 International Journal of Epidemiology 1999;**28**:776–781

Hepatitis A incidence rate estimates from a pilot seroprevalence survey in Rio de Janeiro, Brazil

Claudio José Struchiner,[a] Liz Maria de Almeida,[b] Raymundo Soares de Azevedo[c] and Eduardo Massad[c]

Background To assess the impact of water sanitation and sewage disposal, part of a major environmental control programme in Rio de Janeiro, we carried out seroprevalence studies for Hepatitis A virus (HAV) in three micro-regions in Rio de Janeiro. Each region varied with regard to level of sanitation. We are interested in assessing the discriminating power of age-specific prevalence curves for HAV as a proxy for improvement in sanitation. These curves will serve as baseline information to future planned surveys as the sanitation programme progresses.

Methods Incidence rate curves from prevalence data are estimated parametrically via a Weibull-like survival function, and non-parametrically via maximum likelihood and monotonic splines. Sera collected from children and adults in the three areas are used to detect antibodies against HAV through ELISA.

Results We compare baseline incidence curves at the three sites estimated by the three methods. We observe a strong negative correlation between level of sanitation and incidence rates for HAV infection. Incidence estimates yielded by the parametric and non-parametric approaches tend to agree at early ages in the micro-region showing the best level of sanitation and to increasingly disagree in the other two.

Conclusion Our results support the choice of HAV as a sentinel disease that is associated with level of sanitation. We also introduce monotonic splines as a novel non-parametric approach to estimate incidence from prevalence data. This approach outperforms current estimating procedures.

Keywords Hepatitis A, sanitation, water decontamination, waste disposal, current status data, NPMLE, monotonic splines

Accepted 13 January 1999

The Guanabara Bay Decontamination Program (GBDP) in Rio de Janeiro, Brazil, is a major attempt to change a long-term pattern of waste disposal from various sources, including industrial waste and untreated sewage. One of its components is to provide the dense population that lives in close proximity with basic sanitation and water supply. Environmental interventions leading to improved basic sanitation demand considerable economic resources and are justified on the assumption of a direct association between better sanitation and improved quality of life.[1] Several studies have been conducted in order to assess the impact of water supply and sanitation on health as measured by health indicators, such as morbidity and mortality rates.[2–7] Contradictory results regarding this association have been reported so far, casting doubts on a favourable outcome of such health intervention strategies and raising important questions regarding the research methodologies used in field studies.[8–15]

In the light of this it is important that the overall impact of GBDP should be assessed carefully. It is expected that the GBDP will successfully impact on health status, lowering the incidence of diseases transmitted through contact with contaminated water. As a first approach, the problem of representing complex disease transmission could be considerably reduced by choosing a sentinel disease that is sensitive to changes in levels of sanitation and water supply. In this context, Hepatitis A virus is a good candidate since the dependence of its transmission cycle on sanitation is well established,[16–18] highly sensitive and specific

[a] PROCC/Fundação Oswaldo Cruz, Av. Brasil 4365, Residência Oficial, 21045–900, Rio de Janeiro, Brasil and Instituto de Medicina Social (IMS)/UERJ, Av. São Francisco Xavier 524—7° andar, Bloco D, Maracanã 20559–900, Rio de Janeiro, Brasil.

[b] NESC/ Universidade Federal do Rio de Janeiro, Av. Brigadeiro Trompowski s/n, Edificio do HUCFF, 5°. andar, ala sul, Cidade Universitária, Ilha do Fundão, 21941–590, Rio de Janeiro, RJ, Brasil.

[c] LIM01-HCFMUSP and Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, Avenida Doutor Arnaldo 455, São Paulo, 01246–903, Brasil.

