



Revista de Saúde Pública

ISSN: 0034-8910

revsp@usp.br

Universidade de São Paulo
Brasil

Borges Cabral, Maria Deolinda; Raggio Luiz, Ronir
Sensitivity analysis for unmeasured confounders using an electronic spreadsheet
Revista de Saúde Pública, vol. 41, núm. 3, junio, 2007, pp. 446-452
Universidade de São Paulo
São Paulo, Brasil

Available in: <http://www.redalyc.org/articulo.oa?id=67240160017>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System
Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal
Non-profit academic project, developed under the open access initiative

Maria Deolinda Borges Cabral^IRonir Raggio Luiz^{II}

Sensitivity analysis for unmeasured confounders using an electronic spreadsheet

Análise de sensibilidade para um confundidor não mensurado utilizando planilha eletrônica

ABSTRACT

In studies assessing the effects of a given exposure variable and a specific outcome of interest, confusion may arise from the mistaken impression that the exposure variable is producing the outcome of interest, when in fact the observed effect is due to an existing confounder. However, quantitative techniques are rarely used to determine the potential influence of unmeasured confounders. Sensitivity analysis is a statistical technique that allows to quantitatively measuring the impact of an unmeasured confounding variable on the association of interest that is being assessed. The purpose of this study was to make it feasible to apply two sensitivity analysis methods available in the literature, developed by Rosenbaum and Greenland, using an electronic spreadsheet. Thus, it can be easier for researchers to include this quantitative tool in the set of procedures that have been commonly used in the stage of result validation.

KEYWORDS: Statistical interpretation of data. Sensitivity analysis. Confounding (Epidemiology). Observational studies. Electronic spreadsheet.

RESUMO

Em estudos que avaliam o efeito de uma dada variável de exposição e um determinado desfecho de interesse, uma situação de confusão pode ser caracterizada pela falsa aparência de que a variável de exposição produz o desfecho de interesse quando, de fato, o efeito observado se deve a um fator de confundimento que está presente. Entretanto, é pouco freqüente a utilização de técnicas quantitativas para determinar a influência potencial de confundidores não mensurados. A análise de sensibilidade é uma técnica estatística que permite uma medida quantitativa do impacto de uma variável confundidora não mensurada na associação de interesse que está sendo avaliada. O objetivo do artigo foi viabilizar a aplicação, por meio de planilha eletrônica, de dois métodos de análise de sensibilidade disponíveis na literatura, desenvolvidos por Rosenbaum e Greenland. Dessa forma, é possível facilitar ao pesquisador a incorporação desse ferramental quantitativo ao conjunto de procedimentos que já são comuns na etapa de validação dos resultados.

DESCRIPTORES: Interpretação estatística de dados. Análise de sensibilidade. Fatores de confusão (Epidemiologia). Estudos observacionais.

^I Instituto Brasileiro de Geografia e Estatística. Rio de Janeiro, RJ, Brasil

^{II} Núcleo de Estudos de Saúde Coletiva. Faculdade de Medicina da Universidade Federal do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

Correspondence:

Maria Deolinda Borges Cabral
Av. República do Chile, 500 50 andar
20031-170 Rio de Janeiro, RJ, Brasil
E-mail: maria.cabral@ibge.gov.br

INTRODUCTION

There is much epidemiological interest in establishing causes and relationships. While science is concerned with the frequency, distribution and determination of disease factors, methodological procedures have been developed based on statistical models to identify causes of diseases.^{3,4,7} However, these models rely on assumptions that frequently cannot be tested through the observed data, that is the discussion of causality addresses the assessment of the validity of the findings obtained in the studies.

A study is considered valid, with a resulting causal interpretation, if it is bias-free, i.e., there are no systematic errors that explain the association found as an alternative to the causal hypothesis.⁴

In studies assessing the effects of a given exposure variable and a specific outcome of interest, confusion may result from the mistaken impression that the exposure variable produces the outcome of interest when the effect observed is actually due to an existing confounding factor. According to Koopman* (1997), confounding occurs when a non-causal association is observed between the exposure and the outcome of interest in a reference population. Two types of biases resulting from confounding may arise: overt bias, caused by confounders that are measured in the study, and hidden bias, caused by existing unmeasured confounders in the study⁵ (1991).

