

Methodology of a nationwide cross-sectional survey of prevalence and epidemiological patterns of hepatitis A, B and C infection in Brazil

Metodologia de um estudo de inquérito nacional da prevalência e de padrões epidemiológicos das hepatites A, B e C no Brasil

Ricardo Arraes de Alencar Ximenes^{1,2}
 Leila Maria Beltrão Pereira^{1,3}
 Celina Maria Turchi Martelli⁴
 Edgar Merchán-Hamann⁵
 Airton Tetelbom Stein⁶
 Gerusa Maria Figueiredo⁷
 Maria Cynthia Braga⁸
 Ulisses Ramos Montarroyos¹
 Leila Melo Brasil⁹
 Marília Dalva Turchi⁴
 José Carlos Ferraz da Fonseca¹⁰
 Maria Luiza Carvalho de Lima¹¹
 Luis Cláudio Arraes de Alencar¹
 Marcelo Costa¹²
 Gabriela Coral⁶
 Regina Celia Moreira¹³
 Maria Regina Alves Cardoso¹⁴

¹ Faculdade de Ciências Médicas de Pernambuco, Universidade de Pernambuco, Recife, Brasil.

² Departamento de Medicina Tropical, Universidade Federal de Pernambuco, Recife, Brasil.

³ Instituto do Fígado de Pernambuco, Universidade de Pernambuco, Recife, Brasil.

⁴ Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, Brasil.

⁵ Faculdade de Ciências da Saúde, Universidade de Brasília, Brasília, Brasil.

⁶ Fundação Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brasil.

⁸ Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Brasil.

⁹ Fundação de Medicina Tropical do Amazonas, Manaus, Brasil.

Other institutions listed at the end of the paper.

Correspondence

M. C. Braga
 Departamento de Parasitologia, Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Av. Moraes Rego s/n, Recife, PE 50670-420, Brasil.
braga@cpqam.fiocruz.br

Abstract

A population-based survey to provide information on the prevalence of hepatitis viral infection and the pattern of risk factors was carried out in the urban population of all Brazilian state capitals and the Federal District, between 2005 and 2009. This paper describes the design and methodology of the study which involved a population aged 5 to 19 for hepatitis A and 10 to 69 for hepatitis B and C. Interviews and blood samples were obtained through household visits. The sample was selected using stratified multi-stage cluster sampling and was drawn with equal probability from each domain of study (region and age-group). Nationwide, 19,280 households and ~31,000 residents were selected. The study is large enough to detect prevalence of viral infection around 0.1% and risk factor assessments within each region. The methodology seems to be a viable way of differentiating between distinct epidemiological patterns of hepatitis A, B and C. These data will be of value for the evaluation of vaccination policies and for the design of control program strategies.

Morbidity Surveys; Prevalence; Hepatitis Viruses

Introduction

Population based serological studies of viral hepatitis have demonstrated the diversity of epidemiological patterns with regard to the risk of acquiring infection related to personal attributes, place and risk distribution over time. This provides crucial information for the national control program. In the USA, two national surveys of hepatitis B and C enabled the number of chronically infected individuals to be estimated and led to a greater understanding of the risk patterns associated with the use of illegal drugs and sexual behaviour over time^{1,2,3,4,5}.

Currently, the official Brazilian surveillance system requires compulsory notification of cases of hepatitis A, B and C⁶. Data from the passive notification-based surveillance system is not sufficiently reliable to estimate infection rates in the general population since a large number of viral infections are asymptomatic and the aetiology of notified symptomatic cases cannot always be confirmed. Thus, the epidemiological patterns of HBV and HCV infections in Brazil have been mainly assessed using data from blood donors, volunteers and special groups or from surveys conducted in restricted geographical areas^{7,8,9,10,11}. These previous studies suggest wide variations in distribution of the disease from one Brazilian region to another, with high hepatitis B infection and co-circulation of Delta virus in the Amazon region¹². However,

the prevalence of hepatitis in the general Brazilian population is unknown and there is no information regarding the exact extent to which the frequency of infection and the pattern of risk factors vary from one Brazilian region to another. Viral hepatitis is the most important cause of liver disease and hepatitis B and C have the potential to evolve into chronic hepatitis, cirrhosis and hepatocellular carcinoma, thus increasing chronic liver disease mortality and the incidence of related hepatocellular carcinoma^{13,14}. The hepatitis B vaccine has proved to be an important tool in the prevention of infection and the World Health Organization (WHO) recommends the inclusion of the hepatitis B vaccine in national immunization programs¹⁵. In addition, targeting unvaccinated adults at risk of hepatitis B infection has also been proposed, in such a way as to cover anyone seeking protection from hepatitis B, even if no specific risk factor has been identified¹⁶. Currently, Brazilian health policy recommends HBV vaccination, which is provided free of charge to infants, children and adolescents, besides the known risk groups¹⁷. With regard to hepatitis A, strategies adopted to deliver vaccination vary according to the severity of the endemic, and universal vaccination in areas of low and intermediate endemicity has been advocated¹⁸. In Brazil, the HAV vaccine is not universally recommended within the National Immunization Program and only a small proportion of the population has access to it. In a partial analysis of the ongoing Brazilian survey, hepatitis A infection was classified as an intermediate endemicity in the Northeast and Central-West regions of Brazil¹⁹.

