**Previsão do número de casos de AIDS
no estado de São Paulo usando o método
“back-calculation”: uma análise crítica**

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e apresentada por Renée Xavier de Menezes e aprovada
pela Comissão Julgadora

São Paulo, 22 de dezembro de 1994.

Banca Examinadora:

- Prof. Dr. Marcos Nascimento Magalhães - IME-USP
- Prof. Dr. Júlio da Motta Singer - IME-USP
- Prof. Dr. Cláudio José Struchiner - ENSP/FIOCRUZ

Agradecimentos

Agradeço a todos que ajudaram a tornar este trabalho possível, em particular:

Prof. Dr. Júlio da Motta Singer (Instituto de Matemática e Estatística -USP)

Prof. Dr. Cláudio Struchiner (FIOCRUZ)

Prof. Dr. Eduardo Massad (Faculdade de Medicina-USP)

Prof. Dr. Marcelo Burattini (Faculdade de Medicina-USP e Escola Paulista de Medicina)

Prof. Dr. Raymundo Azevedo Neto (Faculdade de Medicina-USP)
pelo apoio, incentivo e produtivas discussões;

Prof. Dr. Marcos N. Magalhães, meu orientador;

A todos os pesquisadores que fazem parte do Núcleo de Pesquisas Epidemiológicas em AIDS - NUPAIDS da Universidade de São Paulo, com quem aprendi e tenho aprendido muito;

Ao Centro de Referência e Treinamento de AIDS do estado de São Paulo - CRT-AIDS/SP, particularmente à Dra. Naila J. Seabra Santos, ao Dr. Artur O. Kalichman e à Dra. Walkyria Pereira Pinto, pelo apoio;

Ao Alan Francisco de Carvalho, sem cuja paciência e enorme incentivo esse trabalho tornar-se-ia muito mais difícil;

A todos os professores e colegas do Departamento de Estatística do IME-USP, de quem sempre recebi incentivo;

Aos meus pais, irmãos, parentes e amigos que, de longe ou de perto, me ajudaram a vencer inúmeros momentos de ansiedade e angústia, lamentavelmente muito comuns na vida;

Aos profissionais de apoio (copiadores, bibliotecários), que tentam tornar o nosso dia-a-dia mais simples, apesar de todo o esforço que fazemos para complicá-lo ;

Resumo

O método back-calculation foi usado para fornecer previsões a curto prazo do número de casos novos de AIDS no estado de São Paulo, bem como para estimar o número total de pessoas portadoras do HIV. Até 1992, o estado de São Paulo era responsável por aproximadamente 60% dos casos notificados no Brasil. Para modelar o tempo de incubação foram usadas duas distribuições estimadas na literatura, ambas pertencentes à família Weibull de distribuições. A densidade de contaminação foi modelada por uma função escada, com três partições diferentes do período considerado (1977-1992). Procurou-se avaliar a sensibilidade dos resultados às suposições feitas. Os dados foram previamente corrigidos para o atraso de notificação através de um modelo Poisson. Supondo que a mediana do tempo de incubação seja igual a 4,3 anos, estima-se que haja 461036 pessoas portadoras do HIV no estado de São Paulo, com intervalo de confiança (428152,493920) com coeficiente de confiança 95%.

Abstract

The back-calculation method was used to obtain short-term projections of the number of new AIDS cases in the state of São Paulo, as well as to estimate the number of people infected with HIV. By 1992, the state of São Paulo was responsible for approximately 60% of the total number of reported cases in Brazil. To model the incubation time two distributions estimated in the literature were used, both of them pertaining to the Weibull family of distributions. The infection curve was modelled by a step-function, using three different partitions of the considered period (1977-1992). We tried to evaluate the sensitivity of the results with respect to the assumptions made. The data were previously corrected for the reporting delay through a Poisson model. Assuming that the median incubation time is 4,3 years, we estimate that there are 461036 people infected with HIV in the state of São Paulo, with a 95% confidence interval to this estimate being (428152,493920).

Índice

Introdução	1
1. Modelos para a epidemia de AIDS	6
1.1 Definições	6
1.2 Modelos	7
1.3 Variáveis importantes no estudo da epidemia	9
1.4 Histórico do método “back-calculation”	10
1.5 Dados	13
2. O método “back-calculation”	15
2.1 Introdução	15
2.2 Descrição do método	15
2.3 Estimação por máxima verossimilhança	17
2.4 Escolha da densidade de contaminação	20
2.5 Correção do atraso de notificação	21
3. Aplicação ao Estado de São Paulo	27
3.1 Dados e Correção do Atraso de Notificação	27
3.2 Escolha da distribuição do tempo de incubação	31
3.3 Densidade de contaminação	33
3.4 Resultados	35
4. Discussão e Conclusões	41
Bibliografia	48
Apêndice: Dados e Resultados	54

Introdução

O objetivo deste trabalho é estudar e analisar criticamente um método de previsão de casos de AIDS, chamado na literatura de “back-calculation”, e aplicá-lo ao estado de São Paulo, responsável hoje por cerca de 60% dos casos de AIDS no Brasil. O método também produz uma estimativa do número total de pessoas que foram contaminadas pelo HIV (“Human immunodeficiency virus”, o agente causador da AIDS) na população em estudo até um certo instante fixo, que é uma medida do tamanho da epidemia de infecção por HIV.

AIDS é a sigla usada para representar “Acquired Immunodeficiency Syndrome”. Em português seria talvez mais correto usar a sigla SIDA, que corresponde a “Síndrome da Imunodeficiência Adquirida”. Porém, a sigla AIDS está de tal forma difundida na literatura que, por uma questão de universalidade, optei por adotá-la neste trabalho.

A expressão em inglês “back-calculation” foi usada neste texto, também por uma questão de universalidade. Caso se deseje, expressões em português tais como “cálculo retroativo” e retrocálculo” (que corresponde à expressão usada por Rude, 1993, em francês - “retrocalcul”) poderiam ser usadas para referenciar o método.

A seguir, é feito um panorama das contaminações pelo HIV e dos casos de AIDS no mundo. Serão comentados sucintamente os modos de transmissão do vírus, avaliando sua importância. Por fim, a atual situação da epidemia de AIDS no Brasil e, particularmente, no estado de São Paulo, será brevemente descrita.

HIV e AIDS no mundo

O vírus conhecido como *vírus da imunodeficiência humana* (HIV) foi imputado em 1983 como o agente causador da AIDS (Barre-Sinoussi et al., 1983). Os exemplares do HIV que foram isolados são extremamente heterogêneos do ponto de vista molecular e biológico, sendo alguns mais virulentos que outros. A predominância de uma variante mais virulenta do vírus em uma região pode implicar numa frequência maior de casos graves que em outras regiões, onde predominem variantes menos virulentas. Além disso, indivíduos contaminados que vivem em locais com alta soroprevalência (proporção de indivíduos com anticorpos) de doenças infecciosas tendem a desenvolver a doença mais rapidamente que indivíduos em regiões com soroprevalência mais baixa, pelo fato de serem expostos mais intensamente a vírus diversos.

Um fator complicador na pesquisa de terapias, tanto para o prolongamento da sobrevida quanto para bloquear o aparecimento da doença, é o fato do HIV poder modificar-se muito rapidamente, mesmo dentro de um único indivíduo.

A evolução da pandemia de AIDS tem sido diferente em cada parte do mundo. Contribuem para esta variação as soroprevalências locais de doenças infecciosas, os hábitos culturais regionais e a virulência da variante do vírus predominante na região.

Desde que os primeiros casos de AIDS foram descritos em 1981, os padrões de transmissão do HIV mudaram, envolvendo grupos que inicialmente não eram afetados (Mann et al., 1993). Nos países industrializados da América do Norte, Europa, Austrália e Nova Zelândia, a maior incidência ocorreu inicialmente entre homossexuais do sexo masculino. As taxas de transmissão em usuários de drogas injetáveis variam segundo as regiões do mundo. Em alguns países o crescimento foi enorme enquanto, em outros, essas taxas permaneceram relativamente baixas. Em algumas regiões, particularmente na América Latina, incluindo o Caribe, tem sido observada uma mudança acentuada da transmissão homossexual para a heterossexual durante a última metade da década de 80. A expectativa é de que o HIV deve se disseminar mais amplamente em populações onde as taxas de doenças sexualmente transmissíveis são altas e, em todos os países do mundo, a transmissão por contato heterossexual deve se tornar a principal forma de propagação do vírus.

