

Estimativas do número de casos de aids no Brasil, corrigidas pelo atraso de notificação

AIDS cases estimates in Brazil corrected for reporting delay

Maria Tereza S. Barbosa

Departamento de Matemática e Estatística /UNI-RIO

e-mail: tereza@malaria.procc.fiocruz.br

Claudio José Struchiner

Financiado parcialmente pelo CNPq, PRONEX/FINEP/FAPERJ

Instituto de Medicina Social/UERJ e Escola Nacional de Saúde Pública/
FIOCRUZ

e-mail: stru@malaria.procc.fiocruz.br

Endereço para correspondência: Av Brasil 4365, antiga residência oficial –
Manguinhos - 21045-900 Rio de Janeiro, RJ – Brasil

Auxílio financeiro

Trabalho financiado pelo Programa Nacional DST/AIDS através do contrato
nº 052/95 firmado com a Fundação Oswaldo Cruz e pela CAPES.

Resumo

Dois modelos estatísticos são propostos para estimar os casos de aids no Brasil já diagnosticados e ainda não notificados até o primeiro semestre de 1996, para as diversas categorias de exposição. O primeiro considerou a distribuição do atraso de notificação como uma função de sobrevida, com dados censurados à direita, que foi estimada a partir do método de Kaplan-Meyer. O segundo ajusta uma regressão de Poisson a uma tabela de contingência onde cada célula representa o número de casos diagnosticados no semestre e notificados com um determinado atraso. Precedendo o ajuste desta regressão, utilizou-se um modelo aditivo generalizado para identificar uma função que ajuste melhor que um modelo linear as relações funcionais em questão. As estimativas obtidas foram comparadas com as notificações ocorridas entre o segundo semestre de 1996 e o primeiro semestre de 1997. O confronto entre as estimativas fornecidas pela regressão de Poisson e o número de casos oficialmente notificados sugere uma possível mudança no comportamento do padrão de notificação. O número de casos oficialmente notificados encontra-se bem próximo à epidemia real, provavelmente devido à distribuição gratuita de medicamentos. As estimativas do modelo de sobrevida, que não leva em consideração mudanças no nível da epidemia nos diversos semestres de diagnóstico, tendem a uma subestimação das categorias que estão em crescimento.

Palavras-chave: Modelos estatísticos. Análise de regressão. Síndrome de imunodeficiência adquirida, epidemiologia. Brasil.

Abstract

We estimated the number of AIDS cases in Brazil from official statistics, after correcting for delays in reporting, by the use of two statistical models. First, we fitted a Kaplan-Meier survival model with right censoring to the distribution of reporting delays. The second approach is based on a Poisson regression model which was fitted to the contingency table having as one dimension the number of AIDS cases diagnosed per calendar time and the other dimension the amount of delay in reporting. Before the latter approach, we explored, by the use of generalized additive models (GAM), the search for convenient smoothing functions that fit the observed functional relationships in question. The estimates we obtained were then compared to the cases actually reported in the second semester of 1996 and in the first semester of 1997. Comparing the observed number of cases and the predicted figures by Poisson regression, a new reporting pattern seems to start to take place. The official reported number of AIDS cases is very close to the actual epidemic, probably due to government policies of free distribution of medication. Survival models that do not account for calendar time changes in the number of diagnosed AIDS cases in each semester seem to underestimate the behavior of the disease in exposure categories where transmission is increasing.

Keywords: Models, statistical. Regression analysis. Acquired Immunodeficiency Syndrome, epidemiology. Brazil.

Introdução

O longo e variável tempo de incubação da infecção pelo HIV e a indeterminação do momento da infecção são fatores que dificultam o acompanhamento e análise dos dados relacionados à epidemia da aids. A obtenção destes dados incluem desde estudos de soroprevalência e soroconversão, que acompanham seguimentos específicos da população, até aqueles que se baseiam nos casos de aids notificados oficialmente.

Os estudos de soroprevalência no Brasil, devido às suas dificuldades, ainda são pouco representativos, e os registros oficiais têm sido, portanto, um dos mais importantes instrumentos para o monitoramento da epidemia. A partir deles é possível identificar mudanças relevantes no comportamento da epidemia, tais como: sua interiorização, o crescimento do contingente de mulheres contaminadas e conseqüentemente da transmissão perinatal, e mudanças quanto à faixa etária mais atingida em função do crescimento da contaminação de usuários de drogas. No entanto, a estimativa da real magnitude da epidemia depende do tempo decorrido entre o diagnóstico do caso e sua notificação, e do número de casos que não foram e não serão notificados. Nos Estados Unidos, em 1994, 50% dos casos eram notificados com menos de dois meses após o diagnóstico e 95% deles dentro de dois anos¹. Já no Brasil, em 1996, avaliou-se que, após seis meses do momento do diagnóstico, apenas 50% tinham sido notificados e que para atingir os 95% dos casos eram necessários três a quatro anos. Para que este tempo diminua, os envolvidos de alguma forma com a demora entre o diagnóstico do caso e a sua contabilidade oficial devem ser sensibilizados quanto aos efeitos que este atraso pode ter nas medidas de controle a serem tomadas para o combate da epidemia. Portanto, o correto preenchimento de uma ficha, o rápido envio da notificação dos casos, um cuidadoso processamento das informações e uma análise dos dados disponíveis, cada vez mais eficiente e ligada à realidade, estão entre os

maiores desafios que se apresentam a qualquer Sistema de Vigilância Epidemiológica.