Table 1 Hepatitis A seroprevalence data by age and micro-regions. Rio de Janeiro, Brazil, 1996

	MR 112			MR 111			MR 9		
	Prevalence			Prevalence			Prevalence		
Age (years)	n	%	95% CI	n	%	95% CI	n	%	95% CI
1	24	12.5	3.3–33.5	21	9.5	1.7–31.8	26	7.7	1.3–26.6
2	41	9.8	3.2–24.1	17	11.8	2.1–37.7	15	0.0	–
3	29	10.3	2.7–28.5	18	11.1	1.9–36.1	26	0.0	–
4	31	32.3	17.3–51.5	21	14.3	3.8–37.4	28	10.7	2.8–20.4
5	23	43.5	23.9–65.1	21	28.6	12.2–52.3	26	0.0	–
6	22	31.8	14.7–54.9	18	16.7	4.4–42.3	14	14.3	2.5–43.9
7	23	39.1	20.5–61.2	27	29.6	14.5–50.3	25	12.0	3.1–32.3
8	26	30.8	15.1–51.9	28	17.9	6.8–37.6	22	27.3	11.6–50.4
9	34	58.8	40.8–74.9	19	21.1	7.0–46.1	27	33.3	17.2–54.0
10	6	16.7	0.9–63.5	23	13.0	3.4–34.7	28	17.9	6.8–37.6
11	9	33.3	9.0–69.1	42	33.3	20.0–49.6	25	12.0	3.1–32.3
12	12	58.3	28.6–83.5	18	27.8	10.7–53.6	28	17.9	6.8–37.6
13	14	71.4	42.0–90.4	31	38.7	22.4–57.7	21	38.1	18.9–61.3
14	8	25.0	4.4–64.4	20	40.0	20.0–63.6	23	30.4	14.1–53.0
15	13	76.9	46.0–93.8	25	48.0	28.3–68.2	28	39.3	22.1–59.3
16	3	0.0	–	7	42.9	11.8–79.8	10	40.0	13.7–72.6
17–20	16	81.2	53.7–95.0	6	33.3	6.0–75.9	15	53.3	27.4–77.7
⩾21	18	100.0	78.1–100.0	–	–	–	4	100.0	39.6–100.0
Total	352			362			391		

serological tests are available,[19] and evidence of infection persists almost indefinitely. Therefore, the comparison of cheap and efficient study designs to assess age- and time-specific long-term evolution of incidence patterns of HAV needs to be addressed.

Repeated epidemiological surveys of antibody prevalence provide useful information for assessing the immune status in populations, and hence incidence rates of infection or disease. They have been recommended[20–23] as an attractive alternative to the more expensive and invasive approach of collecting longitudinal follow-up data on a cohort of individuals.[24,25] Serological surveillance can be used to identify groups at risk, by age or geographical location for example, making it possible to change or improve health control strategies.

In this paper, we report on first estimates of HAV incidence rates from seroprevalence data gathered by carrying out a pilot epidemiological survey in three areas in Rio de Janeiro, Brazil, prior to the construction of a new system of water and sewage drainage. Each region is subject to a varying degree of level of sanitation. We are interested in assessing the discriminating power of age-specific prevalence curves for HAV as a proxy for improvement in level of sanitation. In order to assess the impact of this programme on health, the incidence curves will serve as baseline information for repeated seroprevalence surveys for HAV, expected to take place in the near future as the sanitation programme progresses.

Subjects and Methods

Setting and research design

A cross-sectional pilot study was conducted from August to November of 1996 to estimate the prevalence of antibodies against HAV by age, and to validate the use of eluate of blood collected onto filter paper for epidemiological studies on hepatitis. The study was carried out in three small areas in the state of Rio de Janeiro, Brazil: two of them (MR 111 and 112) in Campos Elyseos (Duque de Caxias) and the third one (MR 9) in Ilha do Governador. The micro-regions were arbitrarily selected, according to available crude prior information, in order to describe a broad range of level of sanitation and water supply. In Sector 112 the studied population comprised a census of all residents 1–9 years old, and, in Sectors 9 and 111, a census of individuals 1–15 years old. Blood samples were also collected from older individuals in each of these sectors, either by request of their families or the individuals themselves, giving the totals shown in Table 1.

Survey team and protocol

Twenty interviewers and 12 individuals trained in nursing practice, supervised by at least two researchers during the field work, visited all dwellings comprising the three districts, and collected the necessary information to draw a map of the area displaying addresses and the age composition of each home. Field workers were instructed to explain the objectives of the study to all participants. Each family provided a written informed consent, and completed a questionnaire containing information regarding house water supply and sewage system.