When analyzing observational studies, the measured potential confounders are usually analytically “adjusted” using statistical techniques such as stratification, pairing, among others. However, quantitative techniques are rarely used to determine the potential impact of unmeasured confounders. According to Greenland² (1996), the random errors and confounders measured in the data generation process often constitute only a fraction of the total error, and are rarely the only important sources of uncertainty. It is thus convenient to develop and use an appropriate statistical tool that allows a quantitative evaluation of such errors, with the sensitivity analysis being a statistical technique that allows the quantitative measurement of the impact of an unmeasured confounding variable on the association of interest that is being assessed.

Although conceptually well-developed, the two sensitivity analysis methods available in the literature developed by Rosenbaum⁶ (1995) and Greenland² (1996) require laborious calculations not handled by currently available software programs. However, such methods may be fully applied through an electronic spreadsheet. The purpose of the present study is to

make it feasible to apply each of these methods using an electronic spreadsheet in order to make it easier for researchers to include this quantitative tool in the set of procedures that have been commonly used in the stage of result validation. The selection of a spreadsheet is prompted by its widespread use.

SENSITIVITY ANALYSIS METHODS

Rosenbaum⁵ and Greenland² developed two sensitivity analysis methods applied to dichotomic variables that allow analyses of the behavior of study results in the event of unmeasured confounders.

Also known as the external adjustment method, the Greenland method tries to quantify the variation in the association observed in a specific study when adjusted for a potential unmeasured confounding variable. The method consists of simulating various plausible values for the confounder prevalences by exposure level, specifically in those individuals who do not show the outcome, as well as the magnitude of association between the confounder and the outcome, then calculating an estimate of the association between the exposure and the outcome “adjusted” for the specified confounding variable for each combination studied.

In contrast to the Greenland method, which considers the classic confounding scheme (i.e., the confounder must be associated with the exposure and be an independent predictor of the outcome), the Rosenbaum method works only with the association between the confounder and the exposure. This method quantifies the magnitude of the association between the unmeasured confounder and the exposure variable required to make the association statistically non-significant. It is found between the exposure and the outcome, assuming that the gap between the confounder and the outcome is enough for the confounding to affect the association between this confounder and the exposure variable.

ELEMENTS AND NOTATIONS FOR THE APPLICATION OF A SENSITIVITY ANALYSIS

To formalize Greenland² and Rosenbaum⁶ methods, a hypothetical study is considered where the exposure, outcome and unmeasured confounder variables are defined as follows:

E: exposure variable $\rightarrow \begin{cases} E = 1, & \text{if the individual is exposed} \\ E = 0, & \text{if the individual is not exposed} \end{cases}$

* Koopman NJ. Stratification of exposure-disease relationships upon a third variable and the assessment of joint effects [monografia na internet]. Ann Arbor; 1997. Available at: <http://www.sph.umich.edu/group/epid/> [Access on 17 May 2006]

Table 1. General scheme of the frequencies observed

Outcome	Exposure		Total
	E = 1	E = 0	
D=1	A ₁₊	A ₀₊	R
D=0	B ₁₊	B ₀₊	N-R
Total	M	N-M	N

for exposure and outcome (1 = present; 0 = not present).

D: outcome variable → $\begin{cases} D = 1, \text{ if the individual shows the outcome} \\ D = 0, \text{ if the individual does not show} \\ \text{the outcome} \end{cases}$

E: unmeasured confounder variable → $\begin{cases} Z = 1, \text{ if the individual presents} \\ \text{the confounder} \\ Z = 0, \text{ if the individual does not present} \\ \text{the confounder} \end{cases}$

Table 1 shows the general scheme for presenting the findings obtained in this hypothetical study.

In order to apply the Greenland method, Table 1 should be “stratified” by the unmeasured confounding variable Z, according to the scheme presented in Table 2.

The following magnitudes are now considered:

P_{Z1} : prevalence of the unmeasured confounding variable among exposed individuals;

P_{Z0} : prevalence of the unmeasured confounding variable among non-exposed individuals;

OR_{DE} : odds ratio between the outcome and the exposure;

OR_{DZ} : odds ratio between the outcome and the confounding variable;

OR_{EZ} : odds ratio between the exposure and the confounding variable.

