table 8**.

The expected value of a_2 (exposed cases from the second stratum) conditioned to the Mantel-Haenszel common OR [2] in Equation (23), upon having solved the quadratic Equation (22) and chosen a result within the range of 0 to 171 (minimum of n_{12} and m_{12}), is equal to 143.9868. As it happened in the first stratum, this expected frequency is close to the observed one, which is equal to 145.

$$(1 - \widehat{OR}_{MH}) \times E^2(a_2 | OR_{MH}) + [n_{02} - m_{12} + \widehat{OR}_{MH}(n_{12} + m_{12})] \times E(a_2 | OR_{MH}) - m_{12}n_{12}\widehat{OR}_{MH} = 0$$

$$(1 - 1.5998) \times E^2(a_2 | OR_{MH}) + [220 - 171 + 1.5998 \times (787 + 171)] \times E(a_2 | OR_{MH}) - 171 \times 787 \times 1.5998 = 0$$

$$-0.5998 \times E^2(a_2 | OR_{MH}) + 1581.5696 \times E(a_2 | OR_{MH}) - 215290.8355 = 0$$

$$\alpha = -0.5998$$

$$\beta = 1581.5696$$

$$\gamma = -215290.8355$$

$$E(a_2 | OR_{MH}) = \frac{-\beta \mp \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

$$= \frac{-1581.5696 \mp \sqrt{1581.5696^2 - 4 \times (-0.5998) \times (-215290.8355)}}{2 \times (-0.5998)}$$

$$= \begin{cases} 2493.0195 \\ 143.9868 \end{cases}$$

$$0 < E(a_2 | OR_{MH}) = 143.9868 \leq \min(m_{12}, n_{12}) = \min(171, 787) = 171$$

The expected frequencies for b_2 , c_2 and d_2 are also close to the observed ones, which are equal to 642, 26, and 194, respectively. These calculations correspond to Equations (24), (25), and (26).

$$E(b_2 | OR_{MH}) = n_{12} - E(a_2 | OR_{MH}) = 787 - 143.9868 = 643.0132$$

$$E(c_2 | OR_{MH}) = m_{12} - E(a_2 | OR_{MH}) = 171 - 143.9868 = 27.0132$$

$$E(d_2 | OR_{MH}) = T_2 - m_{12} - n_{12} + E(a_2 | OR_{MH})$$

$$= 1007 - 171 - 787 - 143.9868 = 192.9868$$

Table 8. Joint and marginal frequencies of stratum 2 (fair air quality).

X_j	Y = 1 = case	Y = 0 = non case	Σ_j
X = 1 = exposed	$a_2 = 145$	$b_2 = 642$	$n_{12} = 787$
X = 0 = non exposed	$c_2 = 26$	$d_2 = 196$	$n_{02} = 220$
Σ_i	$m_{12} = 171$	$m_{02} = 836$	$T_2 = 1007$

Note. X = exposure to indoor tobacco smoke at home, Y = bronchial asthma in pediatric minors. Σ_j = sum per rows ($j = 0$ and 1), Σ_i sum per columns ($i = 0$ and 1).

The sum of the four expected frequencies is equal to the total of 1007 participants in the second stratum, and it can be verified that the expected value of a_2 allows to correctly obtain the OR_{MH} . See Equation (27).

$$T_2 = E(a_2 | OR_{MH}) + E(b_2 | OR_{MH}) + E(c_2 | OR_{MH}) + E(d_2 | OR_{MH})$$

$$= 143.9868 + 643.0132 + 27.0132 + 192.9868 = 1007$$

$$\widehat{OR}_{MH} = \frac{E(a_2 | OR_{MH})(n_{02} - m_{12} + E(a_2 | OR_{MH}))}{(m_{12} - E(a_2 | OR_{MH}))(n_{12} - E(a_2 | OR_{MH}))}$$

$$= \frac{143.9868 \times (787 - 171 + 143.9868)}{(171 - 143.9868)(787 - 143.9868)} = 1.5998$$

The variance of the exposed cases, a_2 , conditioned to the common OR estimated by the Mantel and Haenszel formula [2] in the second stratum, is equal to 19.7235. See Equation (28).