Sample collection, preparation, and laboratory tests

Blood collected by venepuncture was obtained by vacuum containing system (Vacutainer™, Becton Dickinson Company, USA) or by syringes and needles. The sera thus obtained were stored at –20°C. Total antibodies against HAV was determined by competitive enzyme-linked immunosorbent assay, ELISA (HAVAB EIA, Abbott Laboratories, USA).

Table 2 Summary of the classification criteria, which considers socioeconomic, sanitary facilities and household characteristics, used to rank the micro-regions according to level of sanitation

Sector 112	< Proportion of households with kitchen
	< Proportion of households linked to the water system
	> Proportion of households using water of well
	> Proportion of households storaging water into improper places
Sector 111	The highest family mean income (by month)
	The highest education level of household
	> Proportion of households with complete sanitary facilities (latrine, washbasin and shower)
	> Proportion of households with septic tank linked to pluvial system
Sector 9	All households linked to the water system and regular water supply
	All households with regular waste collection

Statistical analysis

Incidence rates from prevalence data are estimated parametrically via a Weibull-like survival function, and non-parametrically via maximum likelihood and monotonic splines. Following the approach described in Grummer-Strawn,[26] we fitted by least squares the parametric survival model given by

$$G(a) = 1 - \rho e^{-\left(\left(\frac{a}{\beta}\right)^{p}\right)}, \quad (1)$$

to the age-specific prevalence data, G(a), assuming a constant variance structure for the error.[27,28] A test for the equality of the parameter estimates for each sector was carried out by the Wald test and maximum likelihood ratio. When ρ equals 1, this is precisely the cumulative distribution function of a Weibull survival model. Parameter ρ denotes the proportion of seropositivity at birth and is interpreted as the proportion of infants still carrying maternal antibodies.

Non-parametric estimates of incidence rates from the marginal age-status distribution in a cross-sectional sample are discussed in Keiding.[22] Looking at the problem as if all observations were either left or right censored, he shows that the non-parametric maximum likelihood estimator of G(a) is defined as the left-continuous derivative of the convex minorant of the cumulative number of people with antibodies against HAV as age increases. Groeneboon[29] first introduced this estimator and Wim Penninx, from Delf University of Technology, developed the related estimating algorithm and kindly made it available to us.

The role played by the convex minorant in the previous approach could be replaced by monotonic splines.[30] Scatterplot smoothers have attracted much attention lately[31,32] since they do not require the imposition by the data analyst of a given parametric structure on the data. For this reason, they have become powerful methods of data description, exploration and ultimately model specification and estimation. In the present context, monotonic scatterplot smoothers are fitted to the data pairs formed by the cumulative number of people with antibodies against HAV, and the cumulative number of people in the study for each age, in a manner similar to the convex minorant approach.[22,29] Cubic regression splines with bounds on the coefficients offer a simple and effective approximation to monotonic, convex or concave, smoothing splines.[33] Once the fit is available, incidence rates are estimated as above, however, derivatives can be obtained directly from the regression splines and are, in this respect, more convenient than the approximate derivatives obtained as the output of the convex minorant approach. We fitted the regression splines with Cosmo, a set of Splus functions developed by Dave Dole and made publicly available through the web (DOLE, D – Scatterplot Smoothing Subject to Monotonicity and Convexity. Working Paper accompanying the software Cosmo available from: http://www.general.uwa.edu.au/u/ddole/cosmo/intro.html, 1997). To our knowledge this approach is original and looks promising.

Results

Table 1 contains prevalence data on the presence of HAV antibody stratified by the three micro-regions MR9, MR111 and MR112 under surveillance, and age. Preliminary analysis of the data gathered through the questionnaires regarding the household sanitary conditions, and socioeconomic status, and summarized[33] in Table 2, indicate that the regions could be ranked according to known correlates of HAV transmission. Micro-region 112 ranked lowest according to all dimensions explored. Micro-region 111 ranked best regarding household conditions, family income, and housewife level of education. Micro-region 9 benefits from improved water supply and garbage collecting systems.