To evaluate the extent to which the population may be infected is a particularly difficult task in a country like Brazil with a total population of approximately 190 million distributed in a huge territory of 8,511,965km². On the other hand, in Brazil there has been a growing accumulation of experience with regard to the conduct of national surveys. For instance, the health supplement of the *Pesquisa Nacional por Amostra de Domicílio* – PNAD (National Survey by Household Sampling), carried out by the Brazilian Institute of Geography and Statistics (IBGE), is possibly the most comprehensive health survey ever conducted in the Brazilian population aged 14 and over, and other successful examples include the *Pesquisa Nacional sobre Saúde e Nutrição* – PNSN (National Survey on Health and Nutrition) and the *Demographic and Health Survey* (DHS)²⁰. It should be emphasized that an additional challenge is posed by serological surveys due to the logistics involved in blood collection, transportation, storage and testing.

Backed by successful experiences and weighing up the relevance of the information for the planning of control strategies, the Brazilian Ministry of Health has launched a nationwide household survey with a view to estimating the prevalence of hepatitis A, B and C and identifying the most frequent HBV and HCV genotypes among residents of Brazilian state capitals and the Federal District. It also envisages to identify exposure through blood route exposure, sexual behaviour risk factors, use of alcohol and illegal drugs and socioeconomic variables accessed at individual, household and area levels. As a secondary objective, a plan has been drawn up to estimate the prevalence of hepatitis D in the North region of the country. This present paper reports the methods used for this first national survey of viral hepatitis infections ever conducted in Latin America.

Material and methods

Study design and population

This study is a population-based household survey by way of which the seroprevalences of the following serological markers for hepatitis A, B and C are currently being estimated: anti-HAV IgG, HBsAg, anti-HBc IgG, anti-HBs, and anti-HCV. Potential biological, epidemiological and socioeconomic risk factors are also being surveyed. The study population comprises all individuals aged 5 to 69, who are resident in private households in all the 26 Brazilian state capitals and in the Federal District during the survey period.

Sampling

- **Domains of study**

Brazil is composed of five different regions: the North, Northeast, Central-West, South and Southeast.

The Northeast region covers nine states (with respective capitals in parentheses): Bahia (Salvador), Sergipe (Aracaju), Alagoas (Maceió), Pernambuco (Recife), Paraíba (João Pessoa), Rio Grande do Norte (Natal), Ceará (Fortaleza), Piauí (Teresina), Maranhão (São Luís). The total population of these capital cities for the 5-9, 10-19 and 20-69 years age groups was 943,033, 2,054,539 and 6,210,495, respectively.

The Central-West region comprises the states of Goiás (Goiânia), Mato Grosso do Sul (Campo Grande) and Mato Grosso (Cuiabá). The population for those same age groups in the central-

western capital cities was 397,740, 866,536 and 2,619,378 inhabitants, respectively.

The Southeast region is comprised of the states of São Paulo (São Paulo), Rio de Janeiro (Rio de Janeiro), Espírito Santo (Vitória) and Minas Gerais (Belo Horizonte), with a population, in the same age groups, of 1,745,661, 3,803,187 and 11,496,336, respectively.

The North region includes the states of Acre (Rio Branco), Amapá (Macapá), Roraima (Boa Vista), Rondônia (Porto Velho), Amazonas (Manaus), Pará (Belém) and Tocantins (Palmas), and the population in age groups 5-9, 10-19 and 20-69 is 350,189, 762,940 and 2,306,227, respectively

The South region covers the states of Rio Grande do Sul (Porto Alegre), Santa Catarina (Florianópolis) and Paraná (Curitiba) with a population for the target age groups of 305,652, 665,908 and 2,012,918, respectively.

The Federal District, capital of Brazil (Brasília city), is located in the Central-West region and the populations for the above mentioned age groups were 182,537, 387,831 and 1,188,413, respectively.

In view of the importance of estimating the prevalence of hepatitis A, B and C infection at national macro-regional level and for different age groups, the population resident in the state capitals and in the Federal District was divided into domains of study as outlined above.

• **Sample size of individuals for each domain of study**

To estimate the prevalence of individuals with hepatitis A, B or C infection, the following formula was used to calculate the sample size for each domain of study:

$$n = \frac{\sigma^2}{\sigma_p^2}$$

where,

n_0 : estimated sample size;

$\sigma^2 = Px(1-P)$: variance of the study characteristic (hepatitis A, B or C) in the population;

σ_p^2 : variance of the sampling distribution of p (prevalence of hepatitis A, B or C).

A confidence level of 95% was used. Tables 1 and 2 present the other parameters used to calculate the estimated sample sizes for the domains of study. In the expectation of a response rate of 80%, the sample sizes were adjusted using $n = n_0 \cdot 0.80$.

The sample sizes were calculated on the basis of expected prevalence of seropositivity for HBsAg, for anti-HCV and for anti-HAV for each macro-region reported in previous studies ^{21,22}. The prevalence of hepatitis A has been estimated, with defined precision, for each macro-region and for the Federal District, for the 5-9 year and 10-19 year age-groups. Similarly, the prevalence of hepatitis B and hepatitis C was estimated, with defined precision, for each macro-region and for the Federal District for the 10-19 and 20-69 year age-groups.

The sampling procedure used in this study was cluster sampling. The observed precision does not therefore depend on the number of individuals in the selected sample alone, but also depends on the degree of homogeneity of the study variable within the clusters (Primary Sampling Units – PSUs), as well as on the average number of individuals selected in each PSU.

Table 1

Parameters used for the sample size calculations for each study domain, estimated sample size (n_0) and estimated adjusted sample size (n) for the study of hepatitis A.

Macro-region	P		σ_p^{2*}		n_0		n	
	5-9 years	10-19 years	5-9 years	10-19 years	5-9 years	10-19 years	5-9 years	10-19 years
North	92.8	92.8	2.7842	2.7842	86	86	108	108
Northeast	76.5	76.5	2.2952	2.2952	342	342	428	428
Southeast	53.6	53.6	1.6082	1.6082	962	962	1,203	1,203
South	51.2	51.2	1.5362	1.5362	1,059	1,059	1,324	1,324
Central-West	76.5	76.5	2.2952	2.2952	342	342	428	428
Federal District	76.5	76.5	2.2952	2.2952	342	342	428	428

* Corresponds to 3% of the proportion.