As estimativas da Organização Mundial de Saúde (OMS, 1991) indicam que, desde o início da epidemia, a África é o continente com maior incidência de infecção pelo HIV em adultos. O total acumulado de adultos infectados neste continente até 1991 era de aproximadamente 6 milhões. Nos países industrializados ocidentais, a prevalência (número de pessoas com a infecção) do HIV teve crescimento mais lento e a incidência parece ter alcançado seu máximo na metade dos anos 80. Nessa ocasião, o total acumulado de casos de infecções pelo HIV era de 1,5 milhões, aproximadamente. A incidência do HIV na América Latina foi estimada em 1 milhão. Vários países do sul e sudeste da Ásia relataram aumento dos casos de infecção pelo HIV no início de 1988 e, até a metade de 1991, mais de 1 milhão de pessoas podem ter se infectado pelo HIV nesta região. A pandemia de AIDS está agora entrando em sua segunda década. Estima-se que cerca de 5-6 milhões de homens e 3-4 milhões de mulheres em todo o mundo estejam infectados pelo HIV.

Ainda segundo a OMS, as projeções da prevalência do HIV estimadas a partir das taxas atuais de incidência mostram que se espera um aumento na transmissão do HIV na África, com um total acumulado de cerca de 10 milhões de infecções em 1995. Já nos países industrializados, as projeções indicam que a transmissão deve permanecer relativamente baixa, e o número acumulado de infecções deve alcançar a marca de 2 milhões em 1995. Nos países da América Latina, embora a transmissão do HIV deva continuar aumentando, este aumento não deve ser tão rápido como o esperado para outros países em desenvolvimento, de tal forma que em 1995 o total acumulado deve alcançar 2 milhões de infecções. Nos países do sul e sudeste asiático, por outro lado, a incidência de AIDS é baixa até agora, mas existe potencial para seu aumento devido às grandes populações envolvidas. Acredita-se que, no final dos anos 90, o coeficiente de incidência anual em alguma dessas regiões exceda o da África (OMS, 1991), alcançando um total acumulado de 2,5 milhões de infecções na metade da década de 90.

As projeções mais conservadoras (OMS, 1991) estimam que um mínimo de 38 milhões de adultos estarão infectados no ano 2000, enquanto projeções mais realistas afirmam que esse número alcançará até 110 milhões de pessoas. Ainda segundo estas projeções, o número acumulado de casos de AIDS no ano 2000 pode alcançar 25 milhões, com 42% desses casos ocorrendo na Ásia e Oceania, 31% na África, 14% na América Latina e Caribe, e 13% na América do Norte.

O impacto desta pandemia vai muito além das estatísticas. A AIDS, dife-

rentemente da malária, sarampo e poliomielite, afeta principalmente adultos jovens e de meia-idade. Não há dúvidas de que, durante as próximas décadas, a AIDS vai se tornar a principal causa de mortalidade em adultos, atingindo os indivíduos na fase mais produtiva da vida, e também uma das principais causas de mortalidade neonatal e infantil nos países em desenvolvimento. A esperança de vida ao nascer já está diminuindo na maioria das áreas afetadas. Além disso, terá um profundo impacto não apenas sobre os sistemas de saúde e bem-estar social, mas também sobre aspectos sócio-econômicos e outros relacionados à qualidade de vida.

Mecanismos de transmissão do HIV

Existem basicamente três maneiras pelas quais o HIV é transmitido, a saber, o contato sexual, o contato direto com sangue ou fluidos orgânicos infectados e a transmissão perinatal (da mãe infectada para o feto). Não existem evidências de infecção ocasional, tais como pelo uso de assento de vasos sanitários ou através de insetos (Levy, 1993).

Os dois principais modos de transmissão são o contato sexual e a contaminação por sangue e/ou derivados. No caso de transmissão perinatal, o momento exato da transmissão do HIV da mãe para seu conceito e o papel da amamentação ainda não foram totalmente esclarecidos, mas se tem notado que aproximadamente 30% das crianças nascidas de mães contaminadas são igualmente contaminadas.

Existem indicações de que a transmissibilidade da infecção pelo HIV varie de maneira significativa durante o curso da infecção em um indivíduo, devendo ser mais intensa durante a fase inicial da infecção, antes do início da produção de anticorpos, e nos estágios mais avançados, quando o sistema imunológico já se encontra deprimido (Goedert et al., 1987 e Longini et al., 1990).

AIDS no Brasil

Desde que o primeiro caso de AIDS foi diagnosticado no Brasil em 1980 até agosto de 1994, 55894 casos foram notificados (Programa Nacional de DST/AIDS, 1994). Destes, 79% são observados apenas na região sudeste.

Nos primeiros 5 anos, a doença ocorreu predominantemente entre homossexuais e bissexuais. Depois esta situação foi mudando gradualmente,

aumentando a proporção de mulheres infectadas, tanto por exposição heterossexual como por uso de drogas injetáveis. Esta mudança resultou em uma alteração na relação homem-mulher de 32:1 em 1983 para 4:1 atualmente.

O risco acumulado de AIDS é distribuído como se segue (Programa Nacional de DST/AIDS, 1994):

- 51% por exposição sexual, com 40% para atividade homo ou bissexual e 11% para atividade heterossexual;
- 27% por exposição parenteral, com 22% por uso de drogas injetáveis e 5% por transmissão relacionada a transfusão;
- 2% por transmissão perinatal;
- 11% sem nenhuma causa atribuída.

A distribuição dos casos relatados atualmente no Brasil por via de transmissão mostra um equilíbrio entre usuários de drogas injetáveis, transmissão heterossexual e homossexual.

O estado de São Paulo é o principal centro econômico da América Latina, com características de liderança cultural e social que promovem uma intensa relação com a América do Norte e a Europa. Portanto, não é surpreendente que os primeiros casos relatados no Brasil tenham ocorrido em São Paulo, nem que 60% do número total de casos relatados no Brasil sejam desse estado.

O padrão de transmissão no estado de São Paulo mudou da exposição predominantemente homo/bissexual (90% dos casos relatados em 1984) para proporções equilibradas de casos relacionados a exposição homo/bissexual, heterossexual e uso de drogas injetáveis (Centro de Vigilância Epidemiológica/SP, 1994). A distribuição etária nos casos de AIDS mostra que aproximadamente 76% têm entre 15 e 39 anos, com 36% entre 20 e 29 anos.

Em 1991, a AIDS era a terceira causa de diminuição da esperança de vida ao nascer na cidade de São Paulo, sendo suplantada apenas por causas de mortalidade infantil e por homicídios.

Organização deste texto

Este texto está dividido em quatro capítulos. No Capítulo 1 são apresentados os aspectos da doença e da sua progressão que tem sido relevantes para o estudo da evolução da pandemia. Paralelamente, é feita uma revisão geral dos trabalhos publicados que estudam a progressão da pandemia de

AIDS. Na seção 1.3, uma revisão bibliográfica mais detalhada é feita sobre os trabalhos que usam o método “back-calculation”. O conjunto de dados que será usado é descrito na seção 1.4.

No Capítulo 2 o método “back-calculation” é apresentado. A escolha da densidade de contaminação é avaliada na seção 2.4 e, na seção 2.5, o problema do atraso de notificação é abordado.

O Capítulo 3 trata da aplicação do método aos dados do estado de São Paulo. A forma como os dados são usados e a correção do atraso de notificação são vistos na seção 3.1. As seções 3.2 e 3.3 tratam das escolhas da distribuição do tempo de incubação e da densidade de contaminação, respectivamente. Finalmente, na seção 3.4 são analisados os resultados obtidos.

Discussão e conclusões estão no Capítulo 4.

No Apêndice estão reunidos os resultados numéricos.

