

The effect of misclassification error on risk estimation in case-control studies

Efeito da má classificação na estimativa de risco em estudo caso-controle

Armando Baena^{I,II}, Isabel Cristina Garcés-Palacio^{III}, Hugo Grisales^{IV}

ABSTRACT: *Introduction:* In epidemiological studies, misclassification error, especially differential misclassification, has serious implications. *Objective:* To illustrate how differential misclassification error (DME) and non-differential misclassification error (NDME) occur in a case-control design and to describe the trends in DME and NDME. *Methods:* Different sensitivity levels, specificity levels, prevalence rates and odds ratios were simulated. Interaction graphics were constructed to study bias in the different settings, and the effect of the different factors on bias was described using linear models. *Results:* One hundred per cent of the biases caused by NDME were negative. DME biased the association positively more often than it did negatively (70 versus 30%), increasing or decreasing the OR estimate towards the null hypothesis. *Conclusions:* The effect of the sensitivity and specificity in classifying exposure, the prevalence of exposure in controls and true OR differed between positive and negative biases. The use of valid exposure classification instruments with high sensitivity and high specificity is recommended to mitigate this type of bias.

Keywords: Classification. Bias. Case-control studies. Odds ratios. Sensitivity and specificity. Computer Simulation.

^IInfection and Cancer Group, School of Medicine, *Universidad de Antioquia* – Medellín, Colombia.

^{II}National School of Public Health, *Universidad de Antioquia* – Medellín, Colombia.

^{III}Epidemiology Group, National School of Public Health, *Universidad de Antioquia* – Medellín, Colombia.

^{IV}Demography and Health Group, National School of Public Health, *Universidad de Antioquia* – Medellín, Colombia.

Corresponding author: Armando Baena. *Universidad de Antioquia*, Calle 70 No. 52-21, Medellín, Colombia. E-mail: arbaza@udea.edu.co

Conflict of interests: nothing to declare – **Financial support:** Estrategia de Sostenibilidad 2013-2014, *Universidad de Antioquia*.

RESUMO: *Introdução:* Em estudos epidemiológicos, o erro de classificação, especialmente o diferencial tem sérias implicações. *Objetivo:* Explicar como se expressa o erro de classificação diferencial (ECD) e não diferencial (ECND) em um estudo de caso-controle e descrever padrões de comportamento. *Métodos:* Simularam-se diferentes níveis de sensibilidade, especificidade, prevalência e *odds ratio* (OR). Construíram-se gráficos de interação para estudar o comportamento do viés nos diferentes cenários e mediante modelos lineares se descreveu o efeito dos diferentes fatores sobre esse viés. *Resultados:* O 100% dos vieses apresentados ante um ECND foram negativos, enquanto que no caso do ECD observou-se que este enviesava a associação positivamente em maior proporção que negativamente (30 versus 70%), aproximando ou afastando a estimação da OR para a hipótese nula. *Conclusões:* O efeito da sensibilidade e a especificidade na classificação da exposição, da prevalência da exposição nos controles e da OR verdadeira sobre o viés relativo difere entre os vieses negativos e positivos. O uso de instrumentos de classificação da exposição validados, com altos níveis de sensibilidade e especificidade, se recomendam para mitigar esse tipo de viés.

Palavras-chave: Classificação. Viés. Estudo de casos e controles. Razão de Chances. Sensibilidade e especificidade. Simulação por computador.

INTRODUCTION

Case-control studies are commonly utilized, especially for studying rare diseases¹⁻³. Both case identification and control selection should be carried out with the greatest possible rigor. Lack of rigor could lead to systematic errors and consequently to invalid results³. Similarly, if the exposure, the disease or both are misclassified, the association measures, the odds ratio (OR) and the conclusions will be biased¹. In any epidemiological study, the inappropriate classification of the exposure or of the disease is known as misclassification error, which can be divided into two types: non-differential misclassification error (NDME) or differential misclassification error (DME). This type of error is usually found in studies that investigate socially unacceptable behaviour or behaviours considered private that can generate shame or stigma⁴, such as sexual behaviour or the use of psychoactive substances⁵. It is known that NDME biases the risk estimate (typically represented by an odds ratio in case-control studies) towards the null hypothesis^{2,6-11}. The behavioural trends of NDME have been extensively described, but DME trends are not yet clear¹². Consider a 2 x 2 contingency table in any type of epidemiological study that represents the results observed among non-diseased and diseased individuals as a function of the exposure variable, where *a* and *b* represent, respectively, the number of diseased and non-diseased exposed individuals and *c* and *d* represent the number of diseased and non-diseased unexposed individuals, respectively. Exposure misclassification error occurs when the diseased group and the non-diseased group are considered to be exposed although they were not and unexposed

although they were^{6, 12}. A similar situation would occur if error arose in the classification of the disease. Equations 1 to 4 show how misclassification error affects the frequencies in the 2 x 2 table. These equations have been derived by us using equation 9 from the Vogel and Geffeller paper¹³.