$$s^2(a_2 | OR_{MH}) = \frac{1}{\frac{1}{E(a_2 | OR_{MH})} + \frac{1}{E(b_2 | OR_{MH})} + \frac{1}{E(c_2 | OR_{MH})} + \frac{1}{E(d_2 | OR_{MH})}}$$

$$= \frac{1}{\frac{1}{143.9868} + \frac{1}{643.0132} + \frac{1}{27.0132} + \frac{1}{192.9868}} = 19.7235$$

Finally, the expected values and the variance conditioned to the common OR of the cases exposed in the third stratum are calculated. The observed joint and marginal frequencies of stratum 3 (poor air quality) are shown in **Table 9**.

The expected value of a_3 (exposed cases from the third stratum) conditioned to the Mantel-Haenszel common OR [2] in Equation (23), upon having solved the quadratic Equation (22) and chosen a result within the range of 0 to 86 (minimum of n_{13} and m_{13}), is equal to 65.2416. As in the first and second strata, this expected frequency is close to the observed one, which is equal to 62.

$$(1 - \widehat{OR}_{MH}) \times E^2(a_3 | OR_{MH}) + [n_{03} - m_{13} + \widehat{OR}_{MH}(n_{13} + m_{13})]$$

$$\times E(a_3 | OR_{MH}) - m_{13}n_{13}\widehat{OR}_{MH} = 0$$

$$(1 - 1.5998) \times E^2(a_3 | OR_{MH}) + [126 - 86 + 1.5998 \times (272 + 86)]$$

$$\times E(a_3 | OR_{MH}) - 86 \times 272 \times 1.5998 = 0$$

Table 9. Joint and marginal frequencies of stratum 3 (poor quality).

X_i	Y = 1 = case	Y = 0 = non case	Σ_j
X = 1 = exposed	$a_3 = 62$	$b_3 = 210$	$n_{13} = 272$
X = 0 = non exposed	$c_3 = 24$	$d_3 = 102$	$n_{03} = 126$
Σ_i	$m_{13} = 86$	$m_{03} = 312$	$T_3 = 398$

Note. X = exposure to indoor tobacco smoke at home, Y = bronchial asthma in pediatric minors. Σ_j = sum per rows ($j = 0$ and 1), Σ_i sum per columns ($i = 0$ and 1).

$$-0.5998 \times E^2(a_3 | OR_{MH}) + 612.7139 \times E(a_3 | OR_{MH}) - 37421.5744 = 0$$

$$\alpha = -0.5998$$

$$\beta = 612.7139$$

$$\gamma = -37421.5744$$

$$E(a_3 | OR_{MH}) = \frac{-\beta \mp \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

$$= \frac{-612.7139 \mp \sqrt{612.7139^2 - 4 \times (-0.5998) \times (-37421.5744)}}{2 \times (-0.5998)}$$

$$= \begin{cases} 956.3577 \\ 65.2416 \end{cases}$$

$$0 < E(a_3 | OR_{MH}) = 65.2416 \leq \min(m_{13}, n_{13}) = \min(86, 272) = 86$$

The expected frequencies for b_3 , c_3 , and d_3 are also close to the observed ones, which are equal to 210, 24, and 102, respectively. These calculations correspond to Equations (24), (25), and (26).

$$E(c_3 | OR_{MH}) = m_{13} - E(a_3 | OR_{MH}) = 86 - 65.2416 = 20.7584$$

$$E(d_3 | OR_{MH}) = T_3 - m_{13} - n_{13} + E(a_3 | OR_{MH})$$

$$= 398 - 86 - 272 - 65.2416 = 105.2416$$

The sum of the four expected frequencies is equal to the total of 398 participants in the third stratum, and it can be verified that the expected value of a_3 allows to correctly obtain the OR_{MH} . See Equation (27).

$$s^2(a_3 | OR_{MH}) = \frac{1}{\frac{1}{E(a_3 | OR_{MH})} + \frac{1}{E(b_3 | OR_{MH})} + \frac{1}{E(c_3 | OR_{MH})} + \frac{1}{E(d_3 | OR_{MH})}}$$

$$= \frac{1}{\frac{1}{65.2416} + \frac{1}{206.7584} + \frac{1}{20.7584} + \frac{1}{105.2416}} = 12.8470$$

Once the expected values of the four frequencies of each stratum and the variances of the exposed cases with regard to the common OR of each stratum are available, the Breslow-Day Q test statistic [4] can be calculated, as shown in Equation (21). This statistic can be calculated using two formulas. The first formula is a chi-square goodness-of-fit statistic and it needs the expected values for the four frequencies of each stratum. See Table 10. See Equation (28).