Table 2

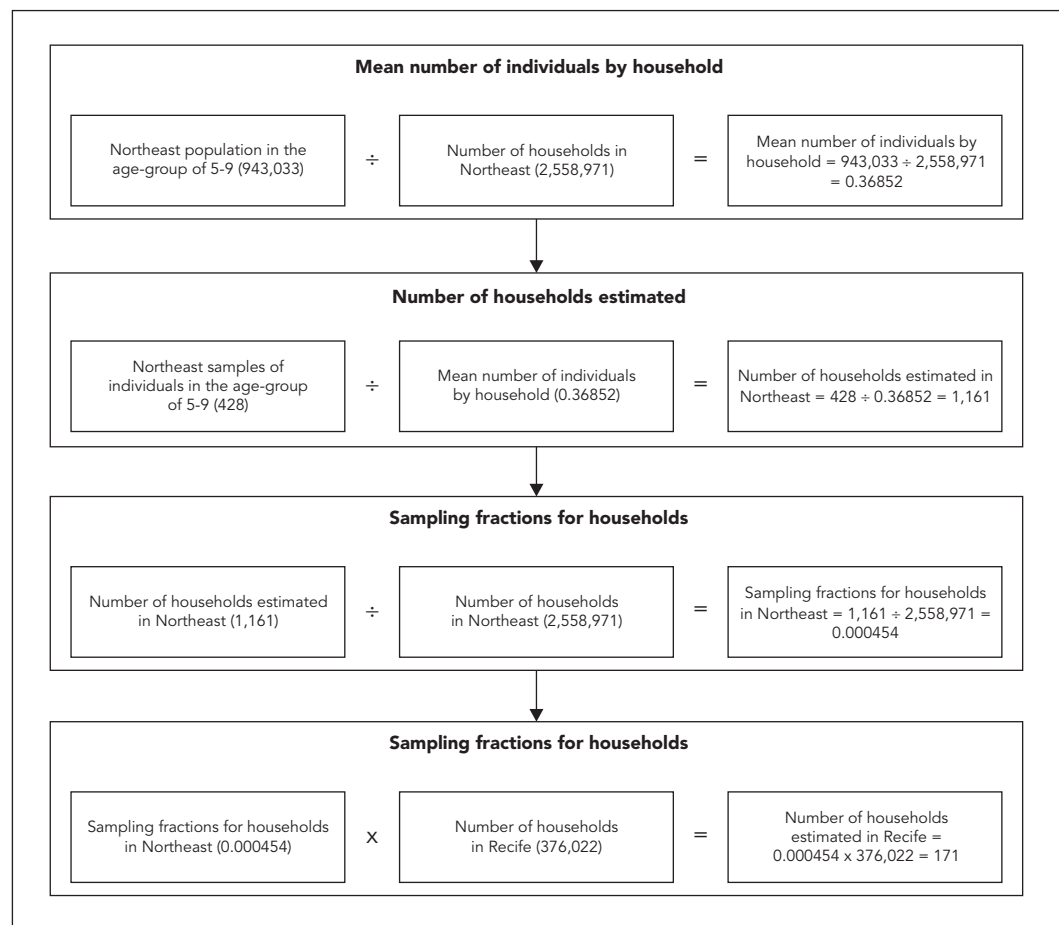
Parameters used for the sample size calculations for each domain of study, estimated sample size (n_0) and estimated adjusted sample size (n) for the study of hepatitis B and C.

Macro-region	P		σ_p^2 *		n_0		n	
	10-19 years	20-69 years	10-19 years	20-69 years	10-19 years	20-69 years	10-19 years	20-69 years
North	0.0212	0.0212	0.00422	0.00422	1,155	1,155	1,443	1,443
North-East	0.0119	0.0119	0.00242	0.00242	2,075	2,075	2,594	2,594
South-East	0.0143	0.0143	0.00292	0.00292	1,723	1,723	2,154	2,154
South	0.0100	0.0100	0.00202	0.00202	2,475	2,475	3,094	3,094
Central-West	0.0104	0.0104	0.00212	0.00212	2,378	2,378	2,973	2,973
Federal District	0.0090	0.0090	0.00272	0.00272	1,124	1,124	1,405	1,405

* Corresponds to 20% of the proportion, with the exception of the Federal District which corresponds to 30%.

Figure 1

Strategy adopted to estimate the number of households to achieve the sample size of individuals – the case of the Northeast region, Brazil.



- **Sample size of households for each domain of study**

The study population in the geographical space was identified on the basis of the fact that the household was occupied at the time the survey was being carried out. A random sample of households was thus selected from each State capital and in the Federal District. All the individuals in the study age-groups resident in these selected households were eligible for inclusion in the samples.

To establish the number of households that should be visited in order to achieve the sample size of individuals, the mean number of individuals in each age-group per household was estimated. The largest sample size of households was that for the lower hepatitis prevalence and less frequent age-group in the population. Taking data from the Northeast, Figure 1 shows the strategy adopted to estimate the number of households to achieve the sample size of individuals.

However, if all individuals from the selected households were enrolled, the sample sizes for the other age-groups would greatly exceed the estimated samples, resulting in excessive cost. For this reason, sub-samples were drawn from the selected households. The results are presented in Table 3.