Capítulo 1

Modelos para a epidemia de AIDS

1.1 Definições

Um indivíduo é dito *suscetível* a contrair o vírus HIV se seu comportamento abre a possibilidade de ele ser contaminado.

Diz-se que um indivíduo é *soronegativo* para uma certa doença se, no exame de uma amostra do seu sangue, não são encontrados anticorpos correspondentes ao vírus desta doença, e o indivíduo é dito *soropositivo* se forem encontrados os anticorpos. A ausência de anticorpos não significa necessariamente não estar contaminado pelo vírus. Especificamente no caso da AIDS, o período desde a contaminação (recebimento do vírus pela primeira vez) até o aparecimento dos anticorpos pode durar desde poucas semanas até vários meses. Alguns dos autores que estudaram a duração deste período são Simmons et al. (1988) e Cooper et al. (1985). Diz-se que houve uma *soroconversão* quando um indivíduo passa de soronegativo para soropositivo. Em outras palavras, o indivíduo soroconverte-se quando passa a produzir anticorpos relativos ao vírus.

O período que vai desde a soroconversão até o diagnóstico da doença será definido como *período de incubação*. Durante este período o indivíduo pode ou não ter sintomas discretos da doença, e já é capaz de transmitir o vírus a outros. Este período também é denominada fase de latência ou

HIV-até-AIDS ¹.

Existem vários métodos que, a partir da incidência anual de AIDS, fazem previsões da incidência futura. As previsões são fundamentais para dimensionar a oferta de leitos, medicamentos, etc. Neste trabalho, vamos utilizar o método de projeção conhecido como “back-calculation”, que será descrito na seção 2.2.

1.2 Modelos

Dois modelos de previsão vinham sendo usados até meados da década de 80. O primeiro consiste em ajustar uma função para a curva de incidência de AIDS e extrapolar os resultados para o futuro. É de simples implementação e, por isso, bastante atrativo, mas a extrapolação de uma função ajustada para fora do período considerado no ajuste é delicada, pois não há como garantir que a progressão da epidemia se dará nas mesmas condições que geraram os dados coletados. Dentre os trabalhos que seguem este caminho, podemos citar McEvoy e Tillett (1985), onde é feita uma regressão da incidência sobre o ano, e Medley et al. (1987), que utiliza uma forma matemática simples (linear ou exponencial) para descrever a curva de prevalência da doença entre pacientes contaminados através de transfusão.

O outro modelo de previsão usa equações diferenciais para modelar a dinâmica da epidemia, descrevendo o fluxo de indivíduos entre as diversas fases da doença, desde não-contaminados suscetíveis, passando pelo diagnóstico, até o óbito. Bailey (1975) é uma boa referência para este tipo de modelo e outra revisão foi feita por Isham (1988). Em Massad (1987) pode ser encontrado um estudo teórico desses modelos, que inclui simulações para a AIDS e outras moléstias infecciosas. Especificamente para a AIDS, seu modelo é baseado no modelo determinístico geral de epidemias de Bailey (1975), diferindo deste por ser multicompartimental e incluir taxas de imigração e mortalidade. Dentre os trabalhos que aplicam o modelo podemos citar: Mode et al. (1988), que avalia o impacto da distribuição do tempo de incubação e de outros parâmetros sobre as projeções feitas a partir de um sistema de equações diferenciais não-lineares; Arcà et al. (1992), onde a disseminação do vírus na população italiana é explicada através de um modelo matemático compartimental, que consiste num sistema de equações diferen-

¹ Alguns autores incluem neste período a fase soronegativa.

ciais; e Heisterkamp et al. (1992), que aplica um modelo determinístico a dados de incidência de AIDS de homens homossexuais e bissexuais de Amsterdam. Infelizmente, estes modelos de equações diferenciais dependem de hipóteses de difícil verificação, como homogeneidade de comportamento dos indivíduos dentro da população (ou dentro de subgrupos dela) e de quantidades difíceis de serem estimadas, tais como a proporção da população que é suscetível.

Aos dois modelos citados foi acrescentado o método "back-calculation", introduzido por Brookmeyer e Gail (1986) para obter projeções a curto prazo da incidência de casos de AIDS. Esta técnica usa a distribuição do tempo de incubação para reconstruir a curva de contaminação a partir da série observada de casos de AIDS. Usa-se o método para calcular projeções da incidência apenas a curto prazo porque ele só leva em consideração casos emergentes de indivíduos já contaminados até o fim do estudo, sem contabilizar novas contaminações. Isso significa que as projeções calculadas são, na realidade, cotas inferiores da incidência. Como o período de incubação do HIV é longo, novas contaminações tendem a afetar pouco as projeções a curto prazo.

O método tem sido aplicado em diversos trabalhos, tais como Anderson et al. (1987), que utiliza dados da Grã-Bretanha, Longini et al. (1992), que lança mão de um modelo em estágios para estimar a distribuição do tempo de incubação, e Mariotto et al. (1992), que aplica o método ao caso da Itália.

As estimativas do número de contaminados pelo HIV produzidas pelo "back-calculation" têm se mostrado consistentes com informações de fontes independentes, obtidas em estudos de prevalência em populações especiais, e simulações indicam boa confiabilidade estatística das projeções, como mostram Rosenberg e Gail (1990) e Rosenberg, Biggar et al. (1991).

Em Gail e Brookmeyer (1988) pode ser encontrado um estudo comparativo dos vários métodos de projeção do curso da epidemia de AIDS.

Dois pontos desta técnica merecem atenção: a escolha da densidade de contaminação e a escolha da distribuição do tempo de incubação. No caso da primeira usualmente atribui-se a ela uma forma paramétrica, estimando os parâmetros a partir dos dados simultaneamente à aplicação do back-calculation.

No caso da distribuição do tempo de incubação, pode-se obter uma sugestão a partir de outros trabalhos. Os estudos que envolvem pacientes contaminados através de transfusão de sangue têm sido importantes para este fim, pois a data de contaminação é conhecida nesses casos. Existem também

modelos para, a partir de informações acerca da evolução da doença nos indivíduos, estimar a distribuição do tempo de incubação. Esses modelos são comentados na seção 1.4.

O trabalho de Gamerman e Migon (1991) se destaca dos demais por usar um método diferente dos já citados e por usar dados da epidemia de AIDS no Brasil. Seu método se baseia na classe de modelos de crescimento exponencial generalizado e nas idéias de modelagem dinâmica não-linear para prever a incidência e a prevalência do número de casos de AIDS notificados. A vantagem de usar modelos dinâmicos é a de acomodar mudanças nas estruturas dos dados através do tempo.

1.3 Variáveis importantes no estudo da epidemia

Pelo fato de a AIDS ter sido descoberta recentemente, ainda se conhece pouco sobre a doença e fatores que influenciam seu desenvolvimento. Alguns destes fatores estão relacionados com as terapias usadas para retardar o desenvolvimento da doença e prolongar a sobrevivência do paciente, como o uso do AZT (zidovudine). Vários autores têm se preocupado com a incorporação do tratamento no modelo de projeção da incidência. Se for utilizado um modelo em estágios para estimar a distribuição do tempo de incubação, o usual é tomar diferentes taxas de transição dependendo do uso ou não de tratamento (consulte, por exemplo, Brookmeyer, 1991, e Longini et al., 1992). O modelo para a distribuição do tempo de incubação denominado "Time Since Infection" (TSI) foi construído especificamente para incorporar o efeito do tratamento (veja Rosenberg et al., 1992 e Gail e Rosenberg, 1992).