$$a' = aSe_D + c(1 - Sp_D) \quad \text{Eq. (1)}$$

$$b' = aSe_N + c(1 - Sp_N) \quad \text{Eq. (2)}$$

$$c' = (1 - Se_D) + cSp_D \quad \text{Eq. (3)}$$

$$d' = b(1 - Se_N) + dSp_N \quad \text{Eq. (4)}$$

In these equations, a and a' indicate, respectively, the true and observed numbers of diseased persons exposed, b and b' the true and observed numbers of non-diseased exposed, c and c' the true and observed numbers of diseased unexposed, and d and d' the true and observed numbers of non-diseased unexposed. Se_D and Se_N represent the sensitivity of classifying those truly exposed among the diseased (Se_D) and non-diseased (Se_N), and Sp_D and Sp_N are the specificities for classifying those truly unexposed among the diseased (Sp_D) and non-diseased (Sp_N). The reasoning behind these equations is that the group of diseased individuals will include a percentage of truly exposed subjects (a) who will be classified as such by a certain instrument (laboratory test, survey, etc.). This percentage corresponds to the sensitivity of the instrument for classifying subjects as exposed (Se_D). In addition, some of the individuals who were truly not exposed (c) will be classified by the instrument as exposed. This percentage represents the instrument's false positive rate ($1 - Sp_D$). Therefore, among diseased individuals, the number of subjects classified by the instrument as exposed ($1 - Sp_D$) will be equal to the fraction of individuals correctly classified, aSe_D , in addition to the fraction of individuals incorrectly classified, $c(1 - Sp_D)$. Similarly, the number of diseased subjects classified by the instrument as unexposed (c') is equal to the fraction of individuals correctly classified as such, cSp_D , plus the fraction of individuals incorrectly classified, $a(1 - Se_D)$. Likewise, the number of subjects classified as exposed and unexposed by an instrument in a non-diseased population (b' and d') is determined by the sensitivity and specificity of the instrument in that population (Se_N and Sp_D). The subjects classified as exposed (b') will correspond to the fraction of individuals correctly classified, bSe_N , plus the fraction of individuals incorrectly classified $d(1 - Sp_N)$. Similarly, the subjects classified as unexposed (d') will equal the fraction of individuals correctly classified, dSp_N , added to the fraction of individuals incorrectly classified, $b(1 - Se_N)$.

Non-differential misclassification error occurs when $Se_D = Se_N$ and $Sp_D = Sp_N$, otherwise, the bias is considered a differential misclassification error¹². The true odds ratio is $OR = ad/cb$, and the observed or estimated odds ratio is $OR' = a'd'/c'b'$. If the sensitivities and specificities in non-diseased and diseased individuals are equal to 100%, the true odds ratio is equal to the

observed ($OR = OR'$), and if they are equal to 50%, the observed odds ratio will be equal to ($OR' = 1$). A similar situation would occur with other risk measures employed in different epidemiological designs. Using equations 5 to 8, it is possible to use the true table frequencies to replace the corresponding sensitivities and specificities for the following formulas derived (also by the authors) from the equations described above:

$$a = \frac{c' = (1 - Sp_D) + a' Sp_D}{(1 - Se_D)(1 + Sp_D) - Se_D Sp_D} \tag{Eq. (5)}$$

$$b = \frac{d' = (1 - Sp_N) + b' Sp_N}{(1 - Se_N)(1 - Sp_N) - Se_N Sp_N} \tag{Eq. (6)}$$

$$c = \frac{a' = (1 - Se_D)(1 - Sp_D) - c'(1 - Sp_D) Se_D}{(1 - Se_D)(1 - Sp_D)^2 - Se_D Sp_D (1 - Sp_D)} \tag{Eq. (7)}$$

$$d = \frac{b' = (1 - Se_N)(1 - Sp_N) - d'(1 - Sp_N) Se_N}{(1 - Se_N)(1 - Sp_N)^2 - Se_N Sp_N (1 - Sp_N)} \tag{Eq. (8)}$$

If the sensitivities and specificities of the instrument employed for classifying exposed/unexposed individuals (non-diseased as well as diseased) are known, then researchers can correct this type of error. However, this is not always possible. Therefore, in this article, the effect of misclassification error (differential and non-differential) on odds ratio estimates is simulated using a case-control design. This effect is simulated as a function of the sensitivity and specificity of the exposure classification given different prevalence rates of the exposure among the controls and different true ORs to describe how misclassification error occurs and to identify possible trends. These trends may guide the discussion of results in this type of studies when there are deficiencies in the classification instruments.

MATERIALS AND METHODS

OUTCOME

The OR was utilized as the measure of risk in our case-control study, and the observed value was calculated using a 2x2 table. To study the effect of misclassification error on the estimation of the OR, we used different scenarios with variations in the sensitivity and specificity of the exposure classification for cases and controls, the prevalence of the exposure among the controls and the OR. The effect was measured according to the bias produced by varying these parameters. By definition, bias is the difference between the estimated OR and the true OR¹⁴. Relative bias (expressed as a percentage) is the quotient between this difference

and the true OR (expressed in hundreds). Negative bias values indicate underestimates of the OR, and positive bias values indicate overestimates. Likewise, bias values close to zero indicate the absence of bias.

STUDY DESIGN

Simulation studies employed in various areas of research are very useful to understand the behaviour of certain phenomena under different virtual scenarios prompted by researchers through some specialized software. In statistical robustness studies these are very common to observe the behaviour of an estimator under different scenarios that could occur in reality. Given the similarity between the simulation studies and experimental studies, this last approach is used to quantify the effect of misclassification in case-control studies¹⁵. For this, we have considered six factors:

1. the sensitivity of the exposure classification among the cases;
2. the specificity of the exposure classification among the cases;
3. the sensitivity of the exposure classification among the controls;
4. the specificity of the exposure classification among the controls;
5. the prevalence of the exposure among the controls; and
6. the odds ratio.

There were five levels for factors 1 – 4 (0, 25, 50, 75 and 100%) and two levels for factors 5 – 6 (5 and 30% for the prevalence and 2 and 7 for the OR). A total of 2,500 scenarios were generated ($4^5 \times 2^2$). One hundred of these scenarios comprised the NDME analysis because the factors 1 and 3 levels as well as the factors 2 and 4 levels were the same. Therefore, the number of factors considered in this analysis was reduced to four. The remainder of scenarios ($n = 2,400$) made up the DME analysis, conserving the six initially stated factors.