The samples for each domain of study were allocated among the State capitals using a uniform sampling fraction ($f_i = n_i / N_i$, where f_i : sampling fraction for the domain i ; n_i : sample size for the domain i and N_i : population of the domain i), so that the sample size from each State capital was proportional to the population of that capital. Table 3 shows the sample sizes of households for the study of the hepatitis A, B and C for each Brazilian State capital and for the Federal District, the proportional distribution of individuals by State capital and Federal District expected to be obtained following the visit to the sampled households and the sampling fractions for each domain of study.

- **Sampling procedures**

The study sample was selected using stratified multi-stage cluster sampling based on the IBGE census data for the year 2000. The census tract is the smallest geographic unit available for which socioeconomic data on residents with similar economic characteristics are available and comprises approximately 300 households (approximately equal to 1,000 inhabitants). To improve the efficiency of sampling, the census tracts were, when necessary, either broken up or grouped together in such a way that the variation coefficient

for their sizes should not exceed 10%. The PSUs thus represent a single census tract, a fraction of a census tract or a group of census tracts.

The PSUs were then ordered according to the number of years of schooling of the head of household to guarantee that all the socioeconomic population groups were represented in the sample. This variable was selected because it is strongly correlated with income and provides more reliable information than the later.

In the first stage of sampling, a systematic sample was taken from this ordered list of PSUs with probability proportional to size in each State capital and in the Federal District. In the second stage, for each PSU drawn up in the first stage, block units were also selected with probability proportional to size. The third stage involved systematic sampling of households within each selected block using a list of residential addresses. All current residents of the selected households within the study age-groups were included in the survey.

The samples were drawn with equal probability from each domain of study in three stages. The overall sampling fraction applied to each state capital was, therefore, $f_{hk} = n_{hk} / N_k$, where f_{hk} is the sampling fraction for macro-region k and age-group h , n_{hk} is the sample size in number of households for the macro-region k and age-group h and N_k is the total number of households in the macro-region k .

For each state capital, the first step in the full sampling procedure was to select the largest sample of households using the following equation:

$$f_k = \frac{aM_i}{\sum M_i} \times \frac{bM_{ij}}{M_i} \times \frac{c}{M_{ij}}$$

where,

f_k : overall sampling fraction for the largest sample size in State capital k ;

a : total number of PSUs to be selected in the first stage;

M_i : number of households in the PSU i ;

b : number of secondary sampling units (SSU) to be drawn from each selected PSU;

M_{ij} : number of households in SSU j of PSU i ;

c : number of households to be drawn from each selected SSU.

To select the smaller samples in each State capital, a system of sub-samples were drawn from the initial sample using the following sampling fraction:

$$f_{hk} = f_k \times \frac{d}{n_i}$$

where,

f_{hk} : overall sampling fraction for age-group h in State capital k ;

Table 3

Proportional distribution of the samples of individuals and households by state capitals and Federal District, Brazil, according to age-groups for the study of hepatitis A, B and C.

Region	Hepatitis A				Hepatitis B and C			
	5-9 years ^a		10-19 years ^b		10-19 years ^c		20-69 years ^d	
	Individual	Household	Individual	Household	Individual	Household	Individual	Household
North	112	285	108	130	1,490	1,738	1,443	575
Rio Branco	7	19	7	9	97	121	97	40
Macapá	8	20	8	9	108	122	108	40
Manaus	40	100	40	46	538	618	538	205
Belém	37	92	37	42	490	561	490	185
Porto Velho	10	27	10	12	128	158	128	52
Boa Vista	6	16	6	7	77	92	77	31
Palmas	4	11	4	5	52	66	52	22
Northeast	428	1,167	428	536	2,594	3,722	2,594	1,231
Maceió	34	92	34	42	204	289	204	96
Salvador	103	296	103	136	624	943	624	312
Fortaleza	90	238	90	110	547	762	547	252
São Luís	37	93	37	42	222	293	222	97
João Pessoa	25	69	25	32	153	220	153	73
Recife	60	171	60	78	364	544	364	180
Teresina	30	77	30	35	182	246	182	81
Natal	30	81	30	37	181	257	181	85
Aracaju	19	52	19	24	118	169	118	56
Southeast	1,203	3,792	1,203	1,741	2,154	3,116	2,154	1,031
Vitória	19	59	19	27	33	48	33	16
Belo Horizonte	143	433	143	199	256	356	256	118
Rio de Janeiro	375	1,242	375	570	671	1,021	671	338
São Paulo	666	2,058	666	945	1,193	1,691	1,193	559
South	1,324	4,396	1,324	2,017	3,094	4,719	3,094	1,561
Curitiba	639	2,041	639	937	1,493	2,189	1,493	724
Florianópolis	138	450	138	206	322	483	322	160
Porto Alegre	547	1,905	548	874	1,279	2,047	1,279	677
Central-West	428	1,296	428	595	2,973	4,112	2,973	1,361
Goiânia	209	649	209	297	1,450	2,060	1,450	682
Cuiabá	92	263	92	121	643	835	643	276
Campo Grande	127	384	127	177	880	1,217	880	403
Federal District	428	1,236	428	566	1,405	1,860	1,405	615

Note: Sampling fractions for households for hepatitis A, B and C according to the study age-groups: North (a 0.0003; b 0.0001; c 0.0019; d 0.0006), Northeast (a 0.0004; b 0.0002; c 0.0014; d 0.0005), Southeast (a 0.0002; b 0.0001; c 0.0013; d 0.0004), South (a 0.0043; b 0.0020; c 0.0039; d 0.0013), Central-West (a 0.0011; b 0.0005; c 0.0061; d 0.0020), Federal District (a 0.0022; b 0.0010; c 0.0034; d 0.0011).

f_k : overall sampling fraction for the largest sample in State capital k ;

d : number of households to be drawn from the largest sample initially selected;

n_i : number of households in the largest sample initially selected.

