



Measures of association in epidemiological studies: smoking mothers and low birth weight children in the city of Campina Grande – PB

Tarsyla Medeiros de Albuquerque¹, Gabriela Albuquerque Batista de Araujo¹, Bruno Leão Caminha¹, Marta Lucia de Albuquerque² and Mácio Augusto de Albuquerque^{3*}

¹Universidade Federal da Paraíba, João Pessoa, Pernambuco, Brazil. ²Faculdade de Ciências Sociais e Aplicadas, Campina Grande, Pernambuco, Brazil. ³Departamento de Estatística, Universidade Estadual da Paraíba, Rua Baraúnas, 351, 58429-500, Campina Grande, Pernambuco, Brazil. *Author for correspondence. E-mail: marcioaa@uepb.edu.br

ABSTRACT. The aim of this study was to present and describe the essential aspects, and to discuss the use of measures of association, relative risk and odds ratio, including formulas for calculating confidence intervals of obtained data from a cohort study of underweight live births of mothers who smoked during pregnancy, whose deliveries were performed in hospitals and maternity wards located in Campina Grande, state Pernambuco. Smoking during pregnancy was analyzed as a potential risk factor for low birth weight among 3612 newborns. In assessing the association of outcome, there were no large numeric differences between the estimates of the relative risk, odds ratios and confidence intervals obtained. It is an acceptable approximation to the relative risk and the odds ratio. It is up to the researcher to choose the most appropriate technique to its subject matter and should be determined according to the surveyed data.

Keywords: cohort study, relative risk, odds ratio.

Medidas de associação em estudos epidemiológicos: mães fumantes e crianças de baixo peso na cidade de Campina Grande –PB

RESUMO. O objetivo deste estudo foi apresentar e descrever os aspectos essenciais, bem como discutir a utilização das medidas de associação, risco relativo e razão de chances, incluindo fórmulas para o cálculo de intervalos de confiança de dados obtidos por meio de um estudo de coorte realizado com nascidos vivos de baixo peso de mães que fumaram durante a gravidez, cujos partos foram realizados em hospitais e maternidades localizados em Campina Grande, Pernambuco. O hábito de fumar durante a gravidez foi analisado como um potencial fator de risco para o baixo peso ao nascer entre 3612 recém nascidos. Na avaliação do desfecho de associação, não se observaram grandes diferenças numéricas entre as estimativas do risco relativo, da razão de chances e do intervalo de confiança obtidos. É uma aproximação aceitável para o risco relativo e para a razão de chances. Cabe ao pesquisador a escolha da técnica mais adequada ao seu objeto de estudo, devendo ser determinada de acordo com os dados pesquisados.

Palavras-chave: estudo de coorte, risco relativo, razão de chances.

Introduction

Smoking is considered a serious public health problem and a leading cause of morbidity and mortality worldwide. The incidence of smoking in women of childbearing age population has increased over the years. The use of cigarettes by this class becomes even more worrisome because studies show that smoking in pregnancy is not harmful only to the mother but also to the fetus.

When a woman smokes during pregnancy, exposes her fetus not only to cigarette smoke components that cross the placenta, but also to changes in oxygenation and placental metabolism, and changes in its own metabolism secondary

related to smoking. Smoking in pregnancy is responsible for 20% of fetuses with low birth weight, 8% of premature births and 5% of all perinatal deaths (Paranhos, Figueiredo Filho, Rocha, & Silva Júnior, 2013).

Moreover, international research conducted in the mid and late 1990s, both in Canada and in the United States, showed that among pregnant women, 20% smoked during pregnancy. Since about 4 to 5 million live births occur annually in North America, it is huge the number of children who will be born exposed to constituents of cigarette smoke due to maternal smoking - not to mention passive exposure to cigarette smoke, even if the mother is non-

smoker - and it has large, comprehensive repercussions for them.. Meta-analysis of 23 studies shows that children of mothers who smoke during pregnancy are around two times more likely to have low birth weight (< 2.500 g) at birth [relative risk (RR) = 1.82; IC95% (confidence interval of 95%): 1.67; 1.97] (Difranza & Lew, 1995). A work by (Pagano & Gauvreau, 2013). in Rio Grande do Sul confirms that, apart from the effects on weight, there is the negative impact of smoking in length and head circumference of newborns, indicating that maternal smoking during pregnancy inversely affected all three anthropometric measurements evaluated at birth that reflect the intrauterine growth.