STATISTICAL ANALYSIS

Interaction graphics were used to study relative bias as a function of the factors that varied, estimating the median bias for each group of combinations. Using linear models, the effect of the varying factors (i.e., the explanatory variables) on relative bias (i.e., the outcome variable) was studied, using a 0.05 level of statistical significance. Because positive relative bias occurred only with DME, which presented a high level of variability, the DME analyses were stratified according to the sign of the bias. Because of the high variability, a natural logarithmic transformation was used for the DME outcome in the linear models. Given the complexity of the DME results, in a subsequent analysis of this type of misclassification error, we considered OR estimates with absolute relative bias less than 15% as having no bias or moderate bias. This cut-off was the minimum value

close to 0% which yielded an adequate sample size to do comparisons. For this analysis, we created a categorical variable with three levels:

1. Negative bias or underestimation of the OR (relative bias values smaller than -15%);
2. Bias absent (bias values greater than -15% but less than 15%); and
3. Positive bias or overestimation of the OR (bias values greater than 15%).

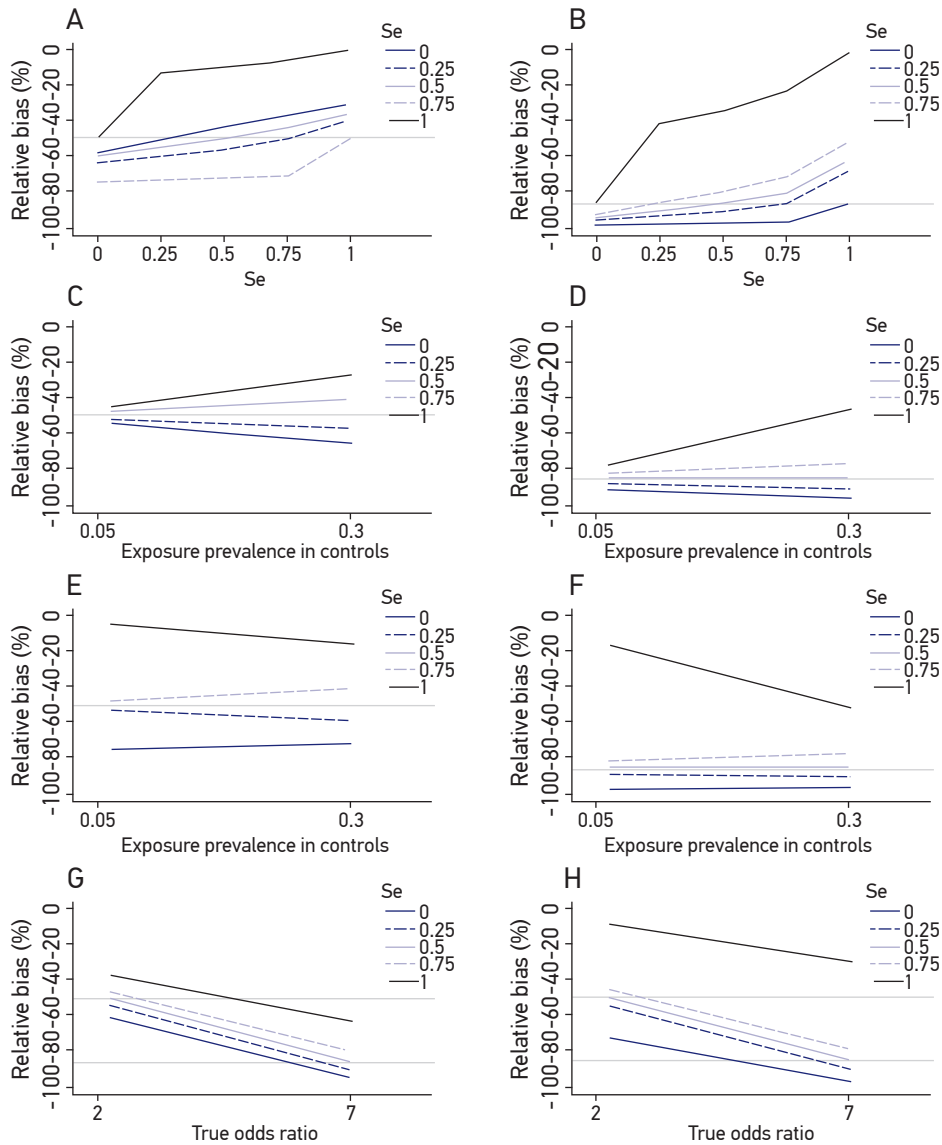
Spearman's correlation coefficient was used to assess the relationships between the different levels of the DME analysis factors and positive, absent or negative bias. P-values less than or equal to 0.05 were considered significant. The R statistical package (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>) was utilized for generating the scenarios, producing the results of the simulation and conducting the data analyses. R scripts are available on request from the corresponding author.

RESULTS

NON-DIFFERENTIAL MISCLASSIFICATION ERROR

The bias attributable to NDME was negative in the whole 100 scenarios and very negative in many cases, indicating that in the presence of NDME in a case-control study, the risk is always underestimated, even beyond the null hypothesis ($OR = 1$). In extreme cases of NDME, the OR becomes the inverse of the true odds ratio if the sensitivity and specificity of the exposure classification tend towards 0%, it approaches one if the sensitivity and specificity of the exposure classification tend towards 50% and it tends towards the true OR if the sensitivity and specificity of the exposure classification tend towards 100% (Figures 1A and 1B). On the other hand, when the specificity of the exposure classification is 100% and the sensitivity is 0%, the OR is underestimated, but never drops below the value that corresponds with the null hypothesis ($OR = 1$). If the specificity tends towards 0%, even if the sensitivity is high, the OR is underestimated below the level of the null hypothesis (Figures 1A and 1B). According to the simulations we conducted, when the specificity level is higher than the sensitivity level, an unbiased estimate of the OR is produced (p-value for the interaction between sensitivity and specificity = 0.014).