Serological markers and co-variables

For this study, the following serological markers were considered:

i) Hepatitis A: total antibodies for hepatitis A were tested using a commercial enzyme immunoassay kit; a positive test indicating current or past HAV infection (ever infected).

ii) Hepatitis B: antibodies for hepatitis B core antigen (anti-HBc) were tested by means of an enzyme-linked immunoassay. Positive results were considered HBV current or past infection (ever infected). Seropositive anti-HBc samples were also tested for hepatitis B surface antigen (HBsAg). When HBsAg-negative, these samples

were tested for hepatitis B surface antibodies (HBsAb). Both tests were performed using the same technique. Borderline results were retested and indeterminate results were excluded from the analysis.

HBsAg positive samples were also tested for HBV-DNA and genotyped. Nested PCR was carried out using the standard technique. Positive tests for antibodies for hepatitis B core antigen (anti-HBc) indicate current or past HBV infection.

iii) Hepatitis C: all the samples were screened using an automated ELISA. Samples with positive or indeterminate results were also tested using a further third-generation ELISA test. HCV-RNA detection using PCR was carried out on every

sample with positive or indeterminate results. A reverse transcriptase-nested PCR was used to detect the HCV genome.

Information on co-variables at individual and household level was obtained using a questionnaire and area level variables retrieved from a subset of the IBGE census data (Table 4).

Laboratory methods

During the household visit, blood samples were collected, transported and stored according to standard procedures. Total antibodies for hepatitis A virus were detected using a commercial enzyme immunoassay kit (HAVAB EIA; Abbott

Table 4

Variables studied at individual, household and area levels.

Individual level	Household level	Area level
Socio-demographic variables	Home ownership	Illiterate head of the household
Gender	Water supply	Illiterate female head of the household
Age	Pipe water inside the house	
Literacy	Regular water supply	Mean level of schooling of head of household (years)
Years of schooling	Sewage disposal	
Paid work last week	Solid waste disposal	15-19 years illiterate individuals
Reported vaccination status		Female head of the household
Blood route transmission		Households with water supply
Hospitalization		Households with waste collection
Blood transfusion		Individuals per household
Surgery		
Dental treatment		
Health care related job		
Tattoo		
Body piercing		
Sharing of cutting objects		
Sexual behaviour		
Initiated sexual life		
Homosexual partner		
Another sexual partner besides current partner		
Previous STD		
Current partner had hepatitis		
Previous sexual relation with known HBV carrier		
Drug use-related factors		
Ever use of smoked drugs		
Ever use of inhaled drugs		
Ever of sniffed drugs		
Ever use of injected drugs		
Previous use of glass syringe		
Alcohol consumption		

Laboratories, Abbott Park, U.S.A.) at central public health laboratories. Laboratory tests were validated and quality control was monitored. Blood samples were tested for hepatitis B core antigen antibodies (anti-HBc) using enzyme-linked immunoassay – ELISA (AxSYM, Abbott Laboratories) at central public health laboratories. Seropositive anti-HBc samples were tested for hepatitis B surface antigen (HBsAg). Both tests were performed using the same technique (AxSYM, Abbott Laboratories). Borderline results were retested and indeterminate results were excluded from the analyses. All samples testing positive for HBV in the North region were also tested for the HDV marker (anti-HD IgG).

To study the prevalence of HCV infection in this population, all the samples were screened by an automated ELISA – Microparticle Enzyme Immunoassay (MEIA) – (AxSYM, Abbott Diagnostic, Germany). Samples with positive or indeterminate results were also tested by another third-generation ELISA test (Diasorin Saluggia Mark, Vercelli, Italy).

The HBV, HDV and HCV positive samples were genotyped.

HBsAg positive samples were also tested for HBV-DNA and genotyped. Nested PCR was carried out as described by Kaneko et al.²³, with some modifications. The S region was amplified according to a standard technique²⁴ for identification of the HBV genotype. Sequencing reactions for characterization of viral strains were performed using PCR according to the method described elsewhere²⁵. Genotyping was analyzed by comparison of the sequences obtained with other known sequences from different HBV genotypes deposited in the GeneBank, using the Edit Seq and MegAlign softwares of the DNASTAR package (Lasergene Inc., U.S.A.).

PCR HCV-RNA detection was carried out on every sample with positive or indeterminate results. A reverse transcriptase-nested PCR was used to amplify a fragment of the 5'UTR of the HCV genome. Any sample with a positive result in at least two out of three tests (PCR and both ELISA) was considered HCV positive.

Data collection: logistics

This survey was a collaborative undertaking involving universities, State Health Secretaries and state-level referral laboratories. The research team responsible included a group of epidemiologists, clinicians, laboratory staff and fieldworkers from each setting, along with a regional and national coordinating group. The Brazilian Ministry of Health and Pan-American Health Organization (PAHO) supported the study. In order

to assure community compliance, local leaders were contacted and information about the survey was provided by local newspapers, radio, and television prior to commencing fieldwork in each locality.

Epidemiologists and clinicians were present at the three levels of coordination, while laboratory and fieldwork coordinators participated at the national and local levels. The training of the regional and local epidemiologist and clinical coordinators, and the local laboratory coordinator was conducted by the national team.

The local teams also included interviewers, phlebotomists, data entry clerks and drivers.

The local epidemiology coordinator was responsible for standardizing and training the interviewers in all procedures involved in data collection and the clinician coordinator standardized and trained the phlebotomists for blood collection. The local clinician was also responsible for the delivery of the laboratory results to the participants, inviting those with positive results to visit the referral service for orientation and, when necessary, start treatment. These coordinators were required to send weekly reports on the development of the study to the regional coordinating group.