The low birth weight occurs due both to premature birth, most common in developing countries, as the constraint of intrauterine growth, most frequently in developing countries, or as a combination of both. Factors associated with low birth weight are: maternal low height and weight, multiple births, low calorie intake, hypertension during pregnancy, maternal smoking, genetic syndromes, physical labor during pregnancy, maternal exposure to toxic substances and inappropriate prenatal. Low birth weight rates differ in various regions of the world, being higher in less developed countries, since they are associated with unfavorable socioeconomic conditions. Underweight rates at birth are estimated at 15% in developing countries and 7% in developed countries. In Brazil, in 2010, the rate was 9.1% (Veloso et al., 2014).

Maternal cigarette smoking in the third trimester of pregnancy is a strong predictor of low percentile of birth weight. Thus, the reduction or complete cessation of smoking for the mother during pregnancy will result in a higher birth weight, regardless of prenatal consumption levels. For each additional cigarette a day that the participant smoked in the third quarter, there was a 27 g reduction in the estimated birth weight. This low birth weight observed in newborns of smoking mothers may have consequences in the long run, since the evidence points to a significant pediatric and adult morbidity (Bernstein et al., 2005).

Considering all the harmful effects of tobacco, both for human health and for the environment, it is imperative to reduce smoking in all population groups. Given the almost universal antenatal care in urban areas of Brazil, pregnancy should be seen as the ideal time to encourage smoking cessation, since at this period there is an intensification of contacts with health professionals, thus providing an

opportunity so there is this incentive. In this sense, it is important to all professionals who are part of maternal and child care to guide the pregnant smoking women, highlighting the great harm on your health and especially about his son, both during intrauterine life, as after birth.

In epidemiological studies, it is common to wonder if certain characteristics of human life, as habits or aspects of the environment where the individual lives, are associated with certain disease, with manifestations of a disease or other events of interest to the researcher. Often the questioning is done to relate the attributes of a person with risk of developing certain event (Arango, 2009). For example, an investigator (physician) may want to assess whether infants born to mothers who smoked during pregnancy had characteristics resulting from contact with chemicals from the cigarette, as low weight and reduced head circumference.

To study a situation as exemplified above, we should be familiar with some basic terms used in epidemiological research, such as 'outcome' and 'risk factor'. Outcome is the name used to designate the event of interest in a search. The outcome may be the emergence of a disease, in a particular symptom, death or any other event that happens in the health-disease process. In the example above, the outcome is 'low weight and reduced head circumference'. But the risk factor (also known as 'factor under study') is the name used in Epidemiology to designate a variable that is supposed to be associated with the outcome. Often, people who have the supposed risk factor are designated as 'exposed'. In that instance, the risk factor is 'born to mothers who smoked during pregnancy'. Finally, we can, in a simplified manner, considering risk as the probability of an individual presenting the outcome (probability of developing low weight and reduced head circumference) at any given time. The risk is usually assessed in epidemiological studies by cumulative incidence.

There are some measures of association that have been developed to evaluate the relationship between the risk factor and the outcome (Margotto, 2015). Among these measures, it can be highlighted the relative risk (RR) and odds ratio (OR). Although they have a common goal (to assess the association between variables 'risk factor' and 'outcome' in epidemiological studies), these measures of association have their own characteristics and should be used in accordance with the employee research design. There are several designs ratings research proposed in the literature, which can be consulted, for details (Zychar, Borda, Moreira, Pereira, & Mário, 2016). This study will be focused on the

calculation of unadjusted estimates (i.e., who do not consider potential confounding factors) for RR and OR with their confidence intervals, indicating what kind of design each of these measures is more widely used.

Material and methods

The sample was composed of live births in hospitals and maternity hospitals in Campina Grande in 2012. Smoking during pregnancy has been studied as a potential risk factor for low birth weight between 3612 newborns. Obtaining measures of association for the epidemiological data are usually presented in tables 2 X 2 or contingency tables as shown in Table 1.