The Greenland method speculates on the plausible values for OR_{DZ} , P_{Z1} and P_{Z0} , and, consequently, it speculates on the possible values for the association between E and Z (OR_{EZ}), because OR_{EZ} is affected by the values of P_{Z1} and P_{Z0} , according to the following formula (1).

$$OR_{EZ} = \frac{P_{Z1} / (1 - P_{Z1})}{P_{Z0} / (1 - P_{Z0})} = \frac{P_{Z1}(1 - P_{Z0})}{P_{Z0}(1 - P_{Z1})} \tag{1}$$

In order to find the values to complete Table 2, the hypothesis formulated is that the odds ratio between E and Z has the same value for both Z strata (Z is the confounding variable for the association between E and D). Thus, speculating about the plausible values for these three (or four) magnitudes, various OR_{DE} values are obtained and “adjusted” for Z, allowing an analysis of existing variations considered epidemiologically relevant that may point out findings other than those obtained.

On the other hand, the Rosenbaum method^{5,6} speculates on the Γ value, the magnitude associating the unmeasured confounder to the exposure which makes the observed association of interest OR_{DE} statistically non-significant. For dichotomic variables, the method is based on the Mantel-Haenszel statistic (T). This is a test statistic normally used in analyses where a third variable is taken into consideration that may “mask” the association found between the exposure and the outcome of interest¹ (1981). It considers the total number of exposed individuals showing the outcome ($T = A$ in the hypothetical case presented in Table 1). The calculation of the expectation and the T variance is carried out on an approximate basis by the normal distribution, establishing the values for the marginal patient totals and the exposed individuals as R and M in Table 1. The expectation expression is given by a second level equation on the null hypothesis that the exposure is not associated with the outcome, obtained by the odds ratio between the exposure and the outcome when equal to the speculated value of the association between the exposure and the unmeasured confounder (Γ). The variance calculation considers the expectation value and the A_{1+} , R , M , N and Γ values. Once the expectation and variance values are obtained, the standardized statistical T value (T_{std}) is calculated and the p-value is obtained for the upper limit. For calculating the lower limit p-value, Γ is replaced by $\frac{1}{\Gamma}$ in the odds ratio equation between the exposure and the outcome, and the expectation and variance calculations are reworked. The value sought by this method is the lowest value for Γ , which makes the observed association of interest (OR_{DE}) statistically non-significant at a 95% confidence level. The formulas for calculating the expectation, variance and standardized T were developed by Stevens⁸ (1951).

Table 2. General scheme (expected data) for the sensitivity analysis (Greenland) of an unmeasured dichotomic variable Z.

Outcome	Unmeasured variable					
	Z=1			Z=0		
	E=1	E=0	Total	E=1	E=0	Total
D=1	A ₁₁	A ₀₁	R ₁	A ₁₊ - A ₁₁	A ₀₊ - A ₀₁	R - R ₁
D=0	B ₁₁	B ₀₁	N ₁ - R ₁	B ₁₊ - B ₁₁	B ₀₊ - B ₀₁	(N - R) - (N ₁ - R ₁)

Table 3. Description of the spreadsheet cells when using the Greenland method presented in Figure 1.

Cell	Content
C6	Value observed for the total number of exposed individuals showing the outcome
C7	Value observed for the total number of exposed individuals not showing the outcome
E6	Value observed for the total number of non-exposed individuals showing the outcome
E7	Value observed for the total number of non-exposed individuals not showing the outcome
B12	Speculated value of the odds ratio between the outcome and the confounder
B13	Proportion of the confounder Z among exposed individuals
B22	Speculated value of the odds ratio between the exposure and the confounder
B9	$=(C6 * E7) / (E6 * C7)$
C19	$=(B12 * C6 * C20) / (B12 * C20 + C7 - C20)$
C20	$=B13 * C7$
E19	$=(B12 * E6 * E20) / (B12 * E20 + E7 - E20)$
E20	$=B14 * E7$
I19	$=C6 - C19$
I20	$=C7 - C20$
K19	$=E6 - E19$
K20	$=E7 - E20$
B14	$=B13 / (B22 + B13 - (B22 * B13))$
B22	$=(B13 * (1 - B14)) / (B14 * (1 - B13))$
B23	$=(C19 * E20) / (E19 * C20)$
B24	$=(I19 * K20) / (K19 * I20)$

SENSITIVITY ANALYSIS PERFORMANCE SPREADSHEETS

To make available the two sensitivity analysis methods under consideration, two spreadsheets were developed that allow the calculations to be carried out as required for their application.