The local fieldwork coordinator organized the daily activities and distributed to the interviewers the household addresses, with their respective maps, stating the age-group to be interviewed in each household.

The interviewers together with the phlebotomists were instructed to visit the households and, after certifying that one or more of the residents fell into the required age-group, to explain the purposes of the study and obtain a signature indicating informed consent.

In case of refusal, a second attempt was made by the fieldwork coordinator and, if the refusal was reaffirmed, the household was visited by the clinician. When no-one appeared to be at home in a household, at least two more visits were made on different occasions.

Fieldwork was carried out according to the agreed schedule, starting with the Northeast, Central-West and Federal District (2005-2006); followed by the South and Southeast regions (2007-2008), and, finally, the North region, to be concluded in 2009.

Quality assurance and quality control

Several efforts were made to standardize all procedures in the 26 State capitals and the Federal District. Manuals were written with instructions on how to conduct the fieldwork, interviewing, collecting blood samples, processing and storing

the blood collected and on sampling. Questionnaires were piloted before the final version was written and a sub-sample of households was revisited to check the questionnaire application. For the serological tests, the same commercial kit and the same machine were used in every State capital. Quality control at a referral laboratory was carried out for all laboratory results testing positive for hepatitis B and C, for a sample of 10% of the results testing negative and of those results which tested positive for hepatitis A.

Either the national and/or the regional coordinators supervised the initial stages of the fieldwork in every State capital. The activities carried out in all localities were monitored by the national team, which received weekly reports on the fieldwork and laboratory activities.

Analysis guidelines

Data was double entered for validation and consistency checking before statistical analysis was performed. The overall prevalence of seropositivity with a 95% confidence interval (95%CI) for hepatitis A, B and C infection was calculated (corrected for the design effect) for each macro-region and the Federal District by age-group. For some analyses, macro-regions and/or age-groups were collapsed and, in these cases, weighting was applied, because different sampling fractions had been adopted for the study domains.

The age-specific distribution of the population susceptible to each type of hepatitis infection was estimated using a simple catalytic model to estimate the force of infection.

The putative risk factors for each hepatitis infection were initially investigated using bivariate analyses, separately for each age-group. To estimate the magnitude of these associations, odds ratios with their respective 95%CI were used.

Random-effect multiple logistic regression models were designed to adjust for the effect of each variable on the others.

A multilevel model was proposed for assessing risk at individual, household and area level using Generalized Linear and Latent Mixed Models (GLLAMM) on the Stata 9.2 software (Stata Corp., College Station, U.S.A.).

Ethical aspects

According to *Resolution no. 196/96* of the Brazilian Ministry of Health, this project was submitted to Ethics Committees in all Brazilian regions. All individuals were informed regarding the objectives of the study and signed informed consent was requested from all participants. The results of the laboratory tests were returned confidentially

to the participants. The negative results for all types of hepatitis or positive results for the anti-HAV IgG and/or total anti-HBc and/or anti-HBs (which indicate immunity to hepatitis A and B, respectively) were posted by mail. The results for the positive cases of hepatitis B (positive HBsAg) and hepatitis C (positive anti-HCV) were delivered in person by the coordinating clinician. This clinician was responsible for performing a thorough evaluation of these positive patients and, if recommended, they were referred for specific treatment at the relevant referral service in each State capital or the Federal District. HBV seronegative individuals were referred to the public health services for vaccination.

Strengths and weaknesses of the study

Interpretations of the results of epidemiological studies should be considered in light of the role of chance, i.e., sampling variation, and the possibility of selection or information bias and confounding factors.

The target population consisted of individuals aged from five to sixty-nine years, resident in private households in all Brazilian State capitals and in the Federal District during the survey period. The study sample was randomly selected and adequate parameters were used to estimate its size. For the analysis, corrections were made for the design effect. The results can be extrapolated for the all the State capitals and the Federal District taken as a whole.

Most of the capital cities (19 out of 26) included in the study belong to Metropolitan Regions or to special Integrated Development Regions that include other municipalities. The latter may have poorer social and economic standards, which may in turn influence the occurrence of viral hepatitis. Therefore, our findings may not reflect the situation of the metropolitan regions and should only be taken to apply to the capital cities.

Selection bias was minimized by the following strategies:

- i) The sampling framework used for selecting census tracts and households was provided by the IBGE, which is the national institute responsible for the Brazilian census;
- ii) To avoid refusal to participate in the study, information on the survey was provided through the mass media, especially on local radio and television. Rigorous training of the local survey teams also facilitated their interaction with the study population;
- iii) Comparability of infected and non-infected individuals was assured as they originated from the same population.

Information bias may occur when there is an error in classifying the individuals, but its consequences are different, depending on whether they occur independently – non-differential or not-differential – in the two axes of classification (exposure and outcome) of the individuals²⁶.