Parameter estimates and standard deviations obtained by the parametric modelling approach are presented in Table 3. Figure 1 depicts the age-specific data points and the parametric regression fit for each micro-region. All models indicate that the proportion of seropositivity at birth, ρ, is significantly different from zero justifying the adjustment for the proportion of infants still carrying maternal antibodies. The remaining parameters, α and p, allow us to plot the hazard function under a Weibull model (Figure 2), which is given by

$$h(a) = p\lambda(a\lambda)^{p-1}, \quad (2)$$

where $\lambda = \exp(\alpha/p)$. These two parameters are increasingly more significant for MR9, MR111 and MR112, respectively. Allowing for a single set of three parameters, instead of nine (three for each micro-region), we can test for equality of all (ρ, α, p) among the three micro-regions (line labelled 'Equal' in Table 3). The likelihood ratio test and Wald test yield a χ^2_{6df} of 35.87 ($P < 0.01$) and 47.90 ($P < 0.01$), respectively, meaning that the regions differ in at least one of the parameters.

Both non-parametric incidence estimates are based on data on the cumulative sum of seropositives and study participants. Figure 3 depicts the pooled raw data, as well as the stratified

Table 3 Parameter estimates and standard deviation for the adjusted Weibull model stratified by micro-region of study and for the total samples pooled together. Pooled—all observations pooled together; MR 9—micro-region 9; MR 111—micro-region 111; MR 112—micro-region 112; Equal—fit allowing for a single set of three parameters for all three micro-regions

	ρ		α		*P*	
Model	Value (se)	t value	Value (se)	t value	Value (se)	t value
Pooled	0.819 (0.031)	27.029	–11.666 (2.068)	–5.641	3.953 (0.706)	5.602
MR 9	1.016 (0.132)	7.080	–3.713 (2.372)	–1.565	1.026 (0.787)	1.304
MR 111	0.885 (0.072)	12.290	–5.608 (2.952)	–1.899	1.721 (1.047)	1.644
MR 112	0.822 (0.063)	12.998	–5.810 (1.791)	–3.243	2.097 (0.655)	3.203
Equal	0.901 (0.601)	14.781	–5.211 (1.818)	–2.867	1.681 (0.646)	2.604