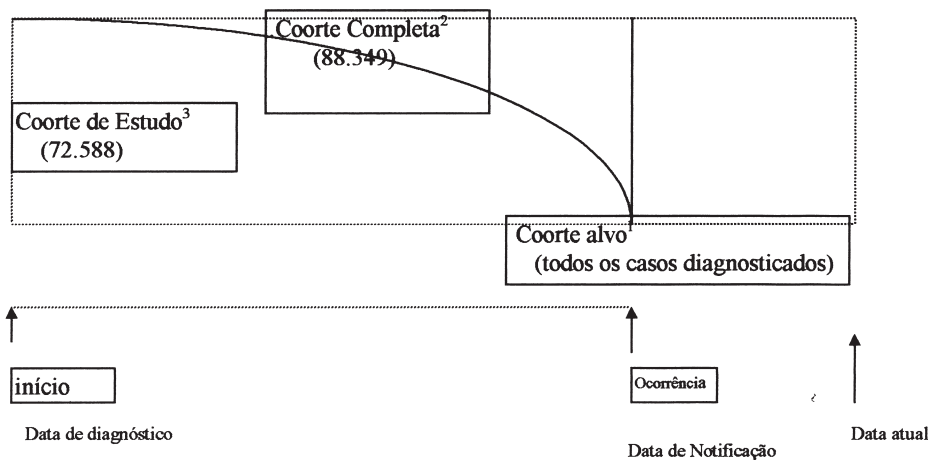
Para se estimar o número de infectados pelo HIV, ou para se dimensionarem corretamente as necessidades de medicamentos para o tratamento da aids, é necessário que se estude o comportamento estatístico do atraso da notificação, e que se proponham mecanismos que estimem o número de diagnosticados e ainda não notificados. Desde o início da epidemia vários métodos têm sido propostos com o objetivo de estimar o número de casos diagnosticados em cada tempo-calendário. Ao tratarem o atraso da notificação em categorias discretas, estes métodos ajustam um modelo loglinear a uma tabela de contingência²⁻⁵. Alternativamente, sob condições de não-estacionaridade do atraso de notificação, o modelo log-logístico tem sido proposto^{6,7}. Ao considerar o atraso da notificação como uma variável contínua, a estimativa da distribuição do atraso de notificação pode ser formulada como uma questão de análise de sobrevivência, utilizando técnicas de tabela de vida com dados censurados à direita⁷. A estreita relação entre estas abordagens foi tratada por Clayton⁸ que ressaltou os aspectos que fazem com que todas elas possam ser consideradas extensões da modelagem de Poisson, diferindo entre si no grau de complexidade considerada na estrutura dos dados.

No Brasil, para corrigir os dados notificados em São Paulo até 1992, Menezes⁹ considerou que os casos diagnosticados e não notificados em cada período de diagnóstico seguiam uma distribuição multinomial incompleta. Barbosa e Struchiner¹⁰ repetiram esta abordagem, mas supuseram que os casos de aids diagnosticados no Brasil até 1992 já haviam sido totalmente notificados, utilizando suas estimativas para corrigir os dados notificados por região brasileira. Bandeira e Silva¹¹ corrigiram os dados notificados na cidade de São Paulo, utilizando uma abordagem não paramétrica. Se um modelo é sempre uma descrição simplificada da realidade, e por isto mes-

mo não existe o melhor, a comparação e avaliação dos resultados obtidos a partir de várias propostas permitem que se tenha mais confiança na ordem de grandeza das estimativas. Este trabalho busca, a partir do estudo do atraso da notificação, entender melhor as diferenças no comportamento da epidemia, existentes entre as diversas categorias de exposição. Serão comparadas as estimativas realizadas quando o atraso da notificação é medido em número de dias e a correção dos casos diagnosticados, realizada a partir da estimativa de uma função de sobrevivência, com as obtidas por uma regressão de Poisson, repetindo a já realizada¹². As estimativas foram obtidas separadamente para cada categoria, sendo avaliadas a partir de seu confronto com os dados notificados até junho de 1997.