Information bias was minimized as follows:

- i) At the time of the interview, neither the interviewers nor the interviewees knew the results of the laboratory tests, as blood samples had not yet been collected. In addition, the laboratory tests were performed using standardized techniques and the laboratory personnel did not have access to the questionnaires. There was thus no way in which the serologic results could influence the application of the questionnaires or *vice versa*, making the possibility of a differential classification error very low. Whereas a differential classification error may lead to overestimation or underestimation of an effect, a non-differential error poses a less significant threat to validity, since bias occurs in a predictable direction, namely, in the direction of the null hypothesis. It is thus possible that the associations found in this study have been underestimated^{26,27,28}. Special care was taken to avoid possible bias introduced by the interviewer. Rigorous training was conducted in each State capital and a pilot of the questionnaire was applied in three capitals situated in different regions of the country (Recife, Manaus and Porto Alegre). Another potential limitation inherent in any household survey, such as this, is the possibility of underreporting or incorrect answers being given to sensitive questions, including those relating to current or former illegal drug use and sexual behaviour. For these reasons, risk assessment for HBV and HCV infections related to exposure to these should be interpreted with caution;
- ii) The codification of the questionnaires was performed after the interview, by the local coordinators;
- iii) With regard to the errors inherent in the laboratory procedures, it should be stressed that the commercial kits from just one manufacturer with a high degree of sensitivity and specificity were used. The same model of machinery was used in all State capitals. In addition, quality control was carried out at a referral laboratory for all positive laboratory results for hepatitis B and C and in a sample of 10% of those testing negative and of those positive for hepatitis A.

Finally, since this is a cross-sectional study, temporal ambiguity and survival bias may occur.

Cross-sectional studies are usually explained as different from follow-up and case-control designs but they can be understood as the case-control analogue of the general population follow-up

study, the cases being a prevalence series of individuals with the disease and the controls being the remainder of the study population²⁶. Odds ratios are often used in the statistical analysis of binary outcomes, even when the disease is not rare. In cross-sectional studies, when the disease is rare, the estimates obtained with the odds ratio are similar to those that would be obtained with the prevalence rate. When the disease is common the value of the odds ratio is numerically larger than the prevalence rate, but it is still a good estimate of the odds ratio of the population.

So far as the possibility of distortion of the measures of association by confounding variables was concerned, all variables were adjusted for age. A multivariate analysis was conducted at different levels (individual, household and census tract) and, in some instances, a multilevel analysis was carried out.

Conclusion

The approach used is a viable way of conducting surveys in countries where regular census data is collected. The proposed methodology is feasibly capable of differentiating various epidemiological patterns of hepatitis A, B and C. It provides population-based information on the prevalence of behavioural and environmental exposure, as well as of viral markers. The methodology adopted is suitable to be adapted to large epidemiological surveys with different goals.

Information provided by estimates of hepatitis A prevalence may prove useful base-line data for monitoring interventions regarding water supply and sewage disposal in different localities. In addition, the data are suitable for conducting cost-benefit studies for the implementation of HAV immunization. With regard to hepatitis B and C, it is possible to estimate the number of individuals who will require special and complex means of diagnosis and medical care (biopsies, surgeries, transplants, cancer treatment, and antiviral and immunological therapy) to be provided by the Unified National Health System. The availability of such data may contribute to the identification of priority groups for HBV and HCV screening. It also enables the health authorities to follow-up asymptomatic individuals identified outside the health care system.

Resumo

Um inquérito de base populacional foi conduzido na população urbana de todas as capitais e do Distrito Federal no Brasil para fornecer informações sobre a prevalência de hepatites virais e fatores de risco, entre 2005 e 2009. Este artigo descreve o delineamento e a metodologia do estudo que envolveu a população com idade entre 5 e 19 anos para hepatite A e 10 a 69 anos para hepatite B e C. As entrevistas e amostras de sangue foram obtidas através de visitas domiciliares e a amostra selecionada a partir de uma amostragem estratificada em múltiplos estágios (por conglomerado) com igual probabilidade para cada domínio de estudo (região e faixa etária). Nacionalmente, 19.280 residências e ~31.000 indivíduos foram selecionados. O tamanho da amostra foi suficiente para detectar uma prevalência em torno de 0,1% e para avaliar os fatores de risco por região. A metodologia apresentou-se viável para distinguir entre diferentes padrões epidemiológicos da hepatite A, B e C. Estes dados serão de valia para a avaliação das políticas de vacinação e para o desenho de estratégias de controle.

Inquéritos de Morbidade; Prevalência; Vírus da Hepatite

Contributors

R. A. A. Ximenes contributed on the conception and design, analysis and interpretation, drafting and revising. L. M. B. Pereira contributed on the conception and design of the study, and drafting and revising the manuscript. C. M. T. Martelli and E. Merchán-Hamann participated in the acquisition, analysis and interpretation of the data, and drafting and revising the manuscript. A. T. Stein contributed in the acquisition and interpretation of data, and revising and final approval of the manuscript. G. M Figueiredo contributed to the interpretation of data, drafting and revising the manuscript. M. C. Bra-

ga participated in the acquisition of data, drafting and revising the article. U. R. Montarroyos participated in the analysis and interpretation of data, and revising the article. L. M. Brasil contributed in the conception and design of the study and revising the final version of the manuscript. M. D. Turchi participated in the conception and design, acquisition of data, interpretation of data and revising the article. J. C. F. Fonseca contributed to the conception and study design and revising the article. M. L. C. Lima, L. C. A. Alencar and M. Costa participated in the conception and design, acquisition of data, revising the article. G. Coral contributed to the conception and design, acquisition of data, interpretation of data and revising the article. R. C. Moreira contributed to the conception and design, interpretation of data and revising the article. M. R. A. Cardoso contributed with the conception and design, analysis and interpretation of data; drafting, revising and final approval of the manuscript.

Acknowledgments

We are grateful for the financial support of the Brazilian Ministry for Health, the Pan American Health Organization and for administrative support from the University of Pernambuco Foundation. The authors were partially supported by National Research Council (CNPq; scholarship 307963/2004-7 to C.M.T.M. and 300917/2006-6 to R.A.A.X. and 303049/2007-3 to M.R.A.C.).