Figure 1 shows the spreadsheet for applying the Greenland method. All cells in this spreadsheet should be completed as described in Table 3. The cells C6, C7, E6, and E7 must be filled out with the data observed in the study and the magnitudes to be speculated should be entered into cells B12, B13 and B14. Once completed as described in Table 1, all findings will be automatically generated by the spreadsheet. Cells B23 and B24 show the odds ratio between the exposure variable and the outcome, between individuals exposed and not exposed to Z, respectively, adjusted for the speculated values in cells B12, B13 and B14. These values are identical as Z is considered a confounding variable for E and D in the method development. Cell B22 provides the

Table 4. Description of spreadsheet cells using the Rosenbaum method, showed in Figure 2.

Cell	Content
C6	Total exposed individuals showing the outcome
C7	Total exposed individuals not showing the outcome
E6	Total non-exposed individuals showing the outcome
E7	Total non-exposed individuals not showing the outcome
C8	$=C6 + C7$
E8	$=E6 + E7$
G6	$=E6 + C6$
G7	$=E7 + C7$
G8	$=G6 + G7$
H6	$=(C6 * E7) / (E6 * C7)$
B12	1
C12	0
D12	$= -G8$
E12	$=B12 * G6 * C8$
B16	$=(B12 * G6 * C8) / G8$
C16	$=1 / ((1 / B16) + (1 / (G6 - B16)) + (1 / (C8 - B16)) + (1 / (G8 - G6 - C8 + B16)))$
G16	$=(ABS(C6 - B16) - 0,5) / \text{ROOT}(C16)$
H16	$=1 - \text{DIST.NORMP}(G16)$
B20	Γ Value to be considered
C20	$=B20 - 1$
D20	$=(B20 - 1) * (C8 + G6) + G8$
E20	$=B20 * G6 * C8$
B24	$=\text{ROOT}(((D20^2) - (4 * C20 * E20)))$
C24	$=(D20 + B24) / (2 * C20)$
D24	$=(D20 - B24) / (2 * C20)$
E24	$=\text{MAXIMUM}(0; (C8 + G6 - G8))$
F24	$=\text{MINIMUM}(G6; C8)$
G24	$=SE(E(C24 > E24; C24 < F24); C24; D24)$
H24	$=1 / ((1 / G24) + (1 / (G6 - G24)) + (1 / (C8 - G24)) + (1 / (G8 - G6 - C8 + G24)))$
C27	$=(ABS(C6 - G24) - 0,5) / \text{ROOT}(H24)$
D27	$=1 - \text{DIST.NORMP}(C27)$

Table 5. Hypothetical data on exposure to factor E and outcome D (1=presence; 0=absence)

Outcome	Exposure		
	E=1	E=0	Total
D=1	63	92	155
D=0	248	930	1,178
Total	311	1,022	1,333