Metodologia

Muitos aspectos da epidemia de aids podem ser esclarecidos a partir de informações estratificadas pela categoria principal de exposição, a respeito das datas dos vários eventos envolvidos. As datas que serão utilizadas referem-se ao dia em que o médico considerou que o paciente apresentava o conjunto de sinais e sintomas definidores do diagnóstico de aids, e ao dia em que este caso passou a fazer parte das estatísticas oficiais nacionais da epidemia. Os dados utilizados referem-se aos casos notificados entre o início do primeiro semestre de 1986 e o final do primeiro semestre de 1996. Os indivíduos já diagnosticados com aids são tratados como uma "coorte" que se constituiu a partir da data de diagnóstico e que só começa a ser observada a partir da data da notificação. Assim, embora os indivíduos entrem na coorte em diferentes tempos-calendários, a data de diagnóstico será definida como o tempo zero e a diferença entre as duas datas, medida em dias ou em semestres, será usada como variável resposta. Assim, em cada tempo-calendário, só podem ser observados os elementos em que esta diferença for menor do que o



1

Figura 1 - Descrição das coortes utilizadas na análise
Figure 1 - Different types of cohorts used in the analysis

¹ coorte alvo - todos os indivíduos com diagnóstico de aids até setembro de 1996

¹ target cohort - all cases of AIDS diagnosed as of September 1996

² coorte completa - todos os indivíduos pertencentes à coorte-alvo já notificados até setembro de 1996

² complete cohort - AIDS cases belonging to the target cohort and already reported as of September 1996

³ coorte de estudo - todos os indivíduos pertencentes à coorte completa cuja data de notificação esteja disponível

³ study cohort - AIDS cases belonging to the complete cohort to whom reporting dates are available.

intervalo total observado. Os outros serão considerados censurados.

Dos 88.349 casos de aids notificados até setembro de 1996, 72.588 possuíam a data da notificação ou alguma “proxy” desta data. Considerando, então, que a coorte alvo, constituída por todos os casos já diagnosticados, não foi totalmente observada, e que a coorte completa e observável seria a referente aos 88.349 casos notificados, definiu-se como coorte de estudo (Figura 1) aquela em que estavam presentes ambas as datas.

Os casos notificados serão corrigidos a partir dos resultados obtidos por uma análise de sobrevivência e por uma regressão de Poisson. Para o modelo de análise de sobrevivência, será definido para cada caso já notificado na data de estudo, o número de dias decorridos entre a data de diagnóstico e a data da notificação. Já o número de dias

entre a data de diagnóstico e a data de estudo será a variável a ser utilizada para os dados censurados. Para a regressão de Poisson, os casos serão distribuídos numa tabela de contingência formada pelo cruzamento do semestre de diagnóstico pelo número de semestres em atraso.

Análise de Sobrevida

A função de sobrevivência será utilizada para fazer previsões periódicas a respeito do número de casos diagnosticados, mas ainda não notificados em cada período e em cada categoria de exposição. O momento de estudo foi definido como sendo 31 de dezembro de 1994 e o objetivo, estudar a distribuição do tempo entre o diagnóstico e a notificação dos 65.223 casos diagnosticados antes desta data e que estavam notificados até 30/06/1996. Este tempo foi medido em

Tabela 1- Número de casos de aids notificados por período de atraso e semestre de diagnóstico.**Table 1 -** AIDS cases reporting by calendar time and delay.

Semestre de Diagnóstico	Semestres atrasados									Total
	Nenhum	Um	Dois	Três	Quatro	Cinco	Seis	Sete	Oito ou mais	
1-86	137	21	12	15	15	36	18	21	134	409
2-86	196	32	11	19	58	64	19	40	257	696
1-87	207	37	37	140	101	71	54	30	278	955
2-87	287	109	185	258	120	89	58	53	385	1.544
1-88	387	322	328	244	88	78	59	48	292	1.846
2-88	1.126	437	210	92	66	58	64	57	180	2.290
1-89	1.446	393	144	68	89	76	91	50	148	2.505
2-89	1.820	356	219	119	121	138	39	53	217	3.082
1-90	1.705	573	332	236	218	90	67	60	133	3.414
2-90	2.295	775	372	312	125	110	98	90	138	4.315
1-91	2.405	897	671	279	147	147	105	58	79	4.788
2-91	2.783	1.476	477	275	195	132	98	53	57	5.546
1-92	3.385	1.525	586	314	165	127	82	29	32	6.245
2-92	2.791	2.322	594	302	235	34	37	116		6.431
1-93	3.521	2.318	623	291	122	52	42			6.969
2-93	3.384	2.830	694	234	100	61				7.303
1-94	3.082	3.359	545	119	88					7.193
2-94	4.221	2.332	280	264						7.097
1-95	4.372	1.578	712							6.662
2-95	2.099	2.058								4.157
1-96	1.064									1.064
Total	42.713	23.750	7.032	3.581	2.053	1.363	931	758	2.330	84.511

número de dias e os 15.263 casos cujas datas de notificação ocorreram depois de dezembro de 1994 foram considerados censurados. Se T é o número de dias em atraso, a probabilidade da notificação ocorrer depois de t é definida como $S(t) = P(T > t)$. Se a suposição é de que cada caso tem a mesma distribuição de atraso e que não existem covariáveis que descrevam diferenças entre eles, $S(t)$ pode ser estimado não parametricamente pelo método de Kaplan-Meier¹³.