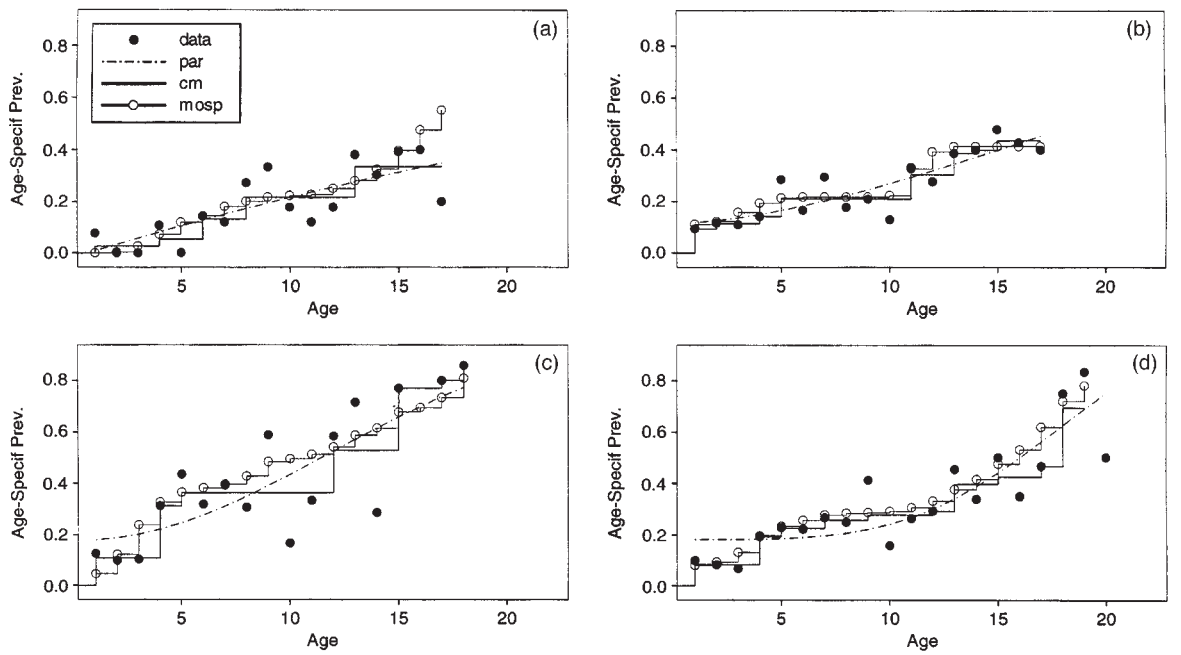

Figure 1 Observed and predicted age-specific prevalence data (A micro-region 9; B micro-region 111; C micro-region 112; D Pooled data)

data—observed data points; par—estimates by parametric approach; cm—estimates by the convex minorant approach; mosp—estimates by monotonic splines.

data for each micro-region. Individual curves are fitted to the data by the convex minorant and monotonic splines approaches (not shown). The left-continuous derivatives of these functions estimate the age-specific prevalence and are shown in Figure 1 together with the observed data points. On the same Figure we also plot the parametric fit.

Incidence curves are presented in Figure 2. We plot h(a) according to the expression above under the parametric approach, and the individual point estimates of this function under the two non-parametric approaches. Also presented in the Figure are the smoothing loess[28] curves for the latter two estimates. Several patterns emerge when analysing these figures: (a) MR112 is subject to the highest incidence rates, particularly at both extremes of the age range, irrespective of the estimating method; (b) MR111 is subject to lower incidence rates than MR112 which seem to peak at age 12; (c) MR9 presents the lowest incidence rates at initial ages and follows the same pattern as MR111 up to age 10 when the estimating methods are no longer consistent; (d) unexpectedly, incidence rates increase with age, starting around age 10, in MR112 and 111 (peaking at age 12), and, possibly, in MR9 (at least if one considers the adjustment by monotonic splines).

Discussion

The methods used to estimate incidence of infection from data on current serological status require a steady-state assumption, so that the only time-like variable is age. When single cross-sectional data are available, interpretation of the rate estimates as obtained in this paper require the often made stationarity or time homogeneity assumptions which, in epidemiological jargon, translates into absence of cohort effects. However, when repeated cross-sectional observations separated by a time interval become available in the near future, as expected, the time homogeneity assumption will no longer be necessary. One needs further to assume that everyone is uninfected at birth and that prevalence can only increase with age, as a marker of having ever had contact with HAV. Some of these assumptions

