
Use of sensitivity analysis to assess the effects on anti-hepatitis A virus antibodies of access to household water supply

M. D. B. CABRAL¹* AND R. R. LUIZ²

¹ *Brazilian Institute of Geography and Statistics – IBGE, Rio de Janeiro, Brazil*

² *Institute of Public Health Studies, UFRJ, Rio de Janeiro, Brazil*

(Accepted 21 February 2007; first published online 20 April 2007)

SUMMARY

The objective of this study was to assess the impact of a possible unmeasured confounding variable in a previously published association between the effects of household water supply and positive results for hepatitis A serology. This was estimated using a path of integration between two methods of sensitivity analysis, called Rosenbaum's method and Greenland's external adjustment. The association between household water supply and positive results for hepatitis A (outcome) serology was insensitive to confounding unless the odds ratio for the association between the confounder and the outcome was ≥ 4 . The integration of the two sensitivity analysis methods presented proved useful when assessing the effects of a potential unmeasured confounder.

INTRODUCTION

The causal association between water quality and the occurrence of waterborne diseases [1–3] is well known and has long been demonstrated. Obviously, the consumption of pathogen-free water implies a reduction in the incidence of waterborne diseases, such as hepatitis A. However, a large proportion of water-communicable diseases, in populations in which these diseases are endemic, is due to the limited amount of water available for domestic consumption and personal hygiene [4, 5]. The potential effects of limited water consumption are less well-known, curtailing personal hygiene even in situations with good-quality drinking water [4]. Seeking evidence of the potentially deleterious effects of the lack of adequate household water supplies, Luiz *et al.* [6] concluded that limited water consumption is an important factor for the

occurrence of hepatitis A, particularly when there is no exposure to a sanitary landfill or an open sewer close to the home.

However, as this is an observational study, despite this association being controlled by a set of analysed covariables, a potential unmeasured confounder (hidden bias) may also account for these findings [7]. For example, individuals who do not have water taps in their homes may have to travel to areas where hepatitis A is endemic far more frequently than individuals who do have water taps in their residence. In this case, the individuals classified as exposed would have more chance of being infected than those who are not exposed, and consequently a higher prevalence of hepatitis A in the group of exposed individuals might reflect only this difference. This variable may be considered as a third non-observed variable that might potentially affect the findings.

According to Greenland [8], most statistical methods used in epidemiological surveys focus on the assessment of random errors and confounders measured

* Author for correspondence: Dr M. D. B. Cabral, IBGE, Rua Luis Vahia Monteiro, 202/102 Jardim Guanabara Rio de Janeiro, Rio de Janeiro 21931-360, Brazil.
(Email: maria.cabral@ibge.gov.br)

during the data generation process, which are frequently only a fraction of the total error, and rarely the sole important source of uncertainty when estimating a causal effect measurement.

Sensitivity analysis is a statistical technique that allows the quantification of the impact of an unmeasured confounder in an association of interest observed in a specific study. Instead of saying that the association found may not imply causation through the possible presence of an unmeasured confounder in the process, the magnitude of the bias may be assessed, as required, to alter or eliminate the observed association.

Cornfield *et al.* [9] were the first to formally establish a sensitivity analysis for an unmeasured confounder in an observational study dealing with the association between smoking and lung cancer. Marcus [10], Modan *et al.* [11] and Leow *et al.* [12] also used a sensitivity analysis to evaluate the presence of unmeasured variables in their studies. More formally, Rosenbaum [13] and Greenland [8] developed two sensitivity analysis methods that allowed analyses of the behaviour of the findings obtained in a study, taking the presence of an unmeasured confounding variable into consideration.

Although developed independently and with different approaches, consideration may be given to merging these two methods in order to use a sensitivity analysis that integrates these two approaches, as both are important for assessing an unmeasured covariable in an observational study.

The purpose of this paper is to assess the impact of a possible unmeasured confounding variable in an association between the effects of access to domestic drinking water and positive results for hepatitis A serology, published by Luiz *et al.* [6], through the integrated application of these two sensitivity analysis methods.