A partir das estimativas de $S(t)$, as probabilidades de atraso serão definidas por $p_i = S(s_{i-1}) - S(s_i)$ sendo os s_1, s_2, \dots, s_9 , os dias acumulados até um semestre de atraso, dois semestres de atraso e assim por diante. A distribuição do atraso por número de semestres ($p_1, p_2, p_3, \dots, p_9$) será usada para corrigir os casos de aids observados. Os números corrigidos de casos diagnosti-

cados em cada semestre, a partir do primeiro semestre de 1992, foram encontrados multiplicando-se o número de casos notificados até o momento, pelo inverso da soma das probabilidades dos atrasos a serem corrigidos.

Modelo de Regressão de Poisson

Os casos de aids para cada categoria de exposição foram classificados em uma tabela de contingência cujas celas Y_{ij} representam o número de casos diagnosticados no semestre I , e notificados J semestres mais tarde. A informação só é disponível para os Y_{ij} tal que $i+j \leq 21$. A Tabela 1 apresenta o número de casos notificados no Brasil, tanto por semestre de diagnóstico quanto pelo número de semestres em atraso, entre 1986 e 1996. Nesta tabela, as celas vazias indicam os casos que ainda não pude-

Tabela 2 - Probabilidades de atraso estimadas pela análise de sobrevivência por categoria de exposição.**Table 2** - *Exposure-specific delay probabilities specific-exposure estimates by survival analysis.*

Número de semestres	Homo	Uso de drogas	Hetero	Bissexual	Receptor de sangue	Peri	Hemo	Ign
nenhum	0.43	0.44	0.35	0.44	0.40	0.342	0.32	0.42
um a dois	0.21	0.28	0.27	0.20	0.22	0.24	0.16	0.24
dois a três	0.08	0.07	0.07	0.08	0.10	0.09	0.10	0.09
três a quatro	0.05	0.04	0.05	0.04	0.06	0.07	0.08	0.05
quatro a cinco	0.03	0.03	0.03	0.03	0.03	0.02	0.04	0.03
cinco a seis	0.03	0.02	0.03	0.02	0.03	0.03	0.04	0.02
seis a sete	0.02	0.01	0.02	0.02	0.03	0.02	0.03	0.02
sete a oito	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.02
oito ou mais	0.04	0.04	0.05	0.03	0.01	0.03	0.04	0.01

ram ser observados devido ao atraso da notificação. Assumindo-se que Y_{ij} tem distribuição de Poisson (m_{ij}), definiu-se:

$$\log(\mu_{ij}) = \alpha_0 + \beta(\text{semestre de diagnóstico}) + \theta(\text{atraso})$$

Visando relaxar a exigência de linearidade, utilizou-se o Modelo Aditivo Generalizado¹⁴ do programa S-PLUS¹⁵. Nesta abordagem, a tendência dos casos de Aids é representada como função de uma transformação não-paramétrica do tipo Spline do semestre de diagnóstico e do tamanho do atraso. Depois de definidas as transformações, ajustou-se pela função **glm** do S-PLUS o modelo:

$$\log(\mu_{ij}) = \beta + f(\text{semestre de diagnóstico}) + g(\text{semestres em atraso}),$$

onde **f** caracterizou a tendência na incidência da aids e **g** o comportamento do atraso, permitindo, assim, o exame em separado do efeito de cada um deles.

Resultados

As probabilidades de atraso, estimadas para cada intervalo de atraso (Tabela 2), mostraram que as categorias de exposição

hemofílico, heterossexual e perinatal apresentam a maior demora na notificação, com mais de 60% dos casos notificados depois de 6 meses de diagnóstico. No entanto, ao observar o gráfico da função $F(t)$, com t definido em número de dias, apresentado na Figura 2, para algumas categorias de exposição, que permite a visualização da probabilidade da notificação ocorrer antes de t dias, percebem-se diferenças importantes no comportamento da notificação destas categorias. A mediana da categoria dos hemofílicos está em torno de 400 dias, enquanto que a da transmissão heterossexual e perinatal está em torno de 250 dias apenas. Um outro padrão de notificação pode ser percebido entre os casos cuja exposição principal foi uso de drogas, transmissão bissexual ou homossexual, onde cerca de 50% dos casos são notificados antes dos 210 dias.