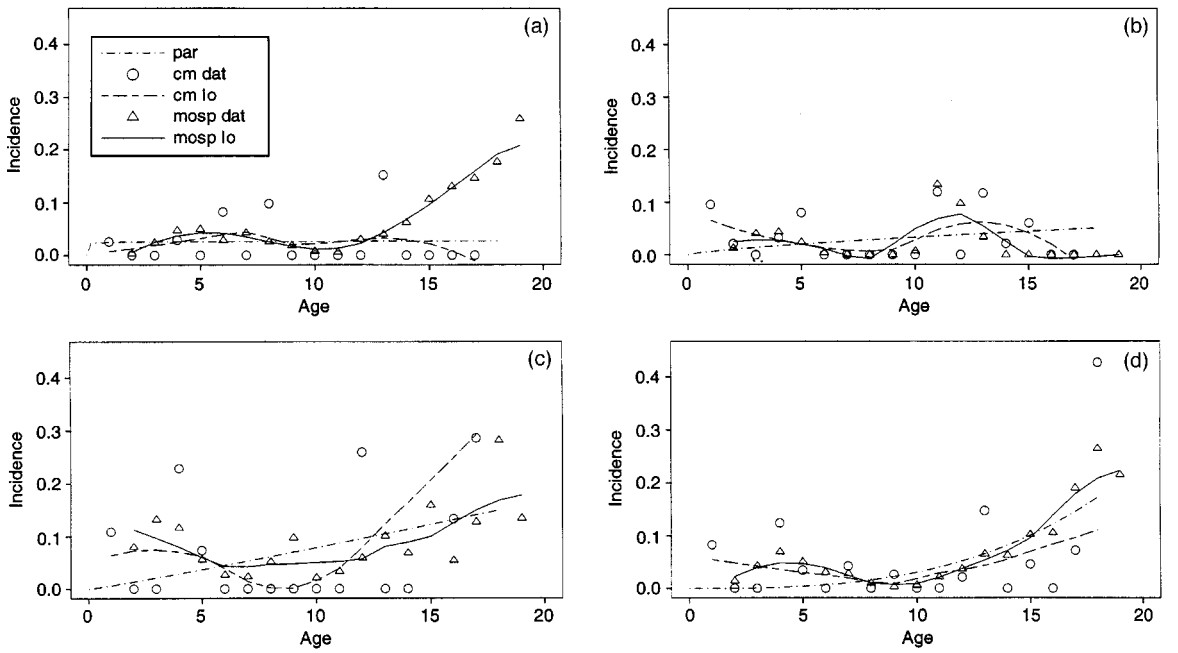

Figure 2 Incidence rates by age (A micro-region 9; B micro-region 111; C micro-region 112; D Pooled data)

par—estimates by parametric approach; cm dat—individual data point estimates by convex minorant approach; cm lo—smoothed lowess curves of the individual data point estimates by the convex minorant approach; mosp dat—individual data point estimates by monotonic splines; mosp lo—smoothed lowess curves of the individual data point estimates by monotonic splines.

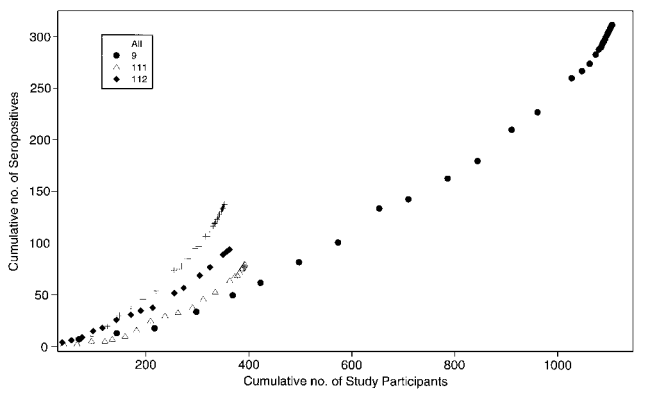

Figure 3 Plot of cumulative number of seropositives against cumulative number of study participants stratified by area of study: All complete data set pooled together; 9 micro-region 9; 111 micro-region 111; 112 micro-region 112

can be relaxed as suggested by Brunet and Struchiner,[21] and time dependencies besides age can be introduced as recently proposed by Marschner.[23] Interpretation of the estimates presented here can also be affected by additional sources of departure from the usual assumptions which include the presence of maternal antibodies at young ages, in- and out-migration from the study areas, and reversibility of immune status.

The dependence of estimating procedures on the subpopulation demographics is discussed by Keiding[22] and Brunet and Struchiner.[21] The classical approach to such a system requires complete historical records of migration fluxes, detailed *ab initio* distribution of contingents in epidemiological categories, and historical records of birth rates. By focusing on the instantaneous relationships between prevalence and incidence rates, Brunet and Struchiner[21] relax most of these assumptions and eliminate any need for knowledge of the history of the system. Thus, birth rates at any time and migration fluxes at previous times are of no concern.

Incidence estimates yielded by the parametric and non-parametric approaches tend to agree at early ages in MR9 and increasingly disagree in MR111 and MR112. Since, in the former approach, parameter ρ specifically describes non-zero prevalence of seropositivity at birth, i.e. the presence of maternal antibodies against HAV, we infer that the non-parametric estimates are not too far off the target in MR9. However, the downward trend at early ages verified for MR111 and MR112 under the non-parametric estimating approaches is puzzling and should be carefully considered. Recent advances in model building and parameter estimation would allow us to account for a more complex dynamics.[34] In this work, however, we have not pursued any further these more elaborated model constructs.