Table 1. Incidence of low birth weight in newborns in hospitals and maternity wards located in Campina Grande, Pernambuco State, according to smoking during pregnancy, in 2012.

Mother classification	Low birth weight		Total	Probability low birth weight
	Yes	No		
Smoker	137(a)	1072(b)	1209(a + b)	137/1209 = 0,113
Non-smoker	155(c)	2248(d)	2403(c + d)	155/2403 = 0,065
Total	292(a + c)	3320(b + d)	3612	292/3610 = 0,081

To study a presumed association between risk factors (smoking during pregnancy) and outcome (low birth weight), a cohort study was conducted. It is longitudinal, retrospective or prospective observational study, in which a defined group of people (cohort) is followed over a period of time. The outcomes are compared from exposure or not an intervention or other factor of interest. Is the study design most appropriate to describe the incidence and natural history of a condition (Glantz, 2013) In a cohort study, we can assess the risk by measuring the cumulative incidence of an outcome by the number of cases (new) occurred during the study period, divided by the population size (or sample) studied.

Relative risk and odds ratio

The relative risk or odds ratio are strength of association measures (effect measures), i.e., measure the association between outcome variable with the exposure variable: as the occurrence probability due to the dependent variable and its relationship with independent variable. The terms association and effect refers to the fact that a variable would have a relationship or exert an effect on another variable. The results are due to a reason, the null value for these is 1 (one). A relative risk or an odds ratio of 1.0 indicates that the probability of disease in the exposed and non-exposed groups are identical;

consequently, there is no association between exposure and disease. A RR or OR higher than 1.0 implies that there is increased risk of disease in individuals exposed, while a value less than 1.0 suggests that there is a reduced risk that the exposed individuals develop the disease (Margotto, 2015).

Relative risk

The relative risk is a measure of strength of association between exposure to the risk factor and the event (outcome), indicating how often the occurrence of the outcome in exposed is higher than that among non-exposed. The RR is defined as the ratio between the incidence of outcome in exposed and the incidence of outcome we unexposed. It is defined as the cohort, a group of individuals/ units of analysis who experienced the same event. Cohort studies, in turn, are observational studies in which individuals are classified/ selected according to the exposure status to a particular event. This type of study is fairly frequent in epidemiology to 'evaluate the incidence of the disease in a given period of time'. Cohort studies can also be used to assess the risks and benefits of the use of certain medication (Oliveira & Parente, 2010).

When one want to compare, in a cohort study, the incidence of an outcome between exposed mothers (smokers) with that obtained from unexposed mothers (non-smokers), usually calculate the relative risk. The RR can be used both to compare cumulative effects as compare incidence densities (Pagano & Gauvreau, 2013). In this study, cumulative incidences were considered, and the RR is calculated as Equation 1:

$$RR = \frac{P(\text{disease}|\text{exposed})}{P(\text{disease}|\text{non-exposed})}$$

$$RR = \frac{\text{outcome of the risk in exposed persons}}{\text{outcome of the risk in unexposed people}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \quad (1)$$

In order to demonstrate the calculation of a 95% confidence interval for the RR (the confidence interval shows the limit within which there is the certainty of the true treatment effect, estimates the magnitude of the association and informs the variability of the estimate through the lower and upper limits), it can be used the method described by the logarithmic transformation (Glantz, 2013). This method assumes that the sampling distribution of RR values has an asymmetrical shape of the log-normal type. Thus, by means of a logarithmic transformation, we obtain a curve with nearly normal way. Using formulas similar to those used

for the calculation of confidence intervals for variables with normal distribution, one can build a confidence interval for the logarithm of the RR, $\ln RR$. To express the confidence intervals in the original scale of the RR, just get the anti-logarithm of the found limits. The antilog is written in the formula $\exp[\ln RR]$ ou $e^{\ln RR}$.