When the exposure prevalence among the controls was high and the sensitivity of the exposure classification was 100%, the OR estimate tended to be unbiased in the presence of NDME. When the sensitivity was lower, the exposure prevalence showed no effect (p-value for the interaction between sensitivity and prevalence = 0.20) (Figures 1C and 1D). The opposite effect was observed when we evaluated the relative bias by combining the specificity of the exposure classification with the prevalence of the exposure among the controls (p-value for interaction between specificity and prevalence = 0.52) (Figures 1E and 1F). Finally, we observed less bias in the estimates of moderate odds ratios. The relative bias increased when the true OR was greater, independent of the sensitivity and specificity (p-values for interaction between sensitivity *versus* OR and specificity *versus* OR = 0.88 and 0.13, respectively) (Figures 1G and 1H).



Se: sensitivity. Sp: specificity. Lines represent levels 100% (continuous), 75% (short discontinuous), 50% (dots), 25% (dot-discontinuous) and 0% (large discontinuous) of Se and/or Sp.

Relative biases below horizontal line are underestimates beyond the null hypothesis (odds ratio = 1): -50% for a true odds ratio equal to 2 (graphs A, C and E), and -86% for a true odds ratio equal to 7 (graphs B, D and F).

Figure 1. Relative bias of the risk estimation (Odds Ratio) in a case-control study under the effect of non-differential misclassification error according to different levels of sensitivity and specificity of exposure classification, exposure prevalence in controls and true odds ratio. (A, C and E) correspond to the median of relative bias of a true odds ratio = 2; and (B, D and F) of a true OR = 7.