Há ainda outros fatores que podem influenciar o desenvolvimento da doença, tais como etnia, sexo, idade, presença de outras doenças associadas, região geográfica e modo de transmissão. Medley et al. (1987) usam dados de indivíduos contaminados por transfusão e consideram sexo e idade como fatores, notando que os períodos de incubação parecem ser mais longos para mulheres que para homens. Em Murphy et al. (1991) também são avaliados os efeitos de sexo e idade através de uma análise de regressão logística com dados de soroprevalência de HTLV-I de uma coorte de jamaicanos acompanhada durante dois anos. Concluem que a soroprevalência de HTLV-I na Jamaica está fortemente ligada a estes fatores, passando de 1,7% na faixa

de 10 a 19 anos para 9,1% na faixa de 70 anos ou mais, entre homens e, de 1,9% (10 a 19 anos) para 17,4% (70 anos ou mais), entre mulheres. Em Mariotto et al. (1992) são usados dados do Istituto Superiore di Sanità, Itália, para comparar a duração do período de incubação entre usuários de drogas injetáveis e homens homossexuais, levando em consideração a idade e o modo de transmissão. A taxa de diagnóstico de AIDS está positivamente relacionada com a idade. O risco anual estimado de a AIDS se desenvolver em indivíduos entre 25 e 34 anos e em indivíduos com 35 anos ou mais foi 2,4 e 3,8 vezes maior que o correspondente a indivíduos entre 16 e 24 anos, respectivamente. Outros estudos, como Brookmeyer e Goedert (1989) e Biggar (1990), também constatarem que a idade influencia a duração do período de incubação. Já o efeito do modo de transmissão sobre a duração da incubação não se mostrou estatisticamente significativo, o que é confirmado em outros estudos, tais como Jason et al. (1989) e Darby et al. (1990).

O fator geográfico exerce influência tanto sobre o processo de disseminação do vírus quanto sobre a duração do período de incubação. Em relação à disseminação do vírus, a influência se dá pelos hábitos da população local. Por exemplo, regiões onde uma porcentagem maior da população usa drogas injetáveis tendem a ter uma incidência maior de contaminados pelo HIV. A associação entre região e duração da incubação parece ser consequência da soroprevalência local de doenças infecciosas e/ou predominância de variantes do vírus HIV com virulências diferentes, opinião também de Medley et al. (1987). Em alguns trabalhos foram encontradas evidências de incubações substancialmente diferentes em populações distintas (veja por exemplo Bacchetti et al., 1993 e Biggar, 1990).

O atraso de notificação, que será definido na seção 2.5, também pode variar de uma região para outra, como notam Zeger et al. (1989).

1.4 Histórico do método “back-calculation”

Este método foi introduzido por Brookmeyer e Gail (1986) para calcular projeções da incidência de AIDS (para uma referência mais detalhada pode-se consultar também Brookmeyer e Gail, 1988). O método supõe que são conhecidas a distribuição do tempo de incubação e a forma paramétrica da densidade de contaminação. A eficiência do método dependerá fortemente, portanto, da adequação das escolhas feitas.

No caso da densidade de contaminação não há necessidade de se escolher uma distribuição específica, bastando optar por uma forma funcional e estimar os parâmetros a partir dos dados. Longini et al. (1992) usaram a distribuição log-logística. Em Marion e Schechter (1993) são comparadas uma função escada, a distribuição log-logística e a distribuição logística. Já Medley et al. (1987) supõem que a contaminação tem crescimento exponencial e, baseado nisto, ajustam uma distribuição Weibull e uma distribuição Gama. Uma função escada foi a forma escolhida por Brookmeyer e Gail (1988), da mesma maneira que Brookmeyer e Damiano (1989), Rosenberg, Biggar et al. (1991), Mariotto et al. (1992) e Rude (1993). Kuo et al. (1991) utilizaram uma função cuja forma é coerente com a indicada por modelos matemáticos para epidemias. A vantagem da escolha da função escada está no fato de que este tipo de função serve para aproximar qualquer função contínua. Logo, qualquer que seja a forma verdadeira da densidade de contaminação na população considerada, ela poderá ser razoavelmente bem aproximada por uma função escada e, quanto maior for o número de patamares, melhor será a aproximação.

Para escolher uma distribuição para o tempo de incubação podem-se tomar sugestões de outros trabalhos. Podemos citar os estudos que envolvem pacientes contaminados através de transfusão, como hemofílicos. Estes estudos têm sido importantes porque nestes casos é possível determinar com exatidão a data da contaminação.

A utilização de distribuições sugeridas por estudos de outras localidades deve ser feita com cautela. O primeiro motivo é que existem diversas variedades de vírus HIV. Além disso, diferenças entre as soroprevalências de doenças infecciosas de duas regiões podem determinar diferenças nas durações dos períodos de incubação de pessoas contaminadas nelas residentes.

Diversos trabalhos têm aplicado o método "back-calculation". Anderson et al. (1987) utilizaram o método para prever a incidência de AIDS entre homossexuais da Grã-Bretanha, tomando a distribuição Weibull como distribuição do tempo de incubação. A distribuição Weibull tem sido uma das mais usadas para modelar a duração do período de incubação, tendo sido também a escolha de Brookmeyer e Gail (1986, 1988), Medley et al. (1987) (que usaram também a distribuição gama), Brookmeyer e Damiano (1989), Brookmeyer e Goedert (1989), Rosenberg, Biggar et al. (1991) e Kuo et al. (1991) (que ajustaram também a distribuição log-logística), entre outros.

A distribuição do tempo de incubação pode ser estimada a partir de dados

de acompanhamento de coortes de pacientes. Um dos modelos que podem ser usados para este fim se baseia na divisão do período de incubação em estágios. A progressão do indivíduo contaminado pelos estágios é vista como um processo estocástico, e a partir de estimativas dos parâmetros do processo estima-se a distribuição do tempo de incubação. O impacto do tratamento é acomodado separadamente em cada estágio. Esta forma de modelo para a distribuição é chamada de *modelo em estágios* ("staged model"). Este modelo tem sido aplicado em diversos trabalhos. Longini et al. (1989) modelaram o período de incubação como um processo de Markov com 2 estágios², obtendo uma estimativa da distribuição de incubação a partir de uma coorte da cidade de São Francisco (EUA), mas não calcularam projeções da incidência. Supõe-se que o tempo de permanência em cada estágio tenha distribuição exponencial.

O modelo em estágios foi utilizado acoplado ao "back-calculation" pela primeira vez por Brookmeyer e Liao (1990); nesse caso, o método foi estendido para estimar a incidência e a prevalência de contaminações com o HIV, por estágio, e para modelar o tempo de visita a cada estágio foi usada a distribuição Weibull. Em Longini et al. (1992) o modelo em estágios também foi aplicado junto ao "back-calculation", tomando a distribuição Erlang para modelar o tempo de visita a cada estágio. Também Mariotto et al. (1992) utilizaram o modelo em estágios com a distribuição Weibull associada ao "back-calculation", aplicando o método ao caso dos usuários de drogas na Itália. Como a distribuição exponencial é um caso particular de ambas as distribuições Erlang e Weibull, a utilização das duas últimas para modelar o tempo de visita a cada estágio generaliza o processo de Markov. A vantagem de usar a distribuição Erlang é a de manter a propriedade markoviana.

Um outro modelo que tem sido usado para estimar a distribuição do tempo de incubação é o denominado "Time Since Infection" (TSI). A vantagem sobre o modelo em estágios é a de incorporar explicitamente a duração do tratamento. Por outro lado, como a utilização ou não de tratamento está mais ligada à sintomatologia (que determina o estágio da doença) que ao tempo desde a contaminação, associar a influência do tratamento sobre a evolução da doença ao estágio é mais natural, o que torna o modelo em estágios mais intuitivo. Este modelo foi usado por Rosenberg et al. (1992) e por Gail e Rosenberg (1992).

²O processo que descreve a progressão completa da doença tem 5 estágios, e o período de incubação corresponde a 2 deles.

Marion e Schechter (1992) introduziram mais um modelo para estimar a distribuição do tempo de incubação, chamado na literatura de modelo "Dampened Hazard" (DH). Neste modelo, a distribuição do tempo de incubação tem taxa de falha igual à de uma distribuição Weibull até um certo instante t fixo e, a partir de t , esta taxa se mantém constante. Este modelo é equivalente a um modelo em 2 estágios, onde a distribuição do tempo de visita ao primeiro tenha distribuição Weibull e a do segundo tenha distribuição Exponencial.

1.5 Dados

Para utilizar o método "back-calculation", dadas a distribuição do tempo de incubação e uma forma paramétrica da curva de contaminação, é necessário conhecer a incidência passada da doença, em intervalos regulares.