Table 10. Joint frequencies observed in the three strata, expected frequencies, and variances conditioned upon the assumption of a common *OR* estimated by the Mantel-Haenszel formula.

Z_k	X_i	Y_j	O_{ijk}	E_{ijk}	$(O_{ijk}-E_{ijk})^2$	$(O-E)^2/E$	$s^2(O_{ijk})$	$(O-E)^2/s^2$
1	1	1	80	77.8904	4.4504	0.0571	11.1058	0.4007
1	1	0	381	383.1096	4.4504	0.0116	11.1058	
1	0	1	13	15.1096	4.4504	0.2945	11.1058	
1	0	0	121	118.8904	4.4504	0.0374	11.1058	
2	1	1	145	143.9868	1.0265	0.0071	19.7235	0.0520
2	1	0	642	643.0132	1.0265	0.0016	19.7235	
2	0	1	26	27.0132	1.0265	0.0380	19.7235	
2	0	0	194	192.9868	1.0265	0.0053	19.7235	
3	1	1	62	65.2416	10.5079	0.1611	12.8470	0.8179
3	1	0	210	206.7584	10.5079	0.0508	12.8470	
3	0	1	24	20.7584	10.5079	0.5062	12.8470	
3	0	0	102	105.2416	10.5079	0.0998	12.8470	
Σ			2000	2000		1.2707		1.2707

Note. Z_k = confounding variable with three ordinal categories = air quality in the place of residence (k : 1 = good, 2 = fair, and 3 = poor), X_i = exposure with two categories (i : 0 = no and 1 = yes), Y_j = bronchial asthma in pediatric minors with two categories (j : 0 = non case and 1 = case), O_{ijk} = joint absolute frequency of category i of X and category j of Y in stratum k of Z , E_{ijk} = expected absolute frequency conditioned to a common odds ratio estimated by the Mantel-Haenszel formula, $(O_{ijk} - E_{ijk})^2$ = square of the difference between the observed and expected probability, $(O_{ijk} - E_{ijk})^2/E_{ijk}$ probability divided by the expected frequency, $s^2(O_{ijk}|) = Var(O_{ijk}|OR_{MH})$ = variance of the observed frequencies conditioned to a common odds ratio estimated by the Mantel-Haenszel formula, $(O - E)^2/s^2 = (O_{11k} - E_{11k})^2/Var(O_{11k}|OR_{MH})$ = quotient between the square of the difference between the observed frequency and the expected frequency of the exposed cases (numerator) and the variance of the exposed cases conditioned to a common odds ratio estimated by the Mantel-Haenszel formula (denominator), Σ = sum per column.

$$Q_{BD} = \sum_{i=1}^k \frac{(a_i - E(a_i | OR_{MH}))^2}{E(a_i | OR_{MH})} + \frac{(b_i - E(b_i | OR_{MH}))^2}{E(b_i | OR_{MH})} + \frac{(c_i - E(c_i | OR_{MH}))^2}{E(c_i | OR_{MH})} + \frac{(d_i - E(d_i | OR_{MH}))^2}{E(d_i | OR_{MH})}$$

$$Q_{BD} = 0.0571 + 0.0116 + 0.2945 + 0.0374 + 0.0071 + 0.0016 + 0.0380 + 0.0053 + 0.1611 + 0.0508 + 0.5062 + 0.0998 = 1.2707$$

The second formula, given in Equation (30), is a simplified form of the first one, shown in Equation (21), considering that the squares of the differences between the observed and expected frequency for each stratum are equal, as well as the variances conditioned to a common odds ratio. This second for-

mula needs the expected frequencies of the exposed cases of the three strata and their variances conditioned to the Mantel-Haenszel common OR . This mathematical expectation is calculated with Equation (31) and the variance is calculated with Equation (32). See **Table 10**.