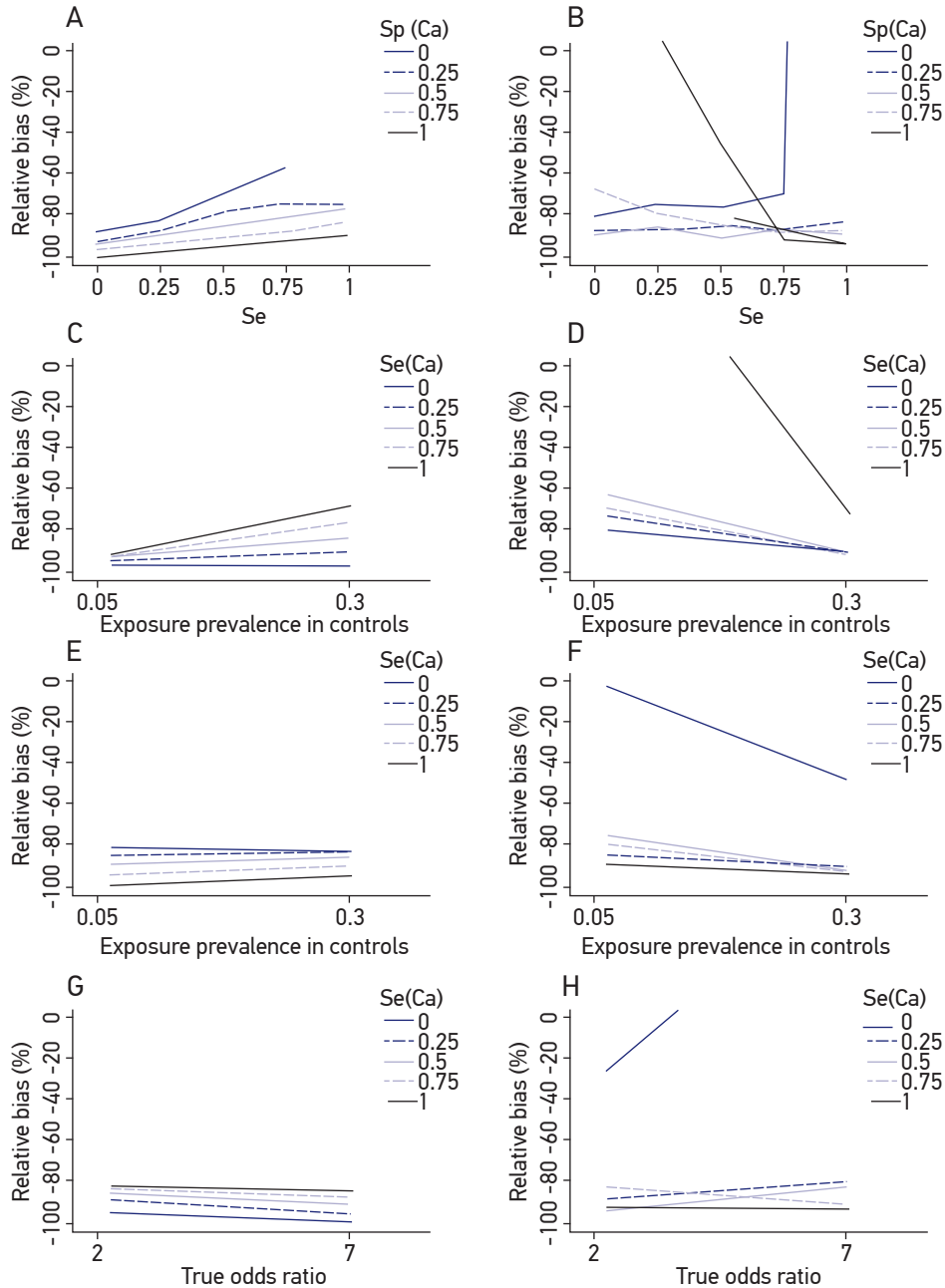
DIFFERENTIAL MISCLASSIFICATION ERROR

Of the 2,400 scenarios in the ECD analysis, 767 (32%) overestimated (i.e., positive bias) and 1,633 (68%) underestimated (i.e., negative bias) the true odds ratio. As opposed to the NDME analysis, the results of the DME are complex, not only because of the number of factors involved in the simulations but also because they do not produce a trend that clearly describes the effect of DME in relation to the combinations of factors. Therefore, the DME analysis is presented in a discriminatory manner according to the type of bias: negative or positive. Figures 2 to 4 correspond to the medians of the relative biases for different combinations of the six factors included in this analysis. High levels of sensitivity and specificity of the exposure classification among the cases (Se_{Ca} and Sp_{Ca} , respectively) reduce the bias, but they reduce positive bias more than negative bias (Figures 2A and 2B). A similar effect is observed for the sensitivity and specificity of the exposure classification among the controls (Se_{Co} and Sp_{Co} , respectively), although the effect is less clear for Se_{Co} (Figures 3A and 3B). With regard to the prevalence of the exposure among the controls, the bias is lower when the exposure prevalence is high, independent of the sensitivity or specificity of the exposure classification, both for the cases and for the controls (Figures 2C to 2F and 3C to 3F). Regarding the specificity, the effect was the opposite of that observed for NDME. When the levels of sensitivity and specificity were combined with the levels of the true ORs for the cases and the controls, no clear pattern of bias was observed (Figures 2G, 2H, 3G and 3H). After combining Se_{Ca} , Sp_{Ca} , Se_{Co} and Sp_{Co} , we observed that the bias decreased with high levels of specificity more than with sensitivity in both the cases and the controls. However, these trends were not very clear (Figures 4A to 4H).

Table 1 presents the negative, absent or positive bias associated with DME according to each factor. Negative bias percentages decreased with high levels of Se_{Ca} and Sp_{Co} (Spearman's p-values = 0.017). In the other cases, the negative bias percentage increased. The complete opposite occurred with positive bias. The absence of bias was invariably observed for varying levels of sensitivity and specificity of the exposure classification among the cases and the controls (Spearman's p-values > 0.05). A high exposure prevalence among the controls was correlated with increased percentages of absent and negative bias and with a decreased percentage of positive bias, but these were not significant (Spearman's p-values = 1). High ORs were associated with an increase in negative bias and a decrease in positive and absent bias, but these results were not significant (Spearman's p-values = 1).

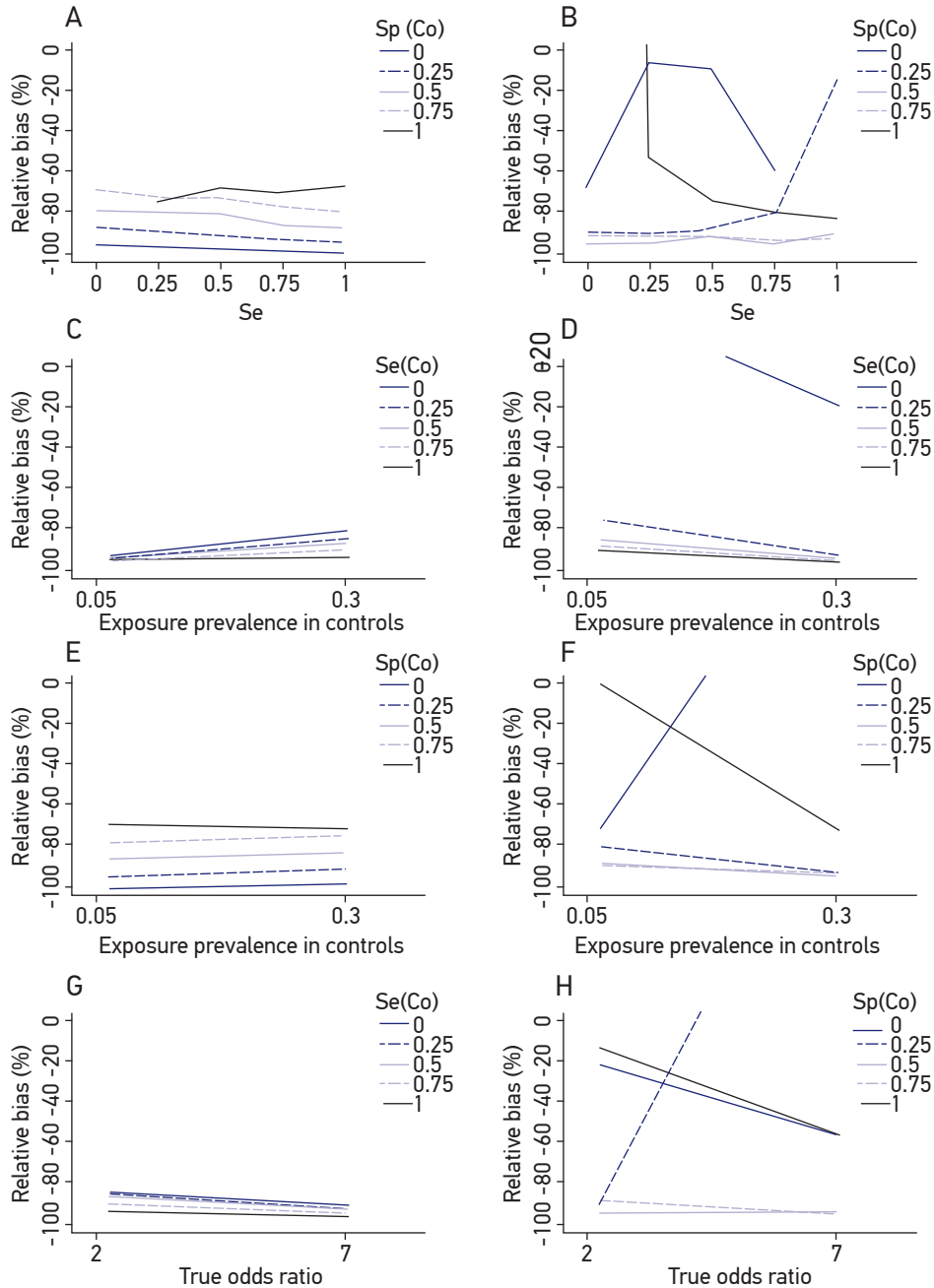
DISCUSSION

In research in general and in epidemiological research in particular, different forms of bias are frequently present and lead to findings that do not reflect precisely what is occurring. Biases are the result of deficiencies in the determination of the sample size, representativeness, data collection, definitions of relevant variables or design. The best-known biases in epidemiological research are selection, survival, loss to follow-up, detection, recall, inclusion/exclusion, confounding and misclassification¹⁶.