According to the method of the logarithmic transformation, the formula for calculating the confidence interval (IC) RR is Equation 2:

$$IC_{RR} = \exp[\ln(RR) \pm Z_{\alpha} \times SE \ln(RR)]$$

where,

$$SE \ln(RR) = \sqrt{\left[\frac{1}{a} - \frac{1}{a+b} \right] + \left[\frac{1}{c} - \frac{1}{c+d} \right]} \quad (2)$$

Z_{α} = two-tailed critical limit for normal distribution

Odds ratio - OR

The odds ratio is another measure often used to compare the odds of an event in two groups. Unlike the relative risk, which compares directly the odds, the odds ratio, as its name suggests, relates the odds of the event in two populations. For women who smoke 21 cigarettes a day or more, but stopped in the last two years, the chance of developing lung cancer on the chance of women who never smoked would be calculated as Equation 3:

$$OR = \frac{P(\text{lung cancer}|\text{stopped})/[1-P(\text{lung cancer}|\text{stopped})]}{P(\text{lung cancer}|\text{non-smoker})/[1-P(\text{lung cancer}|\text{non-smoker})]} \quad (3)$$

In case-control studies, patients are included according to the presence or absence of the outcome. Usually, it is defined a group of cases (outcome) and other controls (with no outcome) and evaluated the exposure (in the past) to potential risk factors in these groups.

In case-control studies, patients are included according to the presence or absence of the outcome.

Due to the fact that the assembly of this type of study is based on the outcome itself, it cannot directly be estimated the incidence of outcome according to the presence or absence of the display, as is usual in cohort studies. This is due to the fact that the proportion of cases / controls or outcome / non-outcome is determined by the researcher. Thus, the occurrence of outcomes in the total study group is not governed by the natural history of the disease, and depends on how many cases and controls the researcher selected (Pagano & Gauvreau, 2013).

Although it cannot be directly estimated the incidence of the disease (outcome) among exposed

and non-exposed in case-control studies, it is possible, however, to estimate the ratio of these incidences, i.e., the RR. For this, the formula used to calculate the RR cohort study should be adapted to be applied in a case-control. This is due to the fact that if the outcome is sufficiently rare in the population (approximately 10% or less), the RR can be estimated, in case-control studies, by reason of 'odds' of exposure between cases and controls. This quantity is often called odds ratio.

If we have dichotomous random variables that represent a disease and exposure, the odds ratio is defined as the chance in favor of disease among exposed individuals divided by chance in favor of disease among non-exposed or Equation 4:

$$OR = \frac{P(\text{disease}|\text{exposed})/1-P(\text{disease}|\text{exhibition})}{P(\text{disease}|\text{unexposed})/[1-P(\text{disease}|\text{unexposed})]} \quad (4)$$

Alternatively, the odds ratio can be defined as the chance of exposure among sick individuals, divided by the chance of exposure among non-sick or Equation 5:

$$OR = \frac{P(\text{exhibition}|\text{sick})/1-P(\text{exhibition}|\text{sick})}{P(\text{exhibition}|\text{not-sick})/[1-P(\text{exhibition}|\text{not-sick})]} \quad (5)$$

These two different exposures to the relative chance are mathematically equivalent, so the odds ratio can be estimated for both cohort studies and case-control.

Our data consists of a sample of $n = (a + b + c + d)$ individuals and are arranged on a contingency table of 2×2 as shown in Table 1.

In this case, we can estimate that Equation 6:

$$P(\text{disease}|\text{exposed}) = \frac{a}{a+c},$$

$$P(\text{disease}|\text{non-exposed}) = \frac{b}{b+d} \quad \text{thus,}$$

$$1 - P(\text{disease}|\text{exposed}) = 1 - \frac{a}{a+c} = \frac{c}{a+c},$$

$$1 - P(\text{disease}|\text{non-exposed}) = 1 - \frac{b}{b+d} = \frac{d}{b+d} \quad (6)$$

With these results, they can Express an odds ratio estimator as:

$$RC = \frac{[a/(a+c)]/[c/(a+c)]}{[b/(b+d)]/[d/(b+d)]} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

When it is calculated a confidence interval for the odds ratio, it is necessary to make the same supposition of original normality. However, a problem arises in which the probability distribution of the odds ratio is skewed to the right. Although it cannot get negative values, the relative chance can take any positive value between zero and infinity. In

contrast, the probability distribution of the natural logarithm of the odds ratio is more symmetrical and approximately normal. So to calculate a confidence interval for the odds ratio, we typically work in logarithmic scale. To ensure that the sample size is sufficiently large, the expected value of each contingency table entry should be at least 5.