Institutions

- ¹⁰ Universidade Federal do Amazonas, Manaus, Brasil.
- ¹¹ Departamento Medicina Social, Universidade Federal de Pernambuco, Recife, Brasil.
- ¹² Hospital de Base do Distrito Federal, Brasília, Brasil.
- ¹³ Instituto Adolfo Lutz, São Paulo, Brasil.
- ¹⁴ Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, Brasil.

References

1. Coleman PJ, McQuillan GM, Moyer LA, Lambert SB, Margolis HS. Incidence of hepatitis B virus infection in the United States, 1976-1994: estimates from the National Health and Nutrition Examination Surveys. *J Infect Dis* 1998; 178:954-9.
2. McQuillan GM, Kruszon-Moran D, Kottiri BJ, Curtin LR, Lucas JW, Kington RS. Racial and ethnic differences in the seroprevalence of 6 infectious diseases in the United States: data from NHANES III, 1988-1994. *Am J Public Health* 2004; 94:1952-8.
3. Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999; 341:556-62.
4. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006; 144:705-14.
5. Stern RK, Hagan H, Lelutiu-Weinberger C, Des Jarlais D, Scheinmann R, Strauss S, et al. The HCV Synthesis Project: scope, methodology, and preliminary results. *BMC Med Res Methodol* 2008; 8:62.
6. Ministério da Saúde. Guia de vigilância epidemiológica. Brasília: Ministério da Saúde; 2005.
7. Souto FJ, Fontes CJ, Gaspar AM. Prevalence of hepatitis B and C virus markers among malaria-exposed gold miners in Brazilian Amazon. *Mem Inst Oswaldo Cruz* 2001; 96:751-5.
8. Nascimento MC, Mayaud P, Sabino EC, Torres KL, Franceschi S. Prevalence of hepatitis B and C serological markers among first-time blood donors in Brazil: a multi-center serosurvey. *J Med Virol* 2008; 80:53-7.
9. Zarife MA, Silva LK, Silva MB, Lopes, GB, Barreto ML, Teixeira MG, et al. Prevalence of hepatitis C virus infection in north-eastern Brazil: a population-based study. *Trans R Soc Trop Med Hyg* 2006; 100:663-8.
10. Tapia-Conyer R, Santos JI, Cavalcanti AM, Urdaneta E, Rivera L, Manterola A, et al. Hepatitis A in Latin America: a changing epidemiologic pattern. *Am J Trop Med Hyg* 1999; 61:825-9.
11. Almeida LM, Werneck GL, Cairncross S, Coeli CM, Costa MC, Coletty PE. The epidemiology of hepatitis A in Rio de Janeiro: environmental and domestic risk factors. *Epidemiol Infect* 2001; 127:327-33.
12. Viana S, Parana R, Moreira RC, Compri AP, Macedo V. High prevalence of hepatitis B virus and hepatitis D virus in the western Brazilian Amazon. *Am J Trop Med Hyg* 2005; 73:808-14.
13. Shepard CW, Simard EP, Finelli L, Fiore AE, Bell BP. Hepatitis B virus infection: epidemiology and vaccination. *Epidemiol Rev* 2006; 28:112-25.
14. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis* 2005; 5:558-67.
15. World Health Organization. Hepatitis B vaccine. Geneva: World Health Organization; 2005.
16. Mast EE, Weinbaum CM, Fiore AE, Alter MJ, Bell BP, Finelli L, et al. A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Advisory Committee on Immunization Practices (ACIP) Part II: immunization of adults. *MMWR Recomm Rep* 2006; 55(RR-16):1-33.
17. Secretaria de Saúde de São Paulo. Hepatitis B vaccine. *Rev Saúde Pública* 2006; 40:1137-40.
18. Jacobsen KH, Koopman JS. Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiol Infect* 2004; 132:1005-22.
19. de Alencar Ximenes RA, Martelli CM, Merchán-Hamann E, Montarroyos UR, Braga MC, de Lima ML, et al. Multilevel analysis of hepatitis A infection in children and adolescents: a household survey in the Northeast and Central-west regions of Brazil. *Int J Epidemiol* 2008; 37:852-61.
20. Malta DC, Leal MC, Costa MFL, Morais Neto OL. Inquéritos Nacionais de Saúde: experiência acumulada e proposta para o inquérito de saúde brasileiro. *Rev Bras Epidemiol* 2008; 11 Suppl 1:159-67.
21. Lyra LG, Damasceno AP, Cotrim P, Mota E, Silva L. Prevalence of antibody to hepatitis B virus in an urban population of Northeast Brazil. *Rev Inst Med Trop São Paulo* 1986; 28:406-9.
22. Fonseca JC. Epidemiologia da infecção pelo vírus da hepatite C no Brasil. *GED Gastroenterol Endosc Dig* 1999; 18 Suppl 1:S3-S8.
23. Kaneko S, Miller RH, Feinstone SM, Unoura M, Kobayashi K, Hattori N, et al. Detection of serum hepatitis B virus DNA in patients with chronic hepatitis using the polymerase chain reaction assay. *Proc Natl Acad Sci U S A* 1989; 86:312-6.
24. Sitnik R, Pinho JR, Bertolini DA, Bernardini AP, Da Silva LC, Carrilho FJ. Hepatitis B virus genotypes and precore and core mutants in Brazilian patients. *J Clin Microbiol* 2004; 42:2455-60.
25. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977; 74:5463-7.
26. Rothman K, Greenland S. *Modern epidemiology*. 2nd Ed. Philadelphia: Lippincott-Raven; 1998.
27. Bross I. Misclassification in 2 x 2 tables. *Biometrics* 1954; 10:478-86.
28. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; 105:488-95.

Submitted on 13/Nov/2009

Final version resubmitted on 01/Jun/2010

Approved on 10/Jun/2010