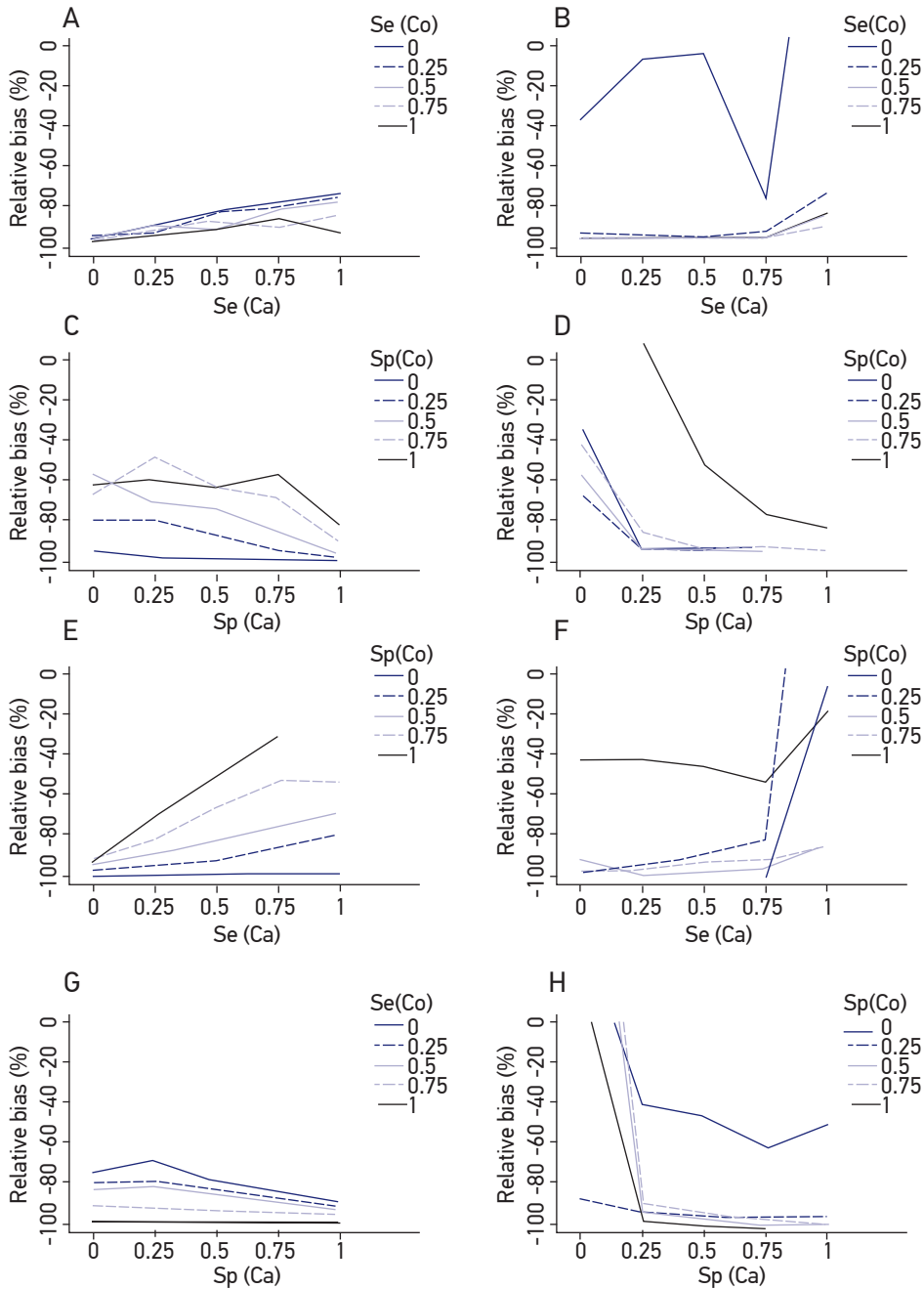
Ca: cases. Se: sensitivity. Sp: specificity. Lines represent levels 100% (continuous), 75% (short discontinuous), 50% (dots), 25% (dot-discontinuous) and 0% (large discontinuous) of Se and/or Sp.

Figure 2. Relative bias of the risk estimation (odds ratio) in a case-control study under the effect of differential misclassification error according to different levels of sensitivity and specificity of exposure classification in cases, exposure prevalence in controls and true odds ratio. (A, C and E) correspond to negative biases (B, D and F) to positive biases.



Co: controls. Se: sensitivity. Sp: specificity. Lines represent levels 100% (continuous), 75% (short discontinuous), 50% (dots), 25% (dot-discontinuous) and 0% (large discontinuous) of Se and/or Sp.

Figure 3. Relative bias of the risk estimation (Odds ratio) in a case-control study under the effect of differential misclassification error according to different levels of sensitivity and specificity of exposure classification in controls, exposure prevalence in controls and true OR. (A, C and E) negative biases; (B, D and F) positive biases.



Ca: cases. Co: controls. Se: sensitivity. Sp: specificity. Lines represent levels 100% (continuous), 75% (short discontinuous), 50% (dots), 25% (dot-discontinuous) and 0% (large discontinuous) of Se and/or Sp.

Figure 4. Relative bias of the risk estimation (Odds ratio) in a case-control study under the effect of differential misclassification error according to different levels of sensitivity and specificity of exposure classification defined for both cases and controls, exposure prevalence in controls and true odds ratio. (A, C and E) negative biases; (B, D and F) positive biases.