MATERIALS AND METHODS

Data and variables

During the preparation of this paper, the data were used from Brazil's Impacts on Health and Quality of Life Assessment Project (PAISQUA; Projeto de Avaliação dos Impactos sobre a Saúde e Qualidade de Vida), which was a sectional study carried out in 1977 by the Collective Health Studies Center (NESC; Núcleo de Estudos de Saúde Coletiva) at the Rio de Janeiro Federal University, for assessing

the Guanabara Bay Clean-Up Program (PDBG; Programa de Despoluição da Baía de Guanabara). This study analyses data on access to household water supply, measured by the presence of water taps in the home (exposure variable) and positive hepatitis A serology (variable outcome), obtained through a sample of 3779 individuals in the Duque de Caxias Municipality, Rio de Janeiro, Brazil. The exposure and outcome variables were dichotomized, respectively, as no water tap *vs.* ≥ 1 water taps, and positive hepatitis A serology *vs.* negative hepatitis A serology. Details of the study in question are presented by Luiz [14].

Sensitivity analysis

The sensitivity analysis method proposed by Greenland [8] – also known as the external adjustment method – attempts to quantify the variation in the association observed in a specific study, when ‘adjusted’ by a potential unmeasured confounder in the study. This method considers the classic confounding scheme, meaning the confounder should be associated with the exposure and should be an independent outcome predictor. It consists of simulating various plausible values for the magnitude of the association between the confounder and the outcome of interest and also for the prevalence of the assumed unmeasured confounder in the exposed and non-exposed groups, calculating a new estimate for the association of interest, *i.e.* the association between the exposure variable and the outcome, adjusted by these values. Thus, by varying the values of the simulated magnitudes, this method allows an assessment of the variations in the observed association of interest, adjusted externally by an unmeasured confounder, for various estimates of this confounder. The values to be used in the assumptions should be based on considerations that are plausible for the study in question, grounded in consultations with specialists, the literature, familiarity with the subject, etc.

The Rosenbaum method [13] works only with the association between the confounder and the exposure. It determines the magnitude of the association between the unmeasured confounding variable and the exposure, sufficient for this confounding variable to be responsible for the association found between the exposure and the outcome of interest, implicitly considering that the confounder is a near perfect outcome predictor, meaning that it considers that the magnitude of the association between the confounder and

Table 1. *Effects of access to household water supply on hepatitis A seroprevalence for the group of individuals not residing close to a sanitary landfill or an open sewer*

Hepatitis A serology (<i>D</i>)	Access to household water (<i>E</i>)		Total	OR _{DE} (95% CI)
	No water tap (<i>E</i> =1)	≥1 water tap (<i>E</i> =0)		
Positive <i>D</i> =1	99	618	717	1.74 (1.30–2.35)
Negative <i>D</i> =0	113	1230	1343	1.00
Total	212	1848	2060	

OR, Odds ratio; CI, confidence interval.

Reference population: Duque de Caxias, RJ, Brazil 1997.

the outcome would be sufficient for the confounding to depend only on the association between this confounder and the exposure variable.

For dichotomic variables, the methodology utilizes the odds ratio given by the Mantel–Haenszel statistic, which is based on the number of individuals exposed and presenting the outcome, being a procedure that is frequently used in analyses where a third variable is considered that may ‘mask’ the association found between the exposure and the outcome of interest [15]. Under the null hypothesis of there being no effective exposure on the outcome of interest, the Rosenbaum method seeks the lowest value for Γ , the magnitude representing the effects of the unmeasured confounding variable on the exposure variable considered, which makes the Mantel–Haenszel statistic statistically non-significant, through the calculation of the *P* values of the upper and lower limits of this statistic, with a confidence level of 95%.

Rosenbaum uses an approach that is focused more on statistical criteria, while Greenland adopts an approach grounded more in the epidemiological considerations of the study. As both methods are important for assessing the potential effects of an unmeasured covariable in an observational study, the proposal is to merge them in order to allow a sensitivity analysis that integrates both approaches.

As a path to integration, it is proposed to use the findings produced through the application of the Rosenbaum method as the starting point for the use of the Greenland method, using the minimum value of Γ , which makes the association between the exposure and the outcome statistically non-significant, as the initial value for the set of values to be assumed for the odds ratio (OR) between the unmeasured confounding variable and the outcome of the Greenland

method. A description of the methodology proposed for the integration of both methods is presented in Cabral [16] and a tool to allow the feasibility of applying each of these methods utilizing an electronic spreadsheet in order to make it easier for researches is presented in Cabral & Luiz [17].