Classification of micro-regions along ordered levels of sanitation was somewhat informal in this work. A better classification scheme and integration with other epidemiological covariates can be achieved by proportional hazard, accelerated failure time, generalized linear and additive (GLM and GAM) modelling approaches, as suggested by Marschner.[23] In addition to the order-based classification used in this paper, it would also be interesting to explore an improved classification in which the distance between classes could also be estimated. Non-parametric methods (convex minorants) are quick, and one could do hypothesis testing using a bootstrap approach. However, they are very inefficient statistically (i.e. they have

large variances which result from not having sqrt(n) rates of convergence). The parametric approaches mentioned in this paragraph could also improve on these shortcomings.

Incidence rates estimated by monotonic splines at older ages, in MR9, differ from the estimates by the other two methods. Visual inspection of the cumulative plot for all three micro-regions (Figure 3) point to a steeper curve for MR9 indicating that the monotonic spline approach seems to better capture this behaviour. Moreover, Keiding[22] reports similar curves for incidence rates increasing with age, starting around age 10, as seen in MR 112 and MR 111 (peaking at age 12), and, possibly, in MR9 (at least if one considers the adjustment by monotonic splines).

Our results support the choice of HAV as a sentinel disease that is associated with level of sanitation. It is expected that future seroprevalence surveys will reflect the impact of water sanitation and sewage disposal as part of the major environmental control programme, which is now being carried out in Rio de Janeiro. The age-specific prevalence and incidence curves estimated in this work would serve as baseline information as the sanitation programme progresses.

Acknowledgements

We are ever indebted to the population of Duque de Caxias City and Colônia Z-10 for being so co-operative. We are grateful to Mr Paulo Eduardo Coletty and Mr Ronir Raggio Luiz for their help during the field work and data analysis and Dr Angelica Arpon Marandino Guimarães for the laboratory analysis. This study is included in the health and life quality impact evaluation project of GBDP (PAISQUA) and is supported by FAPERJ. The authors were partially supported by FAPERJ, PRONEX/FINEP and CNPq.

References

[1] Cvjetanovic B. Health effects and impact of water supply and sanitation. *World Health Stat Q* 1986;**39:**105–17.

[2] Victora CG, Smith PG, Vaughan JP *et al.* Water supply, sanitation and housing in relation to the risk of infant mortality from diarrhoea. *Int J Epidemiol* 1988;**17:**651–54.

[3] Daniels DL, Cousens SN, Makoae LN, Feachem RG. A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho. *Bull World Health Organ* 1990;**68:**455–63.

[4] Harpham T, Stephens C. Urbanization and health in developing countries. *World Health Stat Q* 1991;**44:**62–69.

[5] Knight SM, Toodayan W, Caique WC, Kyi W, Barnes A, Desmarchelier P. Risk factors for the transmission of diarrhoea in children: a case-control study in rural Malaysia. *Int J Epidemiol* 1992;**21:**812–18.

[6] VanDerslice J, Popkin B, Briscoe J. Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bull World Health Organ* 1994;**72:**589–601.

[7] Kindzierski WB, Gabos S. Health effects associated with wastewater treatment, disposal and reuse. *Water Environ Res* 1995;**67:**749–55.

[8] Shuval HI, Tilden RL, Perry BH, Grosse RN. Effect of investments in water supply and sanitation on health status: a threshold-saturation theory. *Bull World Health Organ* 1981;**59:**243–48.

[9] Blum D, Feachem RG. Measuring the impact of water supply and sanitation investments on diarrheal diseases: problems of methodology. *Int J Epidemiol* 1983;**12:**357–65.

[10] Brisco EJ. Interventions studies and the definition of dominant transmission routes. *Am J Epidemiol* 1984;**120:**449–55.