Table 1. Percentages of negative bias, lack of bias (absolute relative bias less than 15%) and positive bias of the estimated odds ratio in case-control studies under the effect of differential misclassification error.

Factor	Negative bias		Lack of bias		Positive bias	
	n	%	n	%	n	%
Sensitivity (Ca)						
0%	381	79.4	15	3.1	84	17.5
25%	363	75.6	13	2.7	104	21.7
50%	333	69.4	19	4.0	128	26.7
75%	294	61.3	20	4.2	166	34.6
100%	221	46.0	16	3.3	243	50.6
r/p-value	-1/0.017		0.6/0.35		1/0.017	
Specificity (Ca)						
0%	175	36.5	19	4.0	286	59.6
25%	268	55.8	28	5.8	184	38.3
50%	334	69.6	18	3.8	128	26.7
75%	383	79.8	13	2.7	84	17.5
100%	432	90.0	5	1.0	43	9.0
r/ p-value	1/0.017		-0.9/0.083		-1/0.017	
Sensitivity (Co)						
0%	273	56.9	17	3.5	190	39.6
25%	299	62.3	17	3.5	164	34.2
50%	319	66.5	18	3.8	143	29.8
75%	340	70.8	16	3.3	124	25.8
100%	361	75.2	15	3.1	104	21.7
r/ p-value	1/0.017		-0.67/0.219		-1/0.017	
Specificity (Co)						
0%	458	95.4	4	0.8	18	3.8
25%	421	87.7	7	1.5	52	10.8
50%	359	74.8	25	5.2	96	20.0
75%	259	54.0	34	7.1	187	39.0
100%	95	19.8	13	2.7	372	77.5
r/ p-value	-1/0.017		0.7/0.233		1/0.017	
Prevalence						
5%	766	63.8	23	1.9	411	34.3
30%	826	68.8	60	5.0	314	26.2
r/ p-value	1/1		1/1		-1/1	
Odds ratio						
2	692	57.7	47	3.9	461	38.4
7	900	75.0	36	3.0	264	22.0
r/ p-value	1/1		1/1		-1/1	

Ca: cases. Co: controls. r: Spearman correlation coefficient.

Misclassification bias refers to errors made when classifying an individual into a certain group, such as diseased individuals or individuals exposed to a certain factor. It occurs when non-validated instruments are used and the sensitivity and specificity of the classification method is unknown. Misclassification errors act in a particular direction, positive or negative¹⁷, underestimating or overestimating the true value of the association measure, such as the relative risk (RR) in cohort studies, the hazard ratio (HR) in longitudinal studies and the odds ratio (OR) in case-control studies. Misclassification error can be differential or non-differential. When DME occurs, the sensitivity and/or the specificity of the classification of the subjects are different among comparison groups³. NDME results in a decrease in statistical power (i.e., the ability of a test to show an association when one really exists), biasing the value of the OR towards the null^{2,6-11}. DME biases the association either towards or away from the null hypothesis¹². The majority of the time, it is impossible to predict the direction of the bias because of the complex framework involving differences in sensitivity, specificity and exposure prevalence between the cases and the controls. Although Chyou¹⁸ claims to have found patterns of the effect of DME, such patterns are described only for certain pre-established scenarios and not for a general scenario as our study intended to. In fact it is one of the limitations Chyou refers in his paper¹⁸. Of course, one aspect to highlight in the description of DME is that there is a strong influence of the prevalence of exposure as well as the estimated effect size. In practice, DME often arises when surveys are administered because it is possible that cases have better memory of the exposure than the controls, which have less motivation to remember. These surveys do not always prompt completely true responses, not because an individual intended to lie but because the individual cannot recall their exposure history³. The immediate consequence is that when the exposure classification is determined, false positives or false negatives arise¹⁹.

This simulation study analyzed the bias in an odds ratio estimate generated in the presence of DME and NDME in a case-control design for different levels of sensitivity (Se) and specificity (Sp) of the exposure classification (in the cases and the controls), the exposure prevalence among the controls and the true OR. The NDME analysis was simpler than the DME analysis because it was possible to observe easily interpretable trends. In the simulation, we found that 100% of the bias produced by the NDME in estimating the OR is negative, weakening the association and driving the estimation of the risk measure towards the null hypothesis. In addition, we found that the unbiased OR estimates in the presence of NDME are achieved faster at higher levels of specificity compared with sensitivity. This could be a consequence in the selection of strategies to classify individuals, preferring those with a greater chance of exclude the presence of the studied event (i.e. more specific). DME biases the association positively or negatively, driving the OR estimate towards or away from the null hypothesis with great uncertainty. This situation, in which the magnitude of the association is under- or overestimated, leads researchers to affirm associations that do not truly exist or, on the contrary, to affirm their non-existence when they are actually present. This could have serious implications for the generation of new knowledge to reporting opposite effects devoid of plausibility. In fact, to give an example, this could be a plausible explanation for the protective effect of cigarette use in breast cancer in certain studies¹⁹. Therefore it should be ideal that

researchers question how cigarette use has been measured and then implement better approaches such as measurements of metabolites in blood or hair, avoiding then self-report²⁰.

In this simulation, it was determined that the DME of the exposure in a case-control study produces positive bias, which is lower compared with the quantity of negative bias (30 *versus* 70%). However, the positive bias is highly variable and reaches unexpected sizes (as high as 50,000%). Another important finding of this study is that the effect of the factors (sensitivity and specificity of the exposure classification, the exposure prevalence among the controls, true OR) on relative bias is completely different for positive bias than for negative bias. There were no similarities between the two types of bias, and the bias seemed to be affected by all of the factors at the same time and not independently, which would indicate a high degree of interaction between them. These interactions are difficult to interpret using graphics because they only describe two-order interactions. However, the linear models revealed second-, third-, fourth-, and fifth-order significant interactions (data not shown). Of the 57 possible interactions in the DME analysis, 22 (39%) were significant for negative bias and 27 (47%) for positive bias, showing the degree of complexity in this analysis in comparison to the NDME analysis, which only revealed two significant interactions among the 11 possible interactions (18%).