A análise dos gráficos apresentados na Figura 3, obtidos pelas regressões Splines, que descrevem as relações existentes entre os casos de aids e o semestre de diagnóstico, permite também identificar particularidades relacionadas ao comportamento da epidemia em cada categoria. Percebe-se, por exemplo, que a função Spline quadrática é crescente apenas até o segundo semestre de 1990 na categoria dos hemofílicos, diferenciando-se da categoria

d

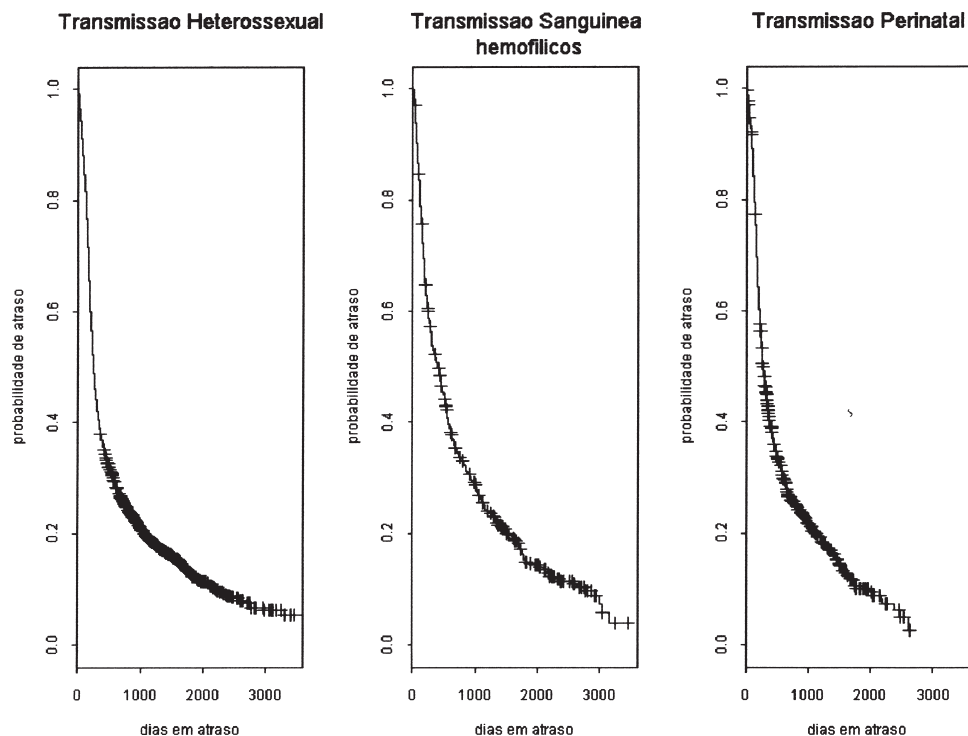


Figura 2 - Probabilidades de atraso estimadas pelo método de Kaplan-Meyer
Figure 2 - Delay probabilities estimates for Kaplan-Meyer method

dos usuários de drogas onde ela cresce até o segundo semestre de 1992, ou da categoria dos heterossexuais que continua crescendo até o segundo semestre de 1993.

Esta análise permitiu a escolha de transformações que linearizam as relações de interesse, satisfazendo assim os pré-requisitos necessários para a utilização da regressão de Poisson. Ao se analisarem estes gráficos percebe-se que as transformações para linearizar as relações vão diferir por categoria para a covariável semestre de diagnóstico. Já o comportamento da covariável semestre de atraso não difere muito entre as diversas categorias, fazendo com que a transformação necessária para a linearização do efeito desta covariável fosse única. A análise do ganho obtido ao se transformar as variáveis, foi realizada a partir da estatística $C_p^{14,15}$, considerada apropriada para comparar ajustes de modelos não aninhados. Esta estatística foi bem menor para todos os modelos e está apresentada na Tabela 3.

Os parâmetros estimados pela regressão de Poisson, realizada após as transformações, sinalizam as diferenças existentes entre as diversas categorias, tanto no nível médio que é refletido pelos interceptos, como no efeito do semestre de diagnóstico e do atraso (Tabela 3). Assim, verifica-se que apesar do número de casos ser maior na categoria de transmissão homossexual, os maiores interceptos correspondem às categorias dos usuários de drogas e dos heterossexuais, provavelmente refletindo o maior crescimento destas categorias. A grandeza dos parâmetros correspondentes ao efeito do atraso acompanha a direção do que já havia sido encontrado na análise de sobrevivência. O maior deles sendo para a categoria de menor atraso, a dos usuários de drogas, e o menor para a categoria dos hemofílicos, a de maior atraso.

A análise de sobrevivência estimou que deveria haver no Brasil 11.104 casos já diagnosticados e ainda não notificados mas,

Tabela 3 - Modelo ajustado, estimativas dos parâmetros com seus desvios e valor da estatística Cp.
Table 3 - Parameters estimates and standard deviation for each fitted model and Cp statistic.