$$\begin{aligned}
 Q_{BD} &= \sum_{i=1}^k \frac{(a_i - E(a_i | OR_{MH}))^2}{s^2(a_i | OR_{MH})} \\
 &= \frac{(80 - 77.8904)^2}{11.1058} + \frac{(145 - 143.9868)^2}{19.7235} + \frac{(62 - 65.2416)^2}{12.8470} \\
 &= 0.400729 + 0.052047 + 0.817930 = 1.270705 < {}_{1-\alpha}\chi_{k-1}^2 = {}_{0.95}\chi_2^2 = 5.991 \\
 P(\chi_2^2 \geq 1.270705) &= 0.529749 > \alpha = 0.05
 \end{aligned}$$

The Breslow-Day statistic with Tarone's correction [5] undergoes a minimal change than without correction. See Equation (30), (31), and (32).

$$a = \sum_{i=1}^k a_i = 80 + 145 + 62 = 287$$

$$E(a | OR_{MH}) = \sum_{i=1}^k E(a_i | m_i, OR_{MH}) = 77.8904 + 143.9868 + 65.2416 = 287.1188$$

$$s^2(a | OR_{MH}) = \sum_{i=1}^k Var(a_i | m_i, OR_{MH}) = 11.1058 + 19.7235 + 12.8470 = 43.6763$$

$$\begin{aligned}
 Q_{BDT} &= Q_{BD} - \frac{(a - E(a | OR_{MH}))^2}{s^2(a | OR_{MH})} \\
 &= \sum_{i=1}^k \frac{a_i - E(a_i | m_i, OR_{MH})}{s^2(a_i | m_i, OR_{MH})} - \frac{(a - E(a | OR_{MH}))^2}{s^2(a | OR_{MH})} \\
 &= 1.270705 - \frac{(287 - 287.1188)^2}{43.6763} \\
 &= 1.270705 - 0.000323 = 1.270382 \\
 P(\chi_2^2 \geq 1.270382) &= 0.529834 > \alpha = 0.05
 \end{aligned}$$

11. Statistical Power and Type-2 Error

When rejecting the null hypothesis, it is important to also report the statistical power or probability of rejecting the null hypothesis conditional on it being false (ϕ). One way to calculate power is by the complement of the cumulative distribution function of the noncentral chi-square distribution or right-tailed probability. The degrees of freedom are 1 ($gl = 1$). The non-centrality parameter is the value of the test statistic: $PNC = Q_{CMH}$. This function is evaluated at the critical point: ${}_{1-\alpha}\chi_1^2$. If the null hypothesis is accepted, the type-2 error or probability of maintaining the null hypothesis conditional on it being false (β) is calculated, which is the probability at the left tail of the previously defined non-central distribution.

Returning to the example in which the null hypothesis of no effect was rejected and using the Q_{CMH} statistic with the Yates' correction, the statistical pow-

er is high:

$$\phi = 1 - \beta = 1 - NC \chi_{gl=1, PNC=9.2259}^2 \left({}_{0.95} \chi_1^2 = 3.8415 \right) = 1 - 0.1406 = 0.8594$$

12. Conclusions

Based on the analyses conducted on the simulated and realistic data presented in this study, it can be concluded that exposure to indoor tobacco smoke at home has a detrimental effect on bronchial asthma in pediatric males, when statistically controlling for the confounding variable of air quality in the urban area where the participants reside. The effect size of tobacco smoke is small, $OR_{MH} = 1.60$, $d = 0.26$, and its effect is homogeneous across the three strata: good, fair, and poor air quality. Exposure is 1.6 times more likely than non-exposure in males 0 - 14 years of age who suffer bronchial asthma compared with age-matched males without bronchial asthma. The implication that would follow from these results is that community preventive campaigns and pediatric interventions should focus on or take into account this health risk, namely smoking by parents or adults in the home. On the other hand, the confounding variable also has a significant effect, making air quality another health issue to be addressed by health authorities. These conclusions are similar to those obtained by Paciência, Cavaleiro-Rufo, and Moreira in a review study [27].