The ratio between these two magnitudes is given in equation (1), where P_A is the confounder prevalence in the group of exposed individuals and P_B is the confounder prevalence in the group of non-exposed individuals, the magnitude of which should also be assumed.

$$OR = \frac{P_A/(1-P_A)}{P_B/(1-P_B)} = \frac{P_A(1-P_B)}{P_B(1-P_A)} \geq \Gamma. \quad (1)$$

Another magnitude that should also be speculated in the use of the Greenland method is the odds ratio between the unmeasured confounder and the outcome. Through assumptions of all these magnitudes, the magnitude may be estimated for the association between the unmeasured confounder and the outcome considered explicitly in the Rosenbaum method, but which is not quantified.

RESULTS

In order to assess the impact of a possible unmeasured confounding variable on the association of the effects of access to domestic drinking water and positive hepatitis A serology [6], the data are considered for the group of individuals who do not live close to a sanitary landfill or an open sewer, as the association between access to water and hepatitis A seroprevalence did not prove significant for the group of individuals living close to a sanitary landfill or an open sewer. The findings are presented in Table 1. The

Table 2. Sensitivity analysis (Rosenbaum technique) of the association with access to household water supply for individuals not living close to a sanitary landfill or an open sewer, in the presence of an unmeasured variable confounder Z

Γ	P value	
	Upper limit	Lower limit
1	<0.001	<0.0001
1.1	0.001	<0.0001
1.2	0.006	<0.0001
1.3	0.026	<0.0001
1.4	0.076	<0.0001

The P value recorded is for the Mantel–Haenszel statistic, based on the number of cases for various possible Γ values.

association considered in the preparation of this paper is the odds ratio.

According to Table 1, an individual with no water tap in the home is 1.74 times more likely to present positive hepatitis A serology than an individual with one or more water taps in the home.

The application of the sensitivity analysis proposed as the integration of the two methods initially uses the Rosenbaum method, whose findings are presented in Table 2.

According to these findings, the lowest Γ value making access to water non-significant for hepatitis A serology is $\Gamma=1.4$, meaning that this association is shown to be ‘sensitive’ to a confounder that increases by 40% the odds of exposure with access to household water supply, and would be a near perfect predictor for hepatitis A seroprevalence.

In order to assess the magnitude of the association between the confounder and the outcome of interest (OR_{ZE}) and also the variations in the estimates of the exposure odds ratios with the outcome (OR_{DE}) adjusted by the unmeasured confounder, the Greenland method is used, taking as possible OR_{ZE} values the values obtained through the Rosenbaum method from 1.4 onwards. This paper assumes that the OR_{ZE} values are fixed at 1.4, 2.0, 4.0 and 5.0, with a set of confounder prevalences among exposed individuals (P_A) varying between 0.90 and 0.10. The values for P_B were calculated in compliance with equation (1). The consideration of a value for $OR_{ZE}=5.0$ means that the confounder increases the odds of an individual being exposed fivefold, which is a value rarely noted in most epidemiological studies.

Table 3 presents the findings of the sensitivity analysis using the Greenland method, based on the findings of the sensitivity analysis obtained by the Rosenbaum method.

The analysis in Table 3 shows that the adjusted OR_{DE} values are quite different from the OR_{DE} value observed (1.74), mainly when the confounder is an important factor for the occurrence of hepatitis A (OR_{DZ} values from 4 upwards). If the OR_{DZ} value is <4 , a stronger association is needed between the confounder and the exposure (OR_{DZ} values also ≥ 4), so that the adjusted OR_{DE} values move away from 1.74. According to the Rosenbaum method, for the association observed between access to household water supply and hepatitis A seroprevalence to be due to an unmeasured confounder, it is necessary for this confounder to increase the odds of exposure by 40%, assuming that it is a near perfect outcome predictor. Analysing the findings presented in Table 3, it is clear that for a 40% increase between the confounder and access to water for the odds of exposure to throw doubt on the association found, the odds ratio between this confounder and the outcome should present a value ≥ 4 , as it is from this value onwards that the adjusted OR_{DE} starts to become statistically non-significant, at a confidence level of 95%. Even if the confounder doubles the odds of an individual being exposed, this alone is not sufficient to negate the observed association of interest. Moreover, the confounder must also present an odds ratio of at least 5, with the outcome variable. The plausibility of the existence of an unmeasured confounder with these characteristics should be assessed carefully. It is difficult to believe that a variable of this importance has not been observed, based on current knowledge of hepatitis A, thus indicating that an association between access to water and positive hepatitis A serology is not likely to be due to an unmeasured confounder in the process.