A AIDS é uma doença de notificação compulsória; logo, em princípio, todos os casos diagnosticados deveriam ser notificados à vigilância sanitária. Porém, existe uma defasagem de tempo entre o momento do diagnóstico e a notificação à vigilância, o que implica subnotificação nos registros mais recentes da incidência. Para corrigir este problema, alguns trabalhos têm procurado estimar a distribuição do tempo de atraso e corrigir os dados. Esta correção é chamada de *correção do atraso de notificação* ("adjustment for reporting delay") e será estudado em maiores detalhes na seção 2.5.

Caso seja necessário estimar também a distribuição do tempo de incubação, deve-se ter dados de acompanhamento de pacientes, registrando em diversos momentos seu estado clínico. Este tipo de informação é tradicionalmente levantado em estudos de coortes, onde pessoas são cadastradas e acompanhadas através de exames periódicos. A partir desta informação, pode-se avaliar a duração do período de incubação e utilizar algum método para estimar sua distribuição. Se for usado o modelo em estágios, por exemplo, os registros devem permitir que se conheça a progressão do indivíduo pelos estágios da doença. Infelizmente, até dezembro de 1993 não havia disponível nenhum estudo dessa natureza envolvendo residentes no Brasil, forçando a utilização de distribuições do tempo de incubação estimadas em outras populações.

Neste trabalho, foram usados dados obtidos junto ao Centro de Referência e Tratamento de AIDS do Estado de São Paulo (CRT-AIDS), a partir de

boletins mensais publicados desde abril de 1988 até março de 1993. Cada caso notificado é registrado segundo o ano em que foi diagnosticado e, a cada mês, o número de notificações já feitas é atualizado, a partir de novas notificações e óbitos. Os dados estão relacionados resumidamente no Apêndice 1, por ano de diagnóstico e ano de notificação.

Capítulo 2

O método “back-calculation”

2.1 Introdução

Neste capítulo são descritos o método “back-calculation” e o método usado para corrigir o atraso de notificação. Nas seções 2.2 e 2.3 o método é detalhado e o procedimento de cálculo das estimativas de máxima verossimilhança é descrito como em Brookmeyer e Gail (1988). Na seção 2.4 são apresentadas algumas das formas paramétricas que têm sido usadas para a distribuição de contaminação. Finalmente, na seção 2.5 é descrita a metodologia de correção do atraso de notificação utilizada para corrigir os dados, como em Brookmeyer e Damiano (1989).

2.2 Descrição do método

Basicamente, o método consiste em estimar o número de pessoas contaminadas pelo HIV que desenvolverão a doença num futuro próximo, a partir da distribuição do tempo de incubação do vírus, da densidade de contaminação e da incidência da doença.

Suponha que foram registrados os casos de AIDS diagnosticados desde o início da epidemia (T_0) até um certo instante T_D . Seja N o número de pessoas contaminadas pelo HIV entre T_0 e T_D . Parte destas pessoas desenvolveu a doença (e, portanto, teve diagnóstico de AIDS) até T_D . Outra parte deve desenvolver a doença após T_D . O principal objetivo deste trabalho é estimar quantas são as pessoas que não desenvolveram a doença até T_D .

Seja $F(\cdot)$ a função de distribuição acumulada do tempo de incubação (que vamos representar por T_{inc}). Considere D intervalos $[T_{j-1}, T_j)$, $j = 1, 2, \dots, D$, e suponha que em cada intervalo observou-se o número de casos de AIDS diagnosticados, X_j , $j = 1, 2, \dots, D$. Note que cada pessoa só tem diagnóstico de AIDS uma vez. Sem perda de generalidade, suponha que $T_0 = -\infty$, de forma que não havia ninguém contaminado antes de T_0 . Sejam $X = X_1 + X_2 + \dots + X_D$ o número total de casos diagnosticados até T_D , e X_{D+1} o número de pessoas contaminadas antes de T_D mas que não tiveram diagnóstico até T_D . Assim, $N = X + X_{D+1}$ é o tamanho mínimo (a cota inferior) da epidemia após T_D , pois no mínimo ainda devem aparecer X_{D+1} casos (situação limite se ninguém mais for contaminado após T_D). Observe que tanto N quanto X_{D+1} são desconhecidos.

Suponha que, para cada indivíduo contaminado, a duração do período de incubação seja independente do instante de contaminação. Então, dado N , a distribuição conjunta de $\mathbf{X} = (X_1, X_2, \dots, X_D, X_{D+1})$ é Multinomial.

Seja $I(s)$ a função densidade de contaminação dos indivíduos contaminados de T_0 até T_D . Ou seja, dado que um certo indivíduo teve diagnóstico de AIDS até T_D , a probabilidade de que tenha sido contaminado em $(t, t + \Delta t)$, $t + \Delta t \leq T_D$ é aproximada por $I(t)\Delta t$, onde Δt é um intervalo de tempo pequeno; além disso,

$$\int_{T_0}^{T_D} I(s)ds = 1.$$

Com base na hipótese de independência entre os indivíduos, dado que houve N contaminações até T_D , os instantes de contaminação podem ser considerados como variáveis aleatórias independentes e identicamente distribuídas (iid), cada uma com densidade $I(s)$. Suponha que $I(s)$ possa ser representada por um membro de uma família paramétrica de parâmetro desconhecido $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, $\theta \in \Theta$, $\Theta \subset \mathbb{R}^k$.

A partir da hipótese de que a duração do período de incubação é independente da data de contaminação, segue que a distribuição de T_{inc} é independente do particular instante em que houve a contaminação. Então, dado que N pessoas foram contaminadas até T_D , a probabilidade de uma pessoa receber diagnóstico de AIDS no intervalo j , $j = 1, 2, \dots, D$, pode ser escrita como

$$\begin{aligned}
\pi_j &\triangleq P \{ \text{pessoa ser diagnosticada em } [T_{j-1}, T_j] \} \\
&= \int_{T_0}^{T_j} P \{ \text{ser infectado em } s \leq T_j \text{ e ser diagnosticado entre } T_{j-1} \text{ e } T_j \} ds \\
&= \int_{T_0}^{T_j} P \{ \text{ser diagnosticado em } [T_{j-1}, T_j] \mid \text{infectado no instante } s \} \\
&\quad P \{ \text{ser infectado no instante } s \} ds \\
&= \int_{T_0}^{T_j} P \{ T_{j-1} - s < T_{\text{inc}} \leq T_j - s \} P \{ \text{ser infectado no instante } s \} ds,
\end{aligned}$$

ou seja,

$$\pi_j = \int_{T_0}^{T_j} \{ F(T_j - s) - F(T_{j-1} - s) \} I(s) ds. \quad (2.1)$$

Condicionalmente a N , $\mathbf{X} = (X_1, X_2, \dots, X_D, X_{D+1})$ tem distribuição multinomial com parâmetros N e $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_D, 1 - \pi.)$, onde $\pi. = \pi_1 + \dots + \pi_D$. Note que, como $F(.)$ é dada e $I(s)$ está sendo parametrizada por $\boldsymbol{\theta}$, a expressão 2.1 relaciona π_j e $\boldsymbol{\theta}$.

Observam-se $X_1, X_2, \dots, X_D$, ou o número de casos novos em cada intervalo, até T_D . O que se quer estimar são N e $\boldsymbol{\pi}$. Como $\boldsymbol{\pi}$ pode ser escrito apenas em função do parâmetro desconhecido $\boldsymbol{\theta}$, o problema se reduz à estimação de N e $\boldsymbol{\theta}$. A estimação do tamanho de uma população multinomial foi abordada na literatura em diversos contextos diferentes (Bishop, Fienberg e Holland 1975, por exemplo). Na seção 2.3 será descrito o procedimento seguido por Brookmeyer e Gail (1988), que obtêm as estimativas de máxima verossimilhança incondicionais (denominadas como “full maximum likelihood estimates”). Outra alternativa poderia ser obter estimativas de máxima verossimilhança condicionais, como na proposta de Sanathanan (1972).