As has been done in this study, it is important to check the assumptions of the test, calculate the statistical power in case the null hypothesis of no effect is rejected or the type-2 error in case it holds, report the effect size in case the Mantel-Haenszel common OR is non-zero, and adjust for the limitations involved in using qualitative variables, one dichotomous as antecedent, another dichotomous as an outcome, and a third statistically controlled polychotomous variable. A broader alternative to the CMH test is conditional logistic regression. This latter statistical technique is used to analyze the relationship between a dependent variable and independent variables in matched or clustered data. It accounts for within-set dependencies and allows for controlling confounding factors. It estimates the odds ratio and uses maximum likelihood estimation to estimate regression coefficients, providing a valuable tool for analyzing matched or clustered data while controlling for confounders.

Acknowledgements

The authors thank the reviewers and editor for their helpful comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Cochran, W.G. (1954) On the Methods for Strengthening the Common χ^2 Tests. *Biometrics*, **10**, 417-451. <https://doi.org/10.2307/3001616>
- [2] Mantel, N. and Haenszel, W. (1959) Statistical Aspects of the Analysis of Data from

- Retrospective Studies of Disease. *Journal of the National Cancer Institute*, **22**, 719-748.
- [3] Woolf, B. (1955) On Estimating the Relation between Blood Group and Disease. *Annals of Human Genetics*, **19**, 251-253. <https://doi.org/10.1111/j.1469-1809.1955.tb01348.x>
- [4] Breslow, N.E. and Day, N.E. (1980) *Statistical Methods in Cancer Research: The Analysis of Case-Control Studies*. International Agency for Research on Cancer Scientific, Lyon.
- [5] Tarone, R.E. (1985) On Heterogeneity Tests Based on Efficient Scores. *Biometrika*, **72**, 91-95. <https://doi.org/10.1093/biomet/72.1.91>
- [6] Agresti, A. (2019) *An Introduction to the Categorical Data Analysis*. 3th Edition, John Wiley and Sons, Hoboken.
- [7] Yilmaz, A.E. and Altunay, S.A. (2022) Post-Hoc Comparison Tests for Odds Ratios. *Electronic Journal of Applied Statistical Analysis*, **15**, 75-94.
- [8] Schwarzer, G. and Rücker, G. (2021) Meta-Analysis of Proportions. In: Evangelou, E. and Veroniki, A.A., Ed., *Meta-Research: Methods and Protocols*, Springer, New York, 159-172. https://doi.org/10.1007/978-1-0716-1566-9_10
- [9] International Business Machines Corporation (2022) IBM SPSS Statistics Base 28. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Base.pdf
- [10] Robins, J.M., Breslow, N. and Greenland, S. (1986) Estimators of the Mantel-Haenszel Variance Consistent in Both Sparse Data and Large-Strata Limiting Models. *Biometrics*, **42**, 311-323. <https://doi.org/10.2307/2531052>
- [11] Böhning, D., Sangnawakij, P. and Holling, H. (2022) Confidence Interval Estimation for The Mantel—Haenszel Estimator of the Risk Ratio and Risk Difference in Rare Event Meta-Analysis with Emphasis on the Bootstrap. *Journal of Statistical Computation and Simulation*, **92**, 1267-1291. <https://doi.org/10.1080/00949655.2021.1991347>
- [12] Cucchetti, A. (2022) Analyzing Categorical Variable: Descriptive Statistics and Comparisons. In: Ceresoli, M., Abu-Zidan, F.M., Staudenmayer, K.L., Catena, F. and Coccolini, F., Eds., *Statistics and Research Methods for Acute Care and General Surgeons. Hot Topics in Acute Care Surgery and Trauma*, Springer, Cham, 67-73. https://doi.org/10.1007/978-3-031-13818-8_6
- [13] Deeks, J.J., Higgins, J.P. and Altman, D.G. (2019) Analysing Data and Undertaking Meta-Analyses. In: Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A., Eds., *Cochrane Handbook for Systematic Reviews of Interventions*, Wiley Online Library, New York, 1-48. <https://training.cochrane.org/handbook/current/chapter-10>
- [14] Kulinskaya, E. and Dollinger, M.B. (2015) An Accurate Test for Homogeneity of Odds Ratios Based on Cochran's Q-Statistic. *BMC Medical Research Methodology*, **15**, Article No. 49. <https://doi.org/10.1186/s12874-015-0034-x>
- [15] Zhang, C., Wang, X., Chen, M. and Wang, T.A. (2021) Comparison of Hypothesis Tests for Homogeneity in Meta-Analysis with Focus on Rare Binary Events. *Research Synthesis Methods*, **12**, 408-428. <https://doi.org/10.1002/jrsm.1484>
- [16] Shim, S.R., Kim, S.J. and Lee, J. (2019) Diagnostic Test Accuracy: Application and Practice Using R Software. *Epidemiology and Health*, **41**, e2019007. <https://doi.org/10.4178/epih.e2019007>
- [17] Almalik, O. and van den Heuvel, E.R. (2018) Testing Homogeneity of Effect Sizes in Pooling 2x2 Contingency Tables from Multiple Studies: A Comparison of Methods. *Cogent Mathematics & Statistics*, **5**, Article ID: 1478698.