DISCUSSION

Hepatitis A seroprevalence is generally due to faecal–oral transmission. Personal hygiene, sanitary conditions and population density are risk factors identified for transmission [18]. Serological surveys indicate that the positive hepatitis A serology prevalences vary from 15% to almost 100% among populations in the less-developed countries [19]. It is known that the incidence and prevalence of hepatitis A are directly related to social and economic

Table 3. Sensitivity analysis (Greenland technique) of the association of access to household water supply (dichotomized as no water tap in the home and ≥ 1 water taps), hepatitis A seroprevalence for individuals not residing close to a sanitary landfill or an open sewer, in the presence of an unmeasured confounding variable

OR _{ZE}	P _A	P _B	OR _{DE} adjusted (95% CI)			
			OR _{DZ} =1.5	OR _{DZ} =2	OR _{DZ} =4	OR _{DZ} =5
1.4	0.90	0.87	1.73 (1.26–2.34)	1.72 (1.26–2.34)	1.70 (1.25–2.30)	1.70 (1.25–2.30)
	0.70	0.63	1.70 (1.20–2.41)	1.67 (1.20–2.38)	1.63 (1.19–2.22)	1.62 (1.16–2.25)
	0.50	0.42	1.69 (1.12–2.54)	1.65 (1.10–2.43)	1.58 (1.11–2.24)	1.56 (1.07–2.26)
	0.30	0.23	1.69 (1.16–2.44)	1.65 (1.23–2.33)	1.55 (1.05–2.36)	1.52 (0.94–2.37)
	0.20	0.15	1.70 (1.22–2.40)	1.67 (1.18–2.38)	1.58 (1.04–2.45)	1.55 (1.05–2.32)
	0.10	0.07	1.72 (1.25–2.36)	1.70 (1.23–2.33)	1.62 (1.16–2.30)	1.59 (1.13–2.25)
2.0	0.90	0.82	1.70 (1.24–2.30)	1.67 (1.22–2.27)	1.63 (1.20–2.21)	1.62 (1.19–2.20)
	0.70	0.54	1.64 (1.16–2.33)	1.58 (1.13–2.25)	1.47 (1.06–2.03)	1.45 (1.04–2.03)
	0.50	0.33	1.63 (1.06–2.43)	1.55 (1.04–2.33)	1.39 (1.04–1.99)	1.35 (0.92–1.96)
	0.30	0.18	1.65 (1.14–2.48)	1.58 (1.08–2.31)	1.41 (1.04–2.24)	1.36 (0.88–2.17)
	0.20	0.11	1.67 (1.20–2.36)	1.61 (1.14–2.30)	1.45 (1.05–2.15)	1.39 (0.94–2.08)
	0.10	0.05	1.70 (1.24–2.33)	1.66 (1.20–2.28)	1.54 (1.05–1.96)	1.49 (1.06–2.11)
4.0	0.90	0.69	1.62 (1.17–2.23)	1.55 (1.13–2.12)	1.45 (0.84–1.67)	1.43 (1.05–1.94)
	0.70	0.37	1.53 (1.13–2.26)	1.41 (1.12–2.05)	1.19 (0.73–1.97)	1.14 (0.81–1.60)
	0.50	0.20	1.53 (1.13–2.36)	1.39 (1.07–2.12)	1.12 (0.89–1.89)	1.05 (0.71–1.54)
	0.30	0.10	1.59 (1.11–2.25)	1.48 (1.12–2.13)	1.19 (0.87–1.86)	1.11 (0.71–1.76)
	0.20	0.06	1.63 (1.09–1.88)	1.54 (1.15–2.02)	1.29 (1.05–2.05)	1.20 (0.81–1.78)
	0.10	0.03	1.69 (1.22–2.28)	1.63 (1.18–2.12)	1.46 (1.00–1.88)	1.39 (0.99–1.89)
5.0	0.90	0.64	1.59 (1.14–2.21)	1.51 (1.10–2.09)	1.38 (0.77–1.55)	1.35 (0.99–1.83)
	0.70	0.32	1.50 (1.01–2.29)	1.35 (0.92–1.99)	1.10 (0.61–1.84)	1.05 (0.74–1.48)
	0.50	0.17	1.51 (1.01–2.32)	1.36 (0.87–2.13)	1.05 (0.75–1.75)	0.98 (0.55–1.80)
	0.30	0.08	1.58 (1.10–2.23)	1.45 (1.00–2.08)	1.14 (0.87–1.85)	1.05 (0.68–1.65)
	0.20	0.05	1.62 (1.17–2.26)	1.53 (1.09–2.15)	1.25 (0.99–1.96)	1.16 (0.79–1.73)
	0.10	0.02	1.68 (1.23–2.29)	1.62 (1.17–2.21)	1.42 (1.03–2.00)	1.35 (0.96–1.89)