2.3 Estimação por máxima verossimilhança

Considere a log-verossimilhança multinomial:

$$\begin{aligned} \log L(N, \theta; \mathbf{X}) = & \log(N!) - \sum_{i=1}^D \log(X_i!) - \log\{(N - X_{\cdot})!\} \\ & + \sum_{i=1}^D X_i \log[\pi_i(\theta)] + (N - X_{\cdot}) \log[1 - \pi_{\cdot}(\theta)] \end{aligned} \quad (2.2)$$

Maximizando esta função em relação a N e θ , encontramos os estimadores de máxima verossimilhança desejados, $\hat{N}$ e $\hat{\theta}$, e $\hat{\pi}$ a partir destes. Assim, a previsão a curto prazo do número de casos novos é $\hat{N}(1 - \hat{\pi}_{\cdot})$, que corresponde ao número de pessoas contaminadas antes de T_D mas não diagnosticadas até T_D .

Para encontrar os estimadores, pode-se por exemplo considerar a log-verossimilhança 2.2 como função contínua de N (já que N é grande) e usar a teoria de verossimilhança tradicional. Aproximando $d \log(N!)/dN$ por $\log N$ (Abramowitz e Stegun, 1970), segue que a equação escore aproximada para N é

$$\hat{N} \cong \frac{X_{\cdot}}{\hat{\pi}_{\cdot}}, \quad (2.3)$$

e as equações escore para θ podem ser obtidas derivando 2.2 em relação a cada θ_j e igualando o resultado a zero, $j = 1, 2, \dots, k$. As estimativas são encontradas resolvendo iterativamente estas últimas equações e 2.3.

Uma outra forma de encontrar as estimativas aproximadas de máxima verossimilhança consiste em propor uma grade $\mathcal{G}$ de valores para N e, para cada valor na grade, maximizar 2.2 em relação a θ . A estimativa de máxima verossimilhança de N seria aproximada pelo valor da grade que produzisse o maior valor da log-verossimilhança, e a estimativa de máxima verossimilhança de θ seria a correspondente àquela log-verossimilhança.

Vamos representar por $LL(N)_{\theta^*}$ a função de log-verossimilhança maximizada em θ , para um certo N fixo. Ou seja,

$$LL(N)_{\theta^*} = \sup_{\theta} \log L(N, \theta; X),$$

onde θ^* é o valor de θ que maximiza a função. Calcula-se o valor de $LL(N)_{\theta^*}$ para cada valor de N na grade estabelecida. As estimativas de N e θ serão os valores $\hat{N}$ e $\hat{\theta}$ que correspondem ao máximo de $LL(N)_{\theta^*}$, isto é, $\hat{N}$ e $\hat{\theta}$ satisfazem

$$LL(\hat{N})_{\hat{\theta}} = \max\{LL(N)_{\theta^*}, N \in \mathcal{G}, \theta^* \in \Theta\}. \quad (2.4)$$

Pode-se usar a equação 2.3 para obter sugestões para os valores da grade.

Pode-se construir um intervalo de confiança para N invertendo o teste da razão de verossimilhança. A idéia é relacionar o teste da razão de verossimilhança com um intervalo de confiança, usando o raciocínio tradicional que relaciona intervalos de confiança a testes. Um intervalo de confiança com coeficiente de confiança $100(1 - \alpha)\%$ ($0 < \alpha < 1$) para N consiste, de acordo com esse raciocínio, dos valores N_0 de N tais que o teste da razão de verossimilhança não rejeite a hipótese nula. Ou seja,

$$2 \{LL(\hat{N})_{\hat{\theta}} - LL(N_0)_{\hat{\theta}}\} \leq \chi_{\alpha}^2,$$

onde χ_{α}^2 é o percentil de ordem $100(1 - \alpha)\%$ da distribuição qui-quadrado com 1 grau de liberdade. Para maiores detalhes, consulte Aitken et al. (1989).

Note que o intervalo obtido é condicional à distribuição do tempo de incubação utilizada. Desta forma, existe uma incerteza associada à distribuição do tempo de incubação que não é refletida neste intervalo de confiança para N .

Para tentar incorporar esta incerteza, um caminho seria atribuir distribuições *a priori* para os parâmetros da distribuição do período de incubação e construir intervalos de credibilidade para estes parâmetros. Usando os limites destes intervalos, seriam calculadas novas estimativas de N e, com base nelas, o intervalo para N poderia ser ampliado. Neste caso, porém, seria difícil avaliar o coeficiente de confiança do intervalo obtido.

Para calcular $\hat{X}_{D+1}$, o número de casos diagnosticados após T_D que foram contaminados até T_D , basta usar a expressão

$$\hat{X}_{D+1} = \hat{N} \int_{T_0}^{T_D} \hat{I}(s) \{1 - F(T_D - s)\} ds. \quad (2.5)$$

Pode-se obter um intervalo de variação para $\hat{X}_{D+1}$ substituindo os limites do intervalo de confiança construído para N na expressão 2.5. Porém, caso o tamanho da amostra não seja grande pode ser trabalhoso associar um coeficiente de confiança ao intervalo obtido, já que não foi levado em consideração o erro associado à estimativa da curva de contaminação.

Deve ser observado que, na expressão 2.5, não são considerados casos diagnosticados cujas contaminações se deram após T_D , o que significa que

$\tilde{X}_{D+1}$ é a estimativa de uma cota inferior do número de casos que devem ser diagnosticados em $[T_D, T_{D+1})$.

Todo o processo de estimação se baseia na expressão 2.1, que relaciona π_j e θ . Por isso, quanto mais simples for a integral que envolve $I(s)$ nesta expressão, mais simples será também o algoritmo de cálculo das estimativas.

2.4 Escolha da densidade de contaminação

Para usar o “back-calculation”, não há necessidade de definir completamente a densidade de contaminação, bastando que se estabeleça uma forma paramétrica para ela. Como foi visto na seção 2.3, os parâmetros podem ser estimados dentro do próprio método. Uma das formas que têm sido utilizadas é a de escada, usada por Brookmeyer e Gail (1988), Brookmeyer e Damiano (1989) e Mariotto et al. (1992), onde são definidos arbitrariamente os pontos de descontinuidade $A_0, A_1, \dots, A_k$, para algum k fixo. As alturas dos degraus, $\theta_1, \theta_2, \dots, \theta_k$, são os parâmetros que precisam ser estimados. Nesse caso,

$$I(s) = \begin{cases} \theta_1, & \text{se } s \in [A_0, A_1) \\ \theta_2, & \text{se } s \in [A_1, A_2) \\ \vdots & \\ \theta_k, & \text{se } s \in [A_{k-1}, A_k) \\ 0, & \text{se } s < A_0 \text{ ou } s \geq A_k \end{cases} \quad (2.6)$$

Note que quanto maior for o número de intervalos, melhor será a aproximação da função escada à “verdadeira” forma da função densidade. Em contrapartida, aumentando o número de parâmetros do modelo aumenta-se também a variância dos estimadores. Em Rosenberg, Biggar et al. (1991) foi usada uma função escada com o número de degraus k fixo, onde além das alturas $\theta_1, \theta_2, \dots, \theta_k$, foram estimados também os pontos de descontinuidade $A_1, A_2, \dots, A_k$.

Outros autores optaram por distribuições de probabilidade tradicionais. Longini et al. (1992) usaram a distribuição log-logística. Em Brookmeyer e Gail (1986) são usadas a distribuição log-logística e uma função escada, e os ajustes obtidos são comparados. O ajuste da log-logística se mostra mais influenciado pelas taxas de contaminação do início da epidemia (das quais se tem mais informação) que pelas do passado recente (das quais se tem menos informação).

A utilização de uma função escada baseia-se na hipótese de a curva de contaminação ser constante em intervalos. Uma generalização desta idéia são as funções “splines”, em que considera-se um polinômio do tempo de grau k em cada intervalo, $k = 0, 1, \dots$. Quando $k = 0$, a “spline” é uma função escada. Quando $k = 1$, a função é chamada na literatura de “piecewise linear function”. Esse tipo de função é mais flexível que a logística e a log-logística porque as estimativas no fim do período observado não dependem das do início da curva. Por isso, na literatura as funções “splines” têm sido referenciadas como *modelos fracamente paramétricos*, enquanto funções tais como a logística e a log-logística têm sido chamadas de *modelos fortemente paramétricos*. Em Brookmeyer e Gail (1994) pode ser encontrada uma descrição mais detalhada dessas funções.