$$OR_{DE} = 2.57$$

	A	B	C	D	E	F	G	H	I	J	K
1	SENSITIVITY ANALYSIS - GREENLAND METHOD (1996)										
2											
3	OBSERVED DATA										
4											
5	Exposed (E=1) Non-exposed (E=0)										
6	Individuals show the outcome (D=1)	A ₁₊ =	63	A ₀₊ =	92						
7	Individuals do not show the outcome (D=0)	B ₁₊ =	248	B ₀₊ =	930						
8											
9	Observed OR _{DE} = 2.57										
10											
11	"SPECULATED" VALUES										
12	OR _{DZ} = 3.00		= odds ratio between the outcome (D) and the unmeasured confounder (Z)								
13	P _{Z1} = 0.90		= prevalence of the unmeasured confounder among exposed individuals (E=1)								
14	P _{Z0} = 0.83		= prevalence of the unmeasured confounder among non-exposed individuals (E=0)								
15											
16	STRATIFIED DATA BY A DICHOTOMOUS UNMEASURED CONFOUNDER Z										
17											
18	FOR Z=1	Exposed (E=1) Non-exposed (E=0)				FOR Z=0	Exposed (E=1)		Non-exposed (E=0)		
19	Individuals show the outcome (D=1)	A ₁₁ =	61	A ₀₁ =	86	Individuals show the outcome (D=1)	A ₁₊ - A ₁₁ =	2	A ₀₊ - A ₀₁ =	6	
20	Individuals do not show the outcome (D=0)	B ₁₁ =	223	B ₀₁ =	768	Individuals do not show the outcome (D=0)	B ₁₊ - B ₁₁ =	25	B ₀₊ - B ₀₁ =	162	
21											
22	OR _{EZ} = 1.90		= odds ratio between the exposure (E) and the unmeasured confounder (Z)								
23	Adjusted OR _{DE} = 2.43		= odds ratio between the outcome (D) and the exposure (E) for Z=1								
24	Adjusted OR _{DE} = 2.16		= odds ratio between the outcome (D) and the exposure (E) for Z=0								

Figure 1. Excel spreadsheet for the application of a sensitivity analysis using the Greenland method.

value of the association between the confounder and the exposure variable (OR_{EZ}).

In turn, the spreadsheet in Figure 2 (Rosenbaum method) should be completed as described in Table 4. When completing the specified cells, the odds ratio value observed between the exposure and the outcome (OR_{DZ}) is calculated automatically in cell H6; T expectation values, T variance, T_{std} statistic and p-value for $\Gamma = 1$ are automatically calculated in cells B16, C16, G16 and H16, respectively. Cell B20 should be completed with the values to be speculated for Γ , when $\Gamma \neq 1$. When completed, T expectation value, T variance, T_{std} statistic

and p-value of the upper limit are automatically calculated in cells G24, H24, C27 and D27, respectively. The expectation values and variance required to calculate the p-value for the lower limit are obtained in the same way, replacing Γ value by $\frac{1}{\Gamma}$ in cell B20.

If there are two or more strata, the expectations and variances for each stratum should be calculated for each Γ value considered. The T statistic will be given by the sum of the exposed individuals showing the outcome for all the strata and, as the T expectation and T variance, the sum of the expectations and variances for all

	A	B	C	D	E	F	G	H
1	SENSITIVITY ANALYSIS - ROSENBAUM METHOD (1995)							
3	OBSERVED DATA							
5		Exposed (E=1)		Non-exposed (E=0)		Total		OR_{DE}
6	Individuals show the outcome (D=1)	A ₁₊ =	63	A ₀₊ =	92	R =	155	2.57
7	Individuals does not show the outcome (D=0)	B ₁₊ =	248	B ₀₊ =	930	N - R =	1,178	
8	Total	M =	311	N - M =	1,022	N =	1,333	
10	$\Gamma = 1$							
11		Γ	a	b	c			
12	First level equation coefficients to calculate T expectation, considering $\Gamma = 1$	1	0	-1,333	48,205			
15	Expectation and T variance calculation	EXP	VAR	Standardized statistical T calculation		T_{pad}	p-value	
16		36.16	24.50			5.32	0.000000	
18	$\Gamma \neq 1$							
19	Second level equation coefficients to calculate T expectation, considering the Γ value	Γ	a	b	c			
20		2	0.9	1752.4	91589.5			
23	Expectation and T variance calculation	RDELTA	R1	R2	MAX	MIN	EXP	VAR
24		1655.65	1893.36	53.75	0	155	53.75	29.89
26	Standardized statistical T and p-value calculation (95% confidence level)	T_{pad}	p-value					
27		1.60064	0.054729					

Figure 2. Excel spreadsheet for the application of a sensitivity analysis using the Rosenbaum method.

the strata respectively. After obtaining T statistic, expectation and T variance values, T_{std} values and p-value are calculated for the upper limit, similar to the formulas described in cells C27 and D27 respectively. Once again, for calculating the p-value of the lower limit, the calculations are repeated, replacing Γ by $\frac{1}{\Gamma}$.