OR, Odds ratio; CI, confidence interval.

The 95% CI was calculated by the Cornfield method [15].

Reference population: Duque de Caxias, RJ, Brazil 1997.

conditions [20]. A large proportion of waterborne diseases in populations where they are endemic is due to the limited amounts of water available for domestic consumption and personal hygiene [4]. The challenge for epidemiologists is to develop indicators that assess the impact of sanitary upgrade projects.

The work by Luiz *et al.* [6] analysed the relationship between access to household water (measured in terms of water taps available in the home) and positive for hepatitis A seroprevalence, as a specific example of the relationship between sanitary conditions and health. These authors concluded that a limited water supply is an important factor for the occurrence of hepatitis A, particularly when there is no exposure to a sanitary landfill or an open sewer close to the home. As this is an observational study, one possibility would be that the association found is the outcome of the presence a confounding variable that is not

measured in the study. This paper shows that the association between access to water and positive hepatitis A serology [6] is relatively insensitive to a non-controlled confounder, meaning the sensitivity analysis used in its assessment suggests that the causality of association hypothesis should not be undermined by the presence of a possible unmeasured confounding variable. However, the fragility of a sectional study is well known when there is a causal assumption, with additional hypotheses being required to assess the causal hypothesis in terms of the prognosis and reverse causality in time.

Considering the current knowledge about hepatitis A, it is unlikely to exist as an unmeasured variable potentially related to the outcome presenting an odds ratio close to 4. Basically, in observational studies, the odds ratio found, once the other variables involved are controlled for, is < 2 [2]. An alternative way

to assess the strength of the association of a given risk factor with an outcome is to estimate the exposure intensity necessary for that factor to produce an association of the same magnitude as that of a well-established risk factor or vice versa [21].

As this is an observational study, the association between access to water and hepatitis A serology will be difficult to measure with complete certainty. However, a sensitivity analysis allows the magnitude to be calculated for the association between the unmeasured confounder and the exposure and outcome variables being studied, in order to assess the potential for the observed association being explained by the presence of this confounder. Although a sensitivity analysis neither confirms nor eliminates the presence of an unmeasured confounder, it may be used by the researcher as a quantitative assessment that can be integrated with the analysis carried out during the validation of findings stage, particularly in observational studies. Moreover, this paper discusses the presence of a single unmeasured confounder, when multiple confounders may exist and blend, altering the association of interest and requiring more complex calculations that are beyond the scope of this study. Furthermore, consideration may be given to using a sensitivity analysis for assessing possible classification errors, as the serological tests performed in order to discover whether a person is positive or negative have a certain specificity and sensitivity, and consequently an associated classification error. For more details about this type of sensitivity analysis, see Greenland [8].