Diferente dos demais, o objetivo do artigo de Marion e Schechter (1993) não é obter projeções da incidência mas sim, avaliar a sensibilidade do método às escolhas da distribuição do tempo de incubação e da forma da densidade de contaminação. No caso da densidade de contaminação, consideram a princípio quatro famílias paramétricas: funções escada, funções lineares em intervalos, funções log-logísticas e logísticas (note que esta última é um caso limite da log-logística quando a origem tende para $-\infty$). Esses autores notam que as funções escada estimadas são bastante sensíveis às escolhas dos pontos de descontinuidade. Comparam os ajustes obtidos com o de um modelo Poisson saturado para o número de casos observados. Baseado no teste da razão de verossimilhança, concluem que a distribuição logística se ajusta melhor que a distribuição log-logística, independentemente da modelagem usada para a distribuição do tempo de incubação.

Já Kuo et al. (1991) usaram uma função cuja forma é consistente com a indicada pelos modelos de sistemas dinâmicos usados em epidemiologia. Esta função tem crescimento inicial exponencial, atinge um máximo e tende para zero quando $t \rightarrow \infty$.

2.5 Correção do atraso de notificação

Desde o momento em que é feito o diagnóstico até a notificação chegar ao serviço de Vigilância Sanitária decorre um certo tempo, chamado de *atraso de notificação* (“reporting delay”), que varia de caso a caso. O atraso faz com que haja subregistro da incidência recente, como pode ser visto nos dados

usados em Arcà et al. (1992) e Zeger et al. (1989), por exemplo.

Alguns autores têm tentado corrigir este problema estimando a distribuição do tempo de atraso e corrigindo os dados, o que é chamado *correção do atraso de notificação* (“adjustment for reporting delay”).

Note que o problema do atraso de notificação é inerente aos dados e, portanto, deve ser considerado sempre que o método de análise usar o número de casos diagnosticados de AIDS. Rude (1993), que usou o mesmo método, contornou esse problema desprezando as informações correspondentes aos dois últimos anos, quando o efeito do atraso é mais significativo. Gartner e Mignon (1991) ajustaram um modelo original que usa diretamente o número de casos notificados de AIDS, ao invés do número de casos diagnosticados. O método “back-calculation” se baseia na maneira como a doença se desenvolve, se referindo especialmente aos instantes de soroconversão e diagnóstico de AIDS, portanto a correção do atraso de notificação nesse caso é imprescindível.

O método de estimação da distribuição do atraso de notificação usado em Brookmeyer e Damiano (1989) será descrito agora. Comentários sobre outros trabalhos que estimam o atraso serão feitos no fim desta seção.

Os instantes em que os diagnósticos são realizados são classificados em intervalos delimitados pelos pontos $T_0, T_1, \dots, T_D$, da mesma forma que na seção 2.2. Assim, o j -ésimo intervalo é $[T_{j-1}, T_j)$, $j = 1, 2, \dots, D$. O instante T_D é o último instante para o qual se tem dados da incidência, ou seja, só conhecemos os casos de AIDS que foram notificados ao sistema de Vigilância Sanitária até T_D .

Os atrasos de notificação são classificados em intervalos $[d_{i-1}, d_i)$, $i = 1, 2, \dots, I$, onde $d_0 = 0$ é o menor atraso possível e d_I é o maior atraso observado. Seja X_{ij} o número de casos diagnosticados em $[T_{j-1}, T_j)$, cujo atraso de notificação está no intervalo $[d_{i-1}, d_i)$. Para cada valor de j , X_{ij} só é conhecido completamente se o extremo superior do intervalo $[T_{j-1}, T_j)$ somado ao maior atraso de notificação considerado for menor ou igual a T_D , ou seja, se $T_j + d_i \leq T_D$. Equivalentemente, se D_j for o maior valor de i tal que $T_j + d_{D_j} \leq T_D$, então para cada j só conheceremos X_{ij} completamente se $1 \leq i \leq D_j$. A Tabela 2.1 ilustra a parametrização usada.

Se um caso tiver atraso de notificação maior que o maior atraso observado (d_I), considera-se que seu atraso é igual a ∞ , ou seja, que nunca será notificado.

Tabela 2.1: Número de observações

Atraso no intervalo	Diagnóstico no intervalo					
	$[T_0, T_1)$	$[T_1, T_2)$	$\dots$	$[T_{j-1}, T_j)$	$\dots$	$[T_{D-1}, T_D)$
$[d_0, d_1)$	X_{11}	X_{12}	$\dots$	X_{1j}	$\dots$	X_{1D}
$[d_1, d_2)$	X_{21}	X_{22}	$\dots$	X_{2j}	$\dots$	X_{2D}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
$[d_{i-1}, d_i)$	X_{i1}	X_{i2}	$\dots$	X_{ij}	$\dots$	X_{iD}
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
	X_{D1}	X_{D2}	$\dots$	X_{Dj}	$\dots$	X_{DD}
Total	$X_{.1}$	$X_{.2}$	$\dots$	$X_{.j}$	$\dots$	$X_{.D}$

Note que, para fazer a correção, os intervalos em que os atrasos são categorizados devem ser do mesmo tamanho que os intervalos usados para os diagnósticos. Suponha, por exemplo, que os casos são observados anualmente mas que os atrasos são categorizados em semestres. Não faria sentido corrigir dados anuais usando uma probabilidade de atraso de até 6 meses. No mínimo haveria a necessidade de agrupar as probabilidades em anos (probabilidade de atraso de até um ano, probabilidade de atraso de 1 a 2 anos, etc.). Por outro lado, suponha que os casos são observados semestralmente mas que os atrasos são categorizados anualmente. Para corrigir os dados, seria usada uma correção anual sobre dados semestrais, o que novamente não faz sentido.

O número total de casos diagnosticados no j -ésimo intervalo e notificados até T_D é:

$$X_{.j} = \sum_{i=1}^{D_j} X_{ij}.$$

Para estimar a distribuição do atraso de notificação, vamos definir:

$$p_i = P\{\text{atraso estar no intervalo } [d_{i-1}, d_i)\}, \quad \text{onde } \sum_{i=1}^I p_i = 1.$$

Note que as probabilidades associadas à distribuição do atraso de notificação, $\mathbf{p} = \{p_1, p_2, \dots, p_I\}$, se referem apenas a casos com atraso menor ou igual a d_I .

Considere a distribuição de X_{ij} quando esta variável tiver sido observada, isto é, quando $1 \leq i \leq D_j$. Para cada j , a distribuição condicional de X_{ij} , dado $X_{.j}$, é multinomial com tamanho de amostra $X_{.j}$ e probabilidades

$$p_{ij} = \frac{p_i}{\sum_{k=1}^{D_j} p_k}, \quad i = 1, 2, \dots, D_j.$$

Estas são as probabilidades condicionais dos atrasos menores ou iguais a d_{D_j} . Assim, a verossimilhança condicional de $\{X_{ij}, i = 1, 2, \dots, D_j, j = 1, 2, \dots, D\}$, dados $X_{.j}, j = 1, 2, \dots, D$, é

$$L_c = \prod_{j=1}^D \left\{ \frac{X_{.j}!}{\prod_{k=1}^{D_j} X_{kj}!} \prod_{i=1}^{D_j} p_{ij}^{X_{ij}} \right\}.$$

Uma alternativa a L_c é a própria verossimilhança incondicional L_u , que envolve o parâmetro desconhecido n_j , o número total de casos diagnosticados no intervalo $[T_{j-1}, T_j)$, $j = 1, 2, \dots, D$:

$$L_u = \prod_{j=1}^D \left\{ \frac{n_j!}{\prod_{i=1}^{D_j+1} X_{ij}!} \left(\prod_{i=1}^{D_j} p_i^{X_{ij}} \right) \left(1 - \sum_{i=1}^{D_j} p_i \right)^{n_j - X_{.j}} \right\},$$

onde X_{D_j+1} é definido como $n_j - X_{.j}$, que representa o número de casos diagnosticados no intervalo j que não foram notificados até T_D . Sanathanan (1972) mostrou que os estimadores $\hat{n} = (\hat{n}_1, \hat{n}_2, \dots, \hat{n}_D)$ e $\hat{p}$ que maximizam L_u são assintoticamente equivalentes aos estimadores baseados em L_c .