Categoria de exposição	Modelo ajustado	α S(α)	β s(β)	θ s(θ)	Cp (antes)	Cp (depois)
Homossexual	$\log(\mu)=\alpha -\beta(\text{sem}-14)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	3.756(0.018)	0.0167(0.0003)	4.667(0.036)	8061.6	4667.6
Usuário de drogas	$\log(\mu)=\alpha +\beta(\text{sem}-14)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	3.814(0.019)	0.027(0.0004)	4.870(0.034)	9845.1	5653.6
Heterossexual	$\log(\mu)=\alpha +\beta(\text{sem}-16)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	4.012(0.022)	0.034(0.0005)	4.387(0.043)	6214.0	4794.9
Bissexual	$\log(\mu)=\alpha +\beta(\text{sem}-15)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	3.056 (0.027)	0.015(0.0003)	4.850(0.051)	4086.9	2465.7
Receptor de sangue	$\log(m)=\alpha +b(\text{sem}-15)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	1.935(0.047)	-0.011(0.0005)	4.267 (0.093)	1014.8	658.8
Perinatal	$\log(\mu)=\alpha +\beta(\text{sem}-16)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	1.965(0.057)	0.0164(0.0007)	4.090(0.112)	997.9	713.9
Hemofílico	$\log(\mu)=\alpha +\beta(\text{sem}-10)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	0.791(0.080)	0.0155(0.002)	3.486(0.176)	596.7	424.0
Ignorado	$\log(\mu)=\alpha +\beta(\text{sem}-15)^2+\theta(\text{raiz}(\exp(-\text{atras})))$	3.667(0.021)	0.018(0.0003)	4.588(0.041)	5721.9	4013.9

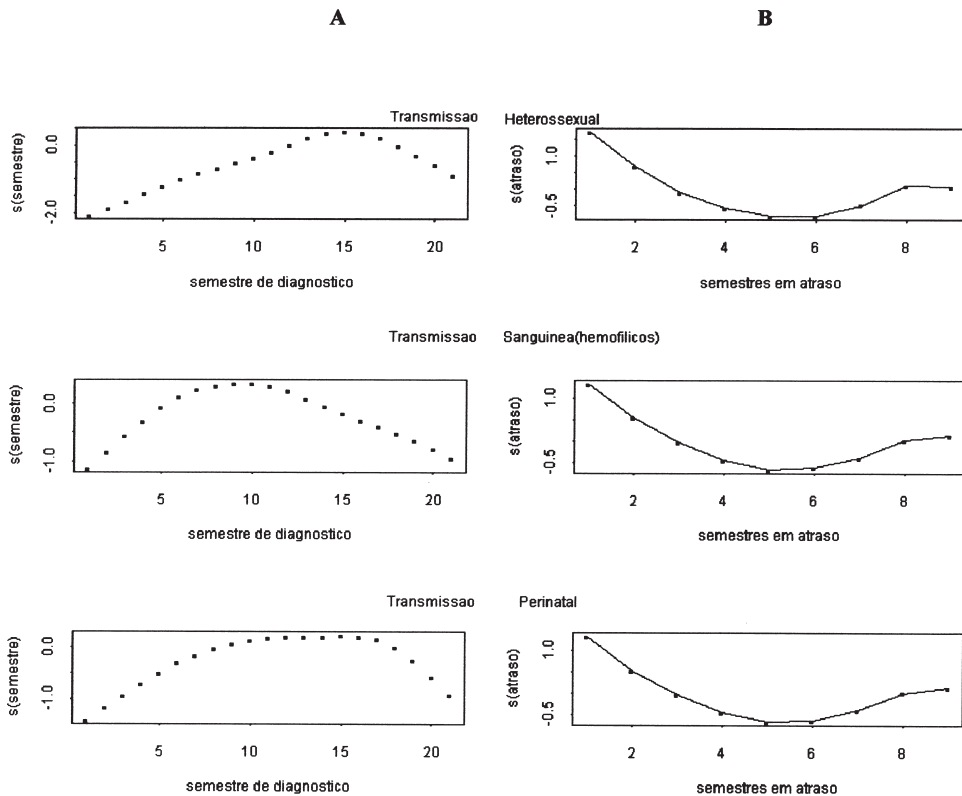


Figura 3 - Funções Splines das relações entre (A) os casos de AIDS e o semestre de diagnóstico e (B) os casos de AIDS e o número de semestres em atraso
Figure 3- Splines Functions of relation of the (A) AIDS cases and calendar-time and (B) AIDS cases and semesters reporting delay

Tabela 4 - Predições dos casos de AIDS realizadas pela Análise de sobrevida e Regressão de Poisson.

Table 4 - AIDS cases after correcting for delay in reporting predicted by survival analysis and Poisson regression.