[11] Esrey AS, Feachem RG, Hughes JM. Interventions for the control of diarrheal diseases among young children: improving water supplies and excreta disposal facilities. *Bull World Health Organ* 1985;**63:**757–72.

[12] Esrey AS, Habicht JP. Epidemiologic evidence for health benefits from improved water and sanitation in developing countries. *Epidemiol Rev* 1986;**8:**117–28.

[13] Okun DA. The value of water supply and sanitation in development: an assessment. *Am J Public Health* 1988;**78:**1463–67.

[14] Huttly SRA. The impact of inadequate sanitary conditions on health in developing countries. *World Health Stat Q* 1990;**43:**118–26.

[15] Cairncross S, Blumenthal U, Kolsky P, Moraes L, Tayeh A. The public and domestic domains in the transmission of disease. *Trop Med Int Health* 1996;**1:**27–34.

[16] Balayan MS. Natural hosts of hepatitis A Virus. *Vaccine* 1992; **10(Suppl.1):**S27–S31.

[17] Shapiro CN, Coleman PJ, McQuillan GM, Alter MJ, Margolis HS. Epidemiology of hepatitis A: seroepidemiology and risk groups in the USA. *Vaccine* 1992;**10(Suppl.1):**S59–S62.

[18] Gust ID, Ruff TA. Hepatitis in the tropics. *Med J Aust* 1993;**159:**691–95.

[19] Gil A, González A, Dal-Ré R, Dominguez V, Astasio P, Aguiar L. Detection of antibodies against hepatitis A in blood spots dried on filter paper. Is this a reliable method for epidemiological studies? *Epidemiol Infect* 1997;**118:**189–91.

[20] Ades AE, Nokes DJ. Modelling age- and time-specific incidence from seroprevalence: toxoplasmosis. *Am J Epidemiol* 1993;**137:**1022–34.

[21] Brunet RC, Struchiner CJ. Rate estimation from prevalence information on a simple epidemiologic model for health interventions. *Theor Popul Biol* 1996;**50:**209–26.

[22] Keiding N. Age-specific incidence and prevalence: a statistical perspective. *J R Stat Soc A* 1991;**154:**371–412.

[23] Marschner IC. A method for assessing age-time disease incidence using serial prevalence data. *Biometrics* 1997;**53:**1384–98.

[24] Moore HÁ, De La Cruz E, Vargas-Mendez O. Diarrheal disease studies in Costa Rica: I. plan and methods of investigation. *Am J Public Health* 1966;**56:**276–86.

[25] Esrey AS, Habicht JP, Latham MC, Sisler DG, Casella G. Drinking water source, diarrheal morbidity, and child growth in villages with both traditional and improved water supplies in rural Lesotho, Southern Africa. *Am J Public Health* 1988;**78:**1451–55.

[26] Grummer-Strawn LM. Regression analysis of current-status data: an application to breast-feeding. *J Am Stat Assoc* 1993;**88:**758–65.

[27] Huet S, Bouvier A, Gruet MA, Jolivet E. *Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus Examples*. Berlin: Springer Verlag, 1996.

[28] *S-PLUS 4 Guide to Statistics*, Data Analysis Products Division, Mathsoft, Seattle, 1997.

[29] Groeneboon P, Wellner JA. *Information Bounds and Nonparametric MLE*. Basel: Birkhauser, 1992.

[30] Ramsay JO. Monotone regression splines in action. *Statist Sci* 1988; **3:**425–61.

[31] Breiman L, Friedman J. Estimating optimal transformations for multiple regression and correlation (with discussion). *J Am Stat Assoc* 1985;**80:**580–619.

[32] Hastie TJ, Tibshirani RJ. *Generalized Additive Models.* New York: Chapman & Hall, 1990.

[33] Almeida LM. Soroprevalencia da hepatite A: µm possível parâmetro para mensuração de efeitos de intervenções ambientais sobre a saúde. MSc Thesis, Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro, Rio de Janeiro.

[34] Struchiner CJ, Brunet RC, Halloran ME, Massad E, Azevedo-Neto RS. On the use of state-space models for the evaluation of health interventions. *J Biol Systems* 1995;**3:**851–65.