Results and discussion

In the present study, it was found in Table 1 that from the total of 3612 live births to mothers who smoke in Campina Grande, in 2012, 1,209 were children of smoking mothers and 2,403 non-smoking mothers, corresponding to 33.47 and 66.53% of smoking mothers and nonsmoking mothers, respectively. It was also observed that 292 newborns have low birth weight and 33.20 has not underweight, corresponding to 8.02 and 91.92% underweight and not underweight respectively.

The relative risk was calculated as shown in Table 1, according Equation 7:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{\frac{137}{1209}}{\frac{155}{2403}} = \frac{0.114}{0.065} = 1.75 \quad (7)$$

The relative risk of 1.75 implies that smoking mothers who have given birth have 75% of newborns born with 75% underweight.

The sampling variability of this finding can be evaluated using a significance test or through confidence interval. In this research, it will be given preference to the confidence interval. In this case, the $RR = 1.75$ represents a present effect in the population and not only in the sample, it can be calculated a confidence interval for this estimate. For a given level of significance, e.g., $\alpha = 0.05$, the confidence interval represents the interval in which must find the parameter, i.e., the true relative risk. If the value 1 (concerning the nullity of association) is not contained in the range, we have a $1 - \alpha$ trust that the population from which our sample was drawn from 1 and is therefore significant finding of the sample, according Equation 8.

$$\begin{aligned} \ln(RR) &= \ln(1.75) = 0.56 \\ E\ln(RR) &= \sqrt{\left[\frac{1}{137} - \frac{1}{1209} \right] + \left[\frac{1}{155} + \frac{1}{2403} \right]} = 0.111 \quad (8) \\ \text{inferior limit} &= \exp(0.56 - 1.96 \times 0.11) = 1.40 \\ \text{upper limit} &= \exp(0.56 + 1.96 \times 0.11) = 2.17 \end{aligned}$$

Thus its confidence interval with 95% confidence for relative risk found in Table 1, according Equation 9:

$$RR = 1.75(IC_{95\%}: 1.40 - 2.17) \quad (9)$$

So it is possible to say that the relative risk is significant, that is, the sample studied reflects a real effect of the risk factor in the population. Thus, it can be said that the risk of low weight for a newborn is about 1.7 times higher if his mother smoking during pregnancy than if she did not smoke, and it is estimated with 95% confidence that true RR is a range between 1.40 and 2.17. Therefore, since the incidence of low birth weight among newborns of smoking mothers is higher than the incidence of low birth weight among those children of non-smokers, it can be said that there is an association between smoking and low birth weight.

The calculation of the odds ratio as shown in Table 1, according Equation 10:

$$OR = \frac{135/155}{1072/2248} = \frac{137}{155} \cdot \frac{2248}{1072} = 1.85 \quad (10)$$

These data suggest that the chance of a smoker have a low weight child is 1.85 times the chance to non-smoking. Therefore, it can be stated that there is an association between smoking and low weight, which does not imply, however, that smoking causes, somehow, low weight; it is possible that children under higher risk are the children with low birth weight.

Using the same set of information, the value obtained for the measurement of the OR association is generally higher than that obtained through the RR model. However, it can be said that to the information in Table 1, the $OR = 1.85$ is a reasonable approximation for the relative risk $RR = 1.75$. As the measured event is smaller, this approach becomes progressively more accurate.

Similarly to the RR, it can be evaluated the variability of the sample through the OR confidence interval calculation. Thus, one can see and accept that the OR of sample values exhibits a log-normal distribution, which is normalized to the logarithmic transformation, according Equation 11.

$$IC_{OR} = \exp[\ln(OR) \pm Z_{\alpha} \times E\ln(OR)]$$

at where,

$$\text{standard error } E\ln(OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad (11)$$

Z_{α} = two-tailed critical limit for normal distribution.