Our analyses were based on the variation of the sensitivity and specificity of the classification of an event of interest. However, we recognize that there are other approaches which additionally include the predictive value of the classification methods such as the one Marshall²¹ proposed. Even so, we believe that the approach presented by us is useful for understanding the effects of the limitations in terms of sensitivity and specificity and even more useful because researchers could easily find explanations for the lack of associations or implausible findings through systematic reviews about the limitations of the instruments used by them. Although knowledge of prior information such as sensitivity and specificity of the exposure classification could allow researchers to detect misclassification bias²² and correct it^{23,24}, we insist that is very useful to be aware about the effects of misclassification and how these could be avoided if researchers validate the methods of classification. As with any other type bias, it will always be best avoid bias rather than having to correct it.

CONCLUSION

It is likely that a misclassification error will occur if the sensitivity and specificity of any instruments used to compare groups are not exactly 100%. If this type of error occurs, it is preferable if it is non-differential. In cases of NDME, there is equilibrium in the bias, and the explanation of the study results is easier. In a case-control study, it is important to pay close attention not only to the identification of cases and to the appropriate selection of controls but also to the classification of the exposure. Thus, the use of validated classification instruments is emphasized so that instruments with high sensitivity and high specificity are selected. The design of the instrument and the appropriate formulation of the questions on a questionnaire are key components for avoiding DME. It is therefore important that

epidemiologists and researchers are close not only to the elaboration and design of surveys, but also to the monitoring of the quality with which these are applied as well as the quality control of other classification procedures and tests laboratory and/or diagnostic tests.

SOFTWARE

Scripts in the form of R code are available on request from the corresponding author.

ACKNOWLEDGMENTS

This work was supported by Estrategia de Sostenibilidad 2013-2014, Universidad de Antioquia. AB is a recipient of a doctoral fellowship grant from Colombian Institute for Scientific and Technological Development (COLCIENCIAS 2010).

REFERENCES

1. Kleinbaum D, Kupper L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods* New York: Van Nostrand Reinhold Company, 1982.
2. Rothman K, Greenland S. *Modern Epidemiology*. 3th ed. Philadelphia: Lippincott Raven Publishers, 2012.
3. dos-Santos I. *Cancer Epidemiology: Principles and Methods*. Lyon, Francia: IARC Library Cataloguing in Publication Data, 1999.
4. Sanders SA, Hill BJ, Yarber WL, Graham CA, Crosby RA, Milhausen RR. Misclassification bias: diversity in conceptualisations about having 'had sex'. *Sex Health* 2010; 7: 31-4.
5. Rikala M, Hartikainen S, Saastamoinen LK, Korhonen MJ. Measuring psychotropic drug exposures in register-based studies - validity of a dosage assumption of one unit per day in older Finns. *Int J Methods Psychiatr Res* 2013; 22(2): 155-65.
6. Bross I. Misclassification in 2x2 tables. *Biometrics*. 1954; 10: 478-86.
7. Newell D. Errors in the interpretation of errors in epidemiology. *Am J Public Health Nations Health* 1962; 52: 1925-8.
8. Keys A, Kihlberg J. Effect of misclassification on estimated relative prevalence of a characteristic: Part I. Two populations infallibly distinguished: Part II. Errors in two variables. *Am J Public Health Nations Health* 1963; 53: 1656-65.
9. Gullen WH, Bearman JE, Johnson EA. Effects of misclassification in epidemiologic studies. *Public Health Rep* 1968; 83: 914-8.
10. Goldberg J. Effects of misclassification on bias in difference between 2 proportions and relative odds in fourfold table. *J Am Stat Assoc* 1975; 70: 7.
11. Weinberg C, Umbach D, Greenland S. When will nondifferential misclassification of an exposure preserve the direction of a trend. *Am J Epidemiol* 1994; 140(6): 565-71.
12. Szklo M, Nieto F. *Epidemiología intermedia. Conceptos y aplicaciones*. España: Díaz de Santos, 2003.
13. Vogel C, Gefeller O. Implications of nondifferential misclassification on estimates of attributable risk. *Methods Inf Med* 2002; 41: 342-8.
14. Daniel W, Cross C. *Biostatistics: A Foundation for Analysis in the Health Sciences*. USA: Wiley, 2013.
15. Salazar JC, Baena A. *Análisis y diseño de experimentos aplicados a estudios de simulación*. Dyna revfacnacminas. 2009; 159: 249-57.
16. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med* 1981; 304(11): 630-3.
17. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. *Int J Epidemiol* 2005; 34: 1370-6.

18. Chyou PH. Patterns of bias due to differential misclassification by case-control status in a case-control study. *Eur J Epidemiol* 2007; 22: 7-17.
19. Hrubá D. [Smoking and breast cancer]. *Klin Onkol.* 2013; 26: 389-93.
20. Joya X, Manzano C, Álvarez AT, Mercadal M, Torres F, Salat-Batlle J, et al. Transgenerational exposure to environmental tobacco smoke. *Int J Environ Res Public Health* 2014; 11: 7261-74.
21. Marshall RJ. Misclassification of exposure in case-control studies: assessment by quality indices. *Epidemiology* 1994; 5: 309-14.
22. Marshall RJ. An empirical investigation of exposure measurement bias and its components in case-control studies. *J Clin Epidemiol* 1999; 52(6): 547-50.
23. Gustafson P, Le ND, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics* 2001; 57: 598-609.
24. Marshall RJ. Assessment of exposure misclassification bias in case-control studies using validation data. *J Clin Epidemiol* 1997; 50(1): 15-9.

Received on: 04/28/2014

Final version presented on: 09/10/2014

Accepted on: 10/08/2014