Categoria de transmissão	Observados entre 86 e junho de 96	Correção (Análise de Sobrevida)	Correção (Regressão de Poisson)	Observado entre 86 e junho de 97
Homossexual	20.453	23.040	22.960	22.479
Usuário de drogas	18.937	20.903	21.402	21.167
Heterossexual	15.561	18.349	18.617	19.212
Bissexual	10.796	11.908	12.113	11.777
Receptor de sangue	2.967	3.307	3.311	3.295
Perinatal	2.033	2.317	2.406	2.516
Hemofílico	809	921	899	844
Ignorado	16.803	18.710	19.153	20.096
Total	88.359	99.455	100.861	101.386

pela regressão de Poisson, este número ficou em 12.052 casos. Ao se confrontarem estes números com os casos notificados após a data deste estudo, verificou-se que até junho de 1997 já existiam 13.027 novas notificações, superando ambas as estimativas. Ao se analisarem estas estimativas por categoria de exposição, percebe-se que a conclusão acima não vale para todas e, segundo os dois modelos, podem-se esperar, ainda, notificações de casos nas categorias classificadas como homossexual ou como bissexual (Tabela 4).

Discussão

Os resultados permitiram identificar diferenças no comportamento da notificação dos casos de aids entre as diversas categorias de exposição e ressaltar outras particularidades que podem afetar as medidas de controle e vigilância da epidemia. No entanto, antes da discussão a respeito das estimativas realizadas, será feita uma análise dos principais aspectos que podem afetar a validade deste estudo, baseando-se nos critérios definidos por Rothman e Greenland¹⁶.

Validade na comparação. Refere-se à suposição de que a distribuição do atraso na coorte alvo é a mesma da coorte completa.

Isto equivale a dizer que o atraso não depende da data de diagnóstico. Esta hipótese foi considerada na escolha dos modelos ajustados, e por não ser a mais provável, pode afetar de alguma maneira as estimativas.

Validade de seguimento. Supõe que a perda das datas de notificação ocorreu de forma aleatória e independente dos motivos que levam ao atraso. Para avaliar esta hipótese, estudaram-se as distribuições da idade, sexo, região e categoria de exposição para os casos com e sem data de notificação, verificando-se que a perda das datas poderia ser considerada aleatória.

Validade nas medidas. Refere-se à precisão do registro da data de diagnóstico e da data de notificação. No entanto, sabe-se que a data de diagnóstico deve apresentar grande variabilidade em seu registro, tanto no que se refere a diferenças no critério de diagnóstico, quanto às diferenças existentes em relação à qualidade do preenchimento desta data. Em relação à data da notificação, além da variabilidade entre as Unidades das Federações dos procedimentos de envio dos casos para a Coordenação do Programa Nacional/DST/AIDS, acrescenta-se a incerteza causada pelas perdas destas datas.

Validade na especificação. Refere-se à

adequação do modelo. Sabe-se que a verdadeira relação entre as variáveis quase nunca é conhecida em estudos de doenças humanas. Além disso, não se conhece o processo exato que levou as pessoas a serem diagnosticadas e notificadas e, assim, pertencerem à coorte de estudo. É esperado, portanto, um certo grau de imprecisão nas análises epidemiológicas aqui realizadas. A comparação dos resultados obtidos a partir de várias suposições (análise de sensibilidade) permite que se avalie a magnitude deste erro.

Os dois modelos estimaram os casos já diagnosticados, mas que deveriam ser notificados em vários semestres à frente, em menor número do que os casos realmente observados logo nos dois semestres seguintes. Esta subestimação pode ser atribuída a uma má especificação do modelo ou à imprecisão das medidas. No entanto, existe também a hipótese de que a distribuição de medicamentos realizada mais amplamente desde 1996 tenha diminuído bastante o atraso da notificação, e tenha levado os casos notificados ao verdadeiro nível da epi-

demia, sendo de maior impacto nas categorias onde o atraso era maior.

Assim, poderia ser explicado porque esta subestimação ocorre apenas entre as categorias de heterossexual, perinatal e ignorado. Ao considerar como verdadeira a hipótese de que os casos notificados já atingiram o verdadeiro nível da epidemia, conclui-se que as estimativas realizadas pelo modelo de Poisson se aproximaram mais da realidade. Na verdade, apenas a suposição de que o comportamento da epidemia seja distinto para as diversas categorias de exposição já faria optar por um modelo que incorporasse além do atraso, o nível da epidemia em cada semestre.

Por outro lado, a visibilidade da epidemia da aids e a relação direta entre o vírus e o comportamento humano individual e coletivo fazem com que seja bastante improvável que o padrão do atraso de notificação não venha se alterando no tempo. São estes múltiplos aspectos de uma epidemia ainda em mutação que tornam necessária a busca cada vez maior de modelos mais flexíveis que permitam captar esta dinâmica.