Replacing the data in Table 1, we have:

$$\ln(OR) = \ln(1.85) = 0.615$$

If any of the contingency table entries is equal to 0, the standard error will be undefined. In this case, add 0.5 to each of the values a, b, c and d corrects the situation and also provides a reasonable estimate; thus, the standard error of the modified estimator is Equations 12 and 13:

Standard error

$$EP\ln(OR) = \sqrt{\frac{1}{(a+0.5)} + \frac{1}{(b+0.5)} + \frac{1}{(c+0.5)} + \frac{1}{(d+0.5)}} \quad (12)$$

$$EP\ln(OR) = \sqrt{\frac{1}{137} + \frac{1}{1072} + \frac{1}{155} + \frac{1}{2248}} = 0.122$$

$$\text{inferior limit} = \exp(0.615 - 1.96 \times 0.087) = 1.46$$

$$\text{upper limit} = \exp(0.615 + 1.96 \times 0.087) = 2.35 \quad (13)$$

this way, the odds ratio

for the table 1 can be expressed as:

$$OR = 1.85 (IC95\%: 1.46 - 2.35)$$

We are 95% confident that the chance that pregnant smokers may give birth to infants with low birth weight is 1.46 to 2.35 times higher than the chance of a newborn do not born with low weight.

Conclusion

Smoking is considered a serious public health problem and a leading cause of morbidity and mortality worldwide. The incidence of smoking in women of childbearing age population has been increasing over the years. The use of cigarettes by this class becomes even more worrisome because studies show that smoking in pregnancy is not harmful only to the mother but also to the fetus. Maternal cigarette smoking in the third trimester of pregnancy is a strong predictor of low percentile of birth weight. Thus, the reduction or complete cessation of smoking during pregnancy will result in a higher birth weight, regardless of prenatal consumption levels.

Measures of association based on risk factors and odds ratios provide information about the strength of association between the study factor and the outcome, allowing it to make an analysis on a comparison of possibilities. Evaluating the outcome of association, there were no numerical differences between the estimates of the relative risk, odds ratios

and confidence intervals obtained. It is an acceptable approximation to the relative risk and the odds ratio. It is up to the researcher to choose the most appropriate technique to his object of study and should be determined according to the surveyed data.

References

- Arango, H. G. (2009). *Bioestatística: teórica e computacional*. Rio de Janeiro, RJ: Guanabara Koogan.
- Bernstein, I. M., Mongeon, J. A., Badger, G. J., Solomon, L. S., Heil, S. H., & Higgins, S. T. (2005). Maternal smoking and its association with birth weight. *Obstetrics & Gynecology*, 106(5), 986-991.
- Difranza, J. R., & Lew, R. A. (1995). Effect of maternal cigarette smoking on pregnancy complications and sudden infant death Syndrome. *The Journal of Family Practice*, 40(4), 385-394.
- Glantz, M. (2013). La Malinche, sus padres y sus hijos. *Taurus*, 3(2), 201-203.
- Margotto, P. R. (2015). Evidência com a experiência, novo paradigma. *Revista de Medicina e Saúde de Brasília*, 4(3), 262-266.
- Oliveira, M. A. P., & Parente, R. C. M. (2010). Estudos de coorte e de caso-controle na era da medicina baseada em evidência. *Brazilian Journal*, 3(3), 115-125.
- Pagano, M., & Gauvreau, K. (2013). *Princípios de bioestatística* (2a. ed.). São Paulo, SP: Cengage Learning.
- Paranhos, R., Figueiredo Filho, D. B., Rocha, E. C., & Silva Júnior, J. A. (2013). Corra que o survey vem aí: Noções básicas para cientistas sociais. *Revista Latinoamericana de Metodología de la Investigación Social*, 3(6), 7-24.
- Veloso, H. J. F., Silva, A. A. M., Bettiol, H., Goldani, M. Z., Filho, F. L., Simões, V. M. F., ... Barbieri, M. A. (2014). Low birth weight in São Luís, northeastern Brazil: trends and associated factors. *BMC Pregnancy and Childbirth*, 14(1), 155-167.
- Zychar, B. C., Borda, C., Moreira, D., Pereira, M., & Mário, R. (2016). Envelhecimento cutâneo induzido pelo tabagismo. *Atas de Ciências da Saúde*, 3(3), 11-26.

Received on January 10, 2016.

Accepted on June 15, 2016.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.