- <https://doi.org/10.1080/25742558.2018.1478698>
- [18] Zaiontz, C. (2022) Cochran-Mantel-Haenszel (CMH) Test. In Real Statistics Using Excel. <https://www.real-statistics.com/chi-square-and-f-distributions/cochran-mantel-haenszel/>
- [19] Meyer, D., Zeileis, A., Hornik, K. and Friendly, M. (2022) Woolf Test for Homogeneity in $2 \times 2 \times k$ Tables. <https://rdrr.io/cran/DescTools/man/WoolfTest.html>
- [20] Vierra, A., Razzaq, A. and Andreadis, A. (2023) Categorical Variable Analyses: Chi-Square, Fisher Exact, and Mantel-Haenszel. In: Eltorai, A.E.M., Bakal, J.A., Newell, P.C. and Osband, A.J., Eds., *Translational Surgery. Handbook for Designing and Conducting Clinical and Translational Research*, Elsevier Science, Amsterdam, 171-175. <https://doi.org/10.1016/B978-0-323-90300-4.00095-1>
- [21] Breslow, N.E. (1996) Statistics in Epidemiology: The Case-Control Study. *Journal of the American Statistical Association*, **91**, 14-28. <https://doi.org/10.1080/01621459.1996.10476660>
- [22] Korthauer, K., Kimes, P.K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., Shukla, C., Alm, E.J. and Hicks, S.C. (2019) A Practical Guide to Methods Controlling False Discoveries in Computational Biology. *Genome Biology*, **20**, Article No. 118. <https://doi.org/10.1186/s13059-019-1716-1>
- [23] Breslow, N.E., Day, N.E., Halvorsen, K.T., Prentice, R.L. and Sabai, C. (1978) Estimation of Multiple Relative Risk Functions in Matched Case-Control Studies. *American Journal of Epidemiology*, **108**, 299-307. <https://doi.org/10.1093/oxfordjournals.aje.a112623>
- [24] Avalos, M., Pouyes, H., Kwemou, M. and Xu, B. (2022) Package ‘clogitLasso’. <https://cran.r-project.org/web/packages/clogitLasso/clogitLasso.pdf>
- [25] World Health Association (2020) Eleventh Revision of the International Classification of Diseases (ICD-11). Geneva. <https://www.who.int/classifications/icd/en/>
- [26] Secretaria de Gobernación (2019) Norma Oficial Mexicana NOM-172-SEMARNAT-2019. Lineamientos para la Obtención y Comunicación del Índice de Calidad del Aire y Riesgos a la Salud [Mexican Official Standard NOM-172-SEMARNAT-2019. Guidelines for Obtaining and Communicating the Air Quality Index and Health Risks]. *Diario Oficial de la Federación*. https://www.dof.gob.mx/nota_detalle.php?codigo=5579387&fecha=20/11/2019#gsc.tab=0
- [27] Paciência, I., Cavaleiro-Rufo, J. and Moreira, A. (2022) Environmental Inequality: Air Pollution and Asthma in Children. *Pediatric Allergy and Immunology*, **33**, e13818. <https://doi.org/10.1111/pai.13818>