Summary

We estimated the number of AIDS cases in Brazil from official statistics, after correcting for delays in reporting by the use of two statistical models. First, we fitted a Kaplan-Meier survival model with right censoring to the distribution of reporting delays. The second approach is based on a Poisson regression model which was fitted to the contingency table having as one dimension the number of AIDS cases diagnosed per calendar time and as the other dimension the amount of delay in reporting. Before the latter approach, we explored, by the use of generalized additive models (GAM), the search for convenient smoothing functions that fit the observed functional relationships in question. Exposure-specific reporting delay probabilities, estimated by sur-

vival analysis (Table 1), showed that the hemophilic, heterosexual and neonatal exposure categories presented the longer delay in reporting, with more than 60% of total reported cases after 6 months of diagnosis. However, when observing Figure 2, where the delay is defined by number of days, we notice important differences in each category's reporting pattern. Among hemophiliacs, the median reporting delay is around 400 days, whereas in the heterosexual and neonatal transmission groups, it is much smaller, around 250 days. Curve fitting by splines regression is presented in Figure 3 and it makes it possible to identify distinct reporting patterns in each exposure category. We observe, for example, that the quadratic spline function

increases only up to the second semester in 1990 among hemophiliacs. However, the curves fitted for drug users and heterosexual categories increased up to the second semester in 1992 and 1993, respectively. Model fitting by Poisson regression to the transformed data suggests significant main effects for semester of disease diagnosis and delay in reporting, as well as different intercepts amongst the various epidemiologic categories (Table 3). Although the number of cases is higher among homosexuals, intercepts are larger among drug users and heterosexuals, probably reflecting a faster increasing epidemic in these categories. The magnitude of the parameters describing the delay effect follows the same direction as in the survival analysis approach. The parameter value is larger when the delay is small, as among drug users, and it is smaller among hemophiliacs. The latter category is subject to a greater delay in re-

porting. The estimates we obtained were then compared to the number of cases actually reported in the second semester of 1996 and in the first semester of 1997. Under the survival analysis approach we estimate that 11,104 cases have not yet been reported. Under the Poisson regression approach this number is now 12,052 cases. Comparing our predictions and the number of cases actually reported after our study had been completed, we notice that, as of June 1997, 13,027 new cases have been reported, surpassing both estimates. The official number of AIDS cases reported is very close to the actual figures, probably due to recent government policies regarding free distribution of medication. Our estimates, stratified by exposure category, suggest that the latter conclusion cannot be generally applied. One should expect additional cases reported for the homosexual and bisexual categories (Table 4).

Referências

- Bacchetti P, Segal MR, Jewell NP. Backcalculation of HIV infection rates. *Stat Sci* 1993; 8: 82-119.
- Brookmeyer R, Damiano A. Statistical methods for short-term projections of AIDS incidence. *Stat Med* 1989; 8: 23-34.
- Harris JE. Reporting delays and the incidence of AIDS. *Am Stat Assoc* 1990; 85: 915-24.
- Zeger SL, See L-C, Diggle PJ. Statistical methods for monitoring the AIDS epidemic. *Stat Med* 1989; 8: 3-21.
- Brookmeyer R, Liao J. Statistical modelling of the AIDS epidemic for forecasting health care needs. *Biometrics* 1990; 46: 1151-63.
- Brookmeyer R, Gail MH. *AIDS epidemiology: A quantitative approach*. New York: Oxford University Press; 1994.
- Sellero CS et al. Reporting delay: A review with a simulation study and application to Spanish AIDS data. *Stat Med* 1996; 15: 305-21.
- Clayton D. The analysis of event history data: A review of progress and outstanding problems. *Stat Med* 1988; 7: 819-41.
- Menezes RX. Previsão do número de casos de AIDS no estado de São Paulo usando o método *back-calculation*: uma análise crítica. [Dissertação de mestrado]. São Paulo: Instituto de Matemática e Estatística da USP; 1994.
- Barbosa MTS, Struchiner CJ. A correção do atraso da notificação por região brasileira. *Bol Epidemiol AIDS* 1996; 9(4).
- Bandeira AC de A, Silva GR. Calculation of HIV infection rates and projection of the number of cases of AIDS in São Paulo, Brazil using a backcalculation method. *Braz J Infect Dis* 1997; 1: 186-95.
- Barbosa MTS, Struchiner CJ. Estimativas do número de casos de Aids: comparação de métodos que corrigem o atraso da notificação. In: Ministério da Saúde. *A epidemia da AIDS no Brasil: situação e tendências*. Brasília; 1997.
- Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. New York: John Wiley and Sons; 1976.
- Hastie TJ, Tibshirani RJ. *Generalized additive models*. London: Chapman and Hall; 1990.
- Statistical Sciences, S-PLUS, Version 3.2. Seattle: MathSoft, Inc.; 1993.
- Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven; 1998.