As the most part of epidemiological studies is observational, the models used depend on assumptions that frequently can not be checked by the observed data, meaning that the discussion of causality addresses the study of the validity of the findings obtained in the studies. The sensitivity analysis could be considered an appropriate tool to enhance causal conclusions in observational studies. The integration of the two sensitivity analysis methods presented proved useful, using the Rosenbaum method to reduce and guide the assumptions to be considered for the association between the confounder and the exposure variable (OR_{ZE}) in the Greenland method, stipulating values for this association that are equal or greater than the Γ value found ($OR_{EZ} \geq \Gamma$ values), and the Greenland method, in order to calculate the variations in the estimates of the OR_{DE} , arising from the consideration of the various possible distributions of the confounder, as well as determining the

magnitude of the near perfect association between the unmeasured confounder and the outcome of interest, implicit but not quantified by the Rosenbaum method. In conclusion, this paper shows that the association between access to water and positive hepatitis A serology found by Luiz *et al.* [6] is unlikely to be due to an unmeasured confounder in the study.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Blum D, Feachem RG.** Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *International Journal of Epidemiology* 1983; **12**: 357–365.
2. **Almeida LM, et al.** The epidemiology of hepatitis A in Rio de Janeiro: environmental and domestic risk factors. *Epidemiology and Infection* 2001; **127**: 327–333.
3. **Jacobsen KH, Koopman JS.** Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiology and Infection* 2004; **132**: 1005–1022.
4. **Cairncross S.** Water quality, quantity and health. In: Drangert JO, Swiderki R, Woodhouse M, eds. *Conference on Safe Water Environments*, Eldoret, Kenya. Linköping, Sweden: Linköping University, 1996, pp. 60–65.
5. **Cvjetanovic B.** Health effects and impact of water supply and sanitation. *World Health Statistics Quarterly* 1986; **39**: 105–117.
6. **Luiz RR, et al.** The relation between anti-hepatitis A virus antibodies and residence water access in Rio de Janeiro, Brazil. *International Journal of Hygiene and Environmental Health* 2003; **206**: 575–582.
7. **Rosenbaum PR, Rubin DB.** Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society* 1983; Series B, **45**: 212–218.
8. **Greenland S.** Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology* 1996; **25**: 1107–1116.
9. **Cornfield J, et al.** Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 1959; **22**: 173–203.
10. **Marcus S.** Using omitted variable bias to assess uncertainty in an AIDS education treatment effect. *Journal of Educational and Behavioral Statistics* 1997; **22**: 193–201.
11. **Modan B, et al.** Cancer incidence in a cohort of infertile women. *American Journal of Epidemiology* 1998; **147**: 1038–1042.
12. **Leow C, et al.** Effects of advanced course-taking on math and science achievement: addressing selection bias using propensity scores. *American Journal of Evaluation* 2004; **25**: 461–478.

13. **Rosenbaum PR.** *Observational Studies*. New York: Springer-Verlag, 1995.
14. **Luiz RR.** Assessing causal effect between domestic water access and hepatitis A: statistical and epidemiological considerations from a cross-sectional study [Thesis]. Rio de Janeiro, Brasil: Universidade Federal do Rio de Janeiro, 2002, 144 pp.
15. **Fleiss JL.** *Statistical Methods for Rates and Proportions*, 2nd edn. New York: Wiley, 1981.
16. **Cabral MDB.** Sensitivity analysis in epidemiological studies [Dissertation]. Rio de Janeiro, Brasil: Universidade Federal do Rio de Janeiro, 2005, 69 pp.
17. **Cabral MDB, Luiz RR.** Ensuring the feasibility of a sensitivity analysis for an unmeasured confounder through an electronic spreadsheet. *Revista de Saúde Pública* (in press).
18. **Moyer L, Warwick M, Mahoney FJ.** Prevention of hepatitis A virus infection. *American Family Physician* 1996; **54**: 107–114.
19. **World Health Organization.** Hepatitis A vaccine. *Weekly Epidemiological Record* 2000; **75**: 38–44.
20. **Vitral CL, et al.** Age-specific prevalence of antibodies to hepatitis A in children and adolescents from Rio de Janeiro, Brazil, 1978 and 1995: relationship of prevalence to environmental factors. *Memórias do Instituto Oswaldo Cruz* 1998; **93**: 1–5.
21. **Szklo M, Nieto FJ.** *Epidemiology: Beyond the Basics*, 2nd edn. Boston: Jones and Bartlett Publishers, 2006.