EXAMPLE OF A SENSITIVITY ANALYSIS APPLICATION

As an example of the use of the spreadsheets presented, a hypothetical observational study is considered, analyzing the association between the exposure to a factor

E and an outcome of interest D, whose findings are presented in Table 5.

In order to verify the behavior of the association found in the presence of a potential unmeasured confounder (Z), it was decided to apply a sensitivity analysis to the observed data. Due to the importance of the two methods available, it is suggested that they be applied in an integrated manner* (2005). Initially, the Rosenbaum method was applied in order to obtain the Γ value making the OR_{DE} value adjusted for the unmeasured confounding variable statistically non-significant. The spreadsheet showed in Figure 2 was used for Γ values equal to 1.0, 1.5, 1.8, 1.9, 2.0 and 3.0, and also for the

* Cabral MDB. Análise de sensibilidade em estudos epidemiológicos [dissertação de mestrado]. Rio de Janeiro: Núcleo de Estudos de Saúde Coletiva da UFRJ; 2005.

corresponding $\frac{1}{\Gamma}$ values, and the findings are showed in Table 6.

According to Table 6, the lowest Γ value making the adjusted OR_{DE} value statistically non-significant at a 95% significance level is $\Gamma = 1.9$. The suggestion is to start the Greenland method by taking the minimum value of Γ in the Rosenbaum method as the initial value for speculating on the value of the association between the exposure variable and the unmeasured confounder Z (OR_{EZ}). Thus, using the spreadsheet in Figure 1 for the OR_{ZE} values set at 1.9, 2.5, and 3.0, the OR_{DZ} values "speculated" at 3.0, 5.0, 10.0 and 15.0, with P_{Z1} values varying between 0.1 and 0.9; and the corresponding P_{Z0} values obtained through formula (2).

$$P_{Z0} = \frac{P_{Z1}}{OR_{EZ} + P_{Z1} - (OR_{EZ} * P_{Z1})} \quad (2)$$

Analyzing the findings showed in Table 7, based on a hypothetical study, it can be noted that the variations in the adjusted OR_{DE} values move away from the observed OR_{DE} value (2.57) when the unmeasured confounder increases the chance of exposure by 2.5 and also presenting an odds ratio with the outcome of at least 10.

FINAL CONSIDERATIONS

The two spreadsheets presented in Figures 1 and 2 are intended to provide an operating tool that streamlines

the application of a sensitivity analysis by researchers, allowing quantitative measurements of the impact of an unmeasured confounding variable on the association of interest that is being assessed.

The spreadsheets provided are easy to use and allow the immediate application of the Greenland² (1996) and Rosenbaum⁵ (1995) methods. The Greenland method approach focuses more on the epidemiological elements of the study, while the Rosenbaum method addresses the statistical significance of the findings observed.

As these two approaches are important for observational studies, the example presented suggests a way of integrating these two techniques in order to direct and reduce the number of calculations required for a sensitivity analysis. The calculations presented for these two methods address the exposure, outcome and dichotomic confounder variables.

It should be stressed that with the Greenland method, should it prove necessary to stratify for a measured confounder, the calculations in the method description should be repeated for each stratum, and the findings obtained for each of them should then be merged. Moreover, the spreadsheet provided to apply the Rosenbaum method may be used when the marginal totals for each stratum are large, i.e., when M , $N - M$, R and $N - R$ are large. Otherwise, other expressions for the exact expectation and variance of the T distribution should be used, which may be found in Rosenbaum⁶ (1995).

REFERENCES

1. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons; 1981.
2. Greenland S. Basic Methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25(6):1107-16.
3. Holland PW. Statistics and causal inference. *J. Am. Stat. Assoc.* 1986; 81(396):945-60.
4. Luiz RR, Struchiner CJ. Inferência Causal em Epidemiologia: o modelo de respostas potenciais. Rio de Janeiro: Editora Fiocruz; 2002.
5. Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med*. 1991;115(11):901-5.
6. Rosenbaum PR. Observational Studies. New York: Springer-Verlag; 1995.
7. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688-701.
8. Stevens WL. Mean and variance of an entry in a contingency table. *Biometrika*. 1951;38(3-4):468-470.