As estimativas baseadas em L_c podem ser obtidas maximizando L_c em relação a p , obtendo então $\hat{p}$, e isto pode ser feito usando rotinas computacionais para regressão Poisson, tais como a do GLIM (Payne, 1985). Por isso, pode ser vantajoso sob o ponto de vista computacional trabalhar com L_c , ao invés de L_u . A estimativa de n_j é então calculada a partir de $\hat{p}$, através da relação $\hat{n}_j = X_{.j} / (1 - \sum_{i=1}^{D_j} \hat{p}_i)$.

A estimativa obtida seria a mesma se L_u fosse maximizada diretamente. A correspondência entre os dois modelos vem do seguinte fato: se X_i , $i = 1, \dots, n$, são variáveis aleatórias independentes e identicamente distribuídas, com distribuição Poisson então, condicionalmente a $\sum_{i=1}^n X_i$, a distribuição conjunta da sequência $(X_i, i = 1, \dots, n)$ é Multinomial.

Para calcular as estimativas $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_I$ a partir da verossimilhança condicional usando o GLIM, declara-se que X_{ij} tem distribuição Poisson

com média $\exp(\alpha_i + \beta_j)$, $i = 1, 2, \dots, D_j$ e $j = 1, 2, \dots, D$. Ou seja, a distribuição do erro é Poisson e a função de ligação é o logaritmo. A partir das estimativas $\hat{\alpha}_i$ e $\hat{\beta}_j$, a estimativa de p_i é

$$\hat{p}_i = \frac{\exp\{\hat{\alpha}_i\}}{\sum_{k=1}^J \exp\{\hat{\alpha}_k\}}.$$

Note que o modelo ajustado supõe que o efeito do ano de diagnóstico é aditivo ao efeito da categoria do atraso, ou seja, dado o ano de diagnóstico, a distribuição condicional do atraso de notificação é a mesma.

O modelo de regressão Poisson também pode ser usado para avaliar o efeito de covariáveis sobre a distribuição do atraso de notificação, através da inclusão de termos adicionais no preditor linear.

Estimada a distribuição do atraso, os dados são corrigidos a partir da seguinte relação: o número de casos diagnosticados em $[T_{j-1}, T_j)$ com atraso menor que d_i é igual ao número total de casos diagnosticados neste intervalo multiplicado pela probabilidade de haver tal atraso. Portanto, o número de casos efetivamente diagnosticados no intervalo $(T_D - d_m, T_D - d_{m-1})$ é encontrado dividindo o número notificado de casos diagnosticados neste intervalo pelo fator

$$\sum_{i=1}^{m-1} \hat{p}_i + \frac{\hat{p}_m}{2}, \quad m = 1, 2, \dots, D_j, \quad j = 1, 2, \dots, D. \quad (2.7)$$

A fim de facilitar a compreensão, vamos imaginar que o número de casos de AIDS diagnosticados por ano em certa população foi observado até o fim de 1992, onde o atraso também foi categorizado anualmente, e que a distribuição do atraso foi estimada pelo método descrito. Os casos diagnosticados em 1991 conhecidos têm atraso de notificação menor que 2 anos. Portanto, o número de casos efetivamente diagnosticados em 1991 pode ser estimado dividindo o número observado por $\hat{p}_1 + \hat{p}_2/2$, onde $\hat{p}_1$ é a probabilidade estimada de atraso de até 1 ano e $\hat{p}_2$ é a probabilidade de atraso entre 1 e 2 anos.

A divisão por 2 no segundo termo de 2.7 é feita porque casos com atraso de notificação classificado no intervalo i , $[I_{i-1}, I_i]$, têm um atraso médio igual ao ponto médio do intervalo, $(I_{i-1} + I_i)/2$, se a distribuição dos atrasos no intervalo for uniforme (sem perda de generalidade, supõe-se que os intervalos tenham comprimentos sejam iguais a 1). Supondo que a probabilidade de atraso seja constante em todo o intervalo, a probabilidade de atraso cor-

respondente à primeira metade do intervalo seria igual à probabilidade de atraso no intervalo todo dividida por 2.

Os primeiros trabalhos que abordaram o problema do atraso de notificação surgiram em meados da década de 80. Curran et al. (1985) foram os primeiros a corrigir os dados para o atraso de notificação, e seus dados corrigidos foram usados mais tarde por Brookmeyer e Gail (1988). Harris (1987) formulou o problema em termos de uma distribuição multinomial incompleta (consulte também Downs et al., 1987), e este método de ajuste foi estendido mais tarde por Brookmeyer e Damiano (1989), chegando então a um método de fácil implementação. Ele se baseia na verossimilhança condicional para estimar a distribuição do tempo de atraso de notificação. A facilidade computacional vem do fato de podermos maximizar a verossimilhança condicional usando algoritmos computacionais padrões para regressão Poisson, tais como aqueles incorporados no GLIM. Além disso, não há necessidade de atribuir qualquer forma paramétrica à distribuição de incidência de AIDS. Também Zeger et al. (1989) estimaram a distribuição do atraso, modelando a incidência de AIDS como um processo de Poisson não-homogêneo. Este modelo pode conduzir a estimativas mais precisas da distribuição do atraso que o modelo usado por Brookmeyer e Damiano (1989), porém neste caso é necessário atribuir uma forma funcional à média do processo de Poisson.

Capítulo 3

Aplicação ao Estado de São Paulo

3.1 Dados e Correção do Atraso de Notificação

Os dados foram obtidos junto ao Centro de Referência e Treinamento de AIDS do Estado de São Paulo (CRT-AIDS), a partir de boletins mensais publicados desde abril de 1988 até março de 1993. Neste trabalho não foi considerada uma eventual subnotificação de casos de AIDS. Uma maneira de avaliar essa subnotificação seria comparar os registros de óbitos em decorrência de AIDS com os diagnósticos notificados, caso a caso. Se a subnotificação for estimada e for constante no tempo, isto é, se a porcentagem de casos subnotificados for sempre a mesma, os resultados obtidos neste trabalho (estimativa do tamanho da epidemia e projeções do número de casos diagnosticados) poderiam ser corrigidos dividindo-os pela porcentagem de notificação.

Cada caso notificado é registrado segundo o ano em que foi diagnosticado e, a cada mês, o número de notificações já feitas é atualizado, a partir de novas notificações e óbitos. Os dados estão relacionados resumidamente no Apêndice 1, por ano de diagnóstico e ano de notificação. Analisando a tabela A.1, verifica-se que uma porcentagem grande (aproximadamente 40%) de casos não são notificados no mesmo ano do diagnóstico.

De acordo com os dados do CRT-AIDS, o primeiro caso de AIDS no estado de São Paulo foi diagnosticado em 1980. Até março de 1993, não

havia nenhum caso notificado com diagnóstico em 1993.

Para casos notificados a partir de maio de 1988 é possível saber o mês de notificação. Por exemplo, se em junho de 1990 havia n_1 casos notificados com diagnóstico em 1989 e em julho este número passou a n_2 , significa que $n_2 - n_1$ casos com diagnóstico em 1989 foram notificados em julho de 1990. Já da data do diagnóstico só se conhece o ano. Por isso, para fins de correção dos dados não interessa estimar a distribuição do atraso mensalmente, posto que a correção será feita sobre dados anuais.

A classificação do atraso foi feita da seguinte maneira: um caso é classificado com atraso i se, dado que foi diagnosticado no ano j , sua notificação ocorreu no ano $i + j$, $i = 0, 1, \dots, I - 1$, $j = 1982, 1983, \dots, 1992$. Casos classificados com atraso I são aqueles notificados pelo menos I anos após o diagnóstico. Note que essa classificação é qualitativa, já que não sabemos exatamente com quanto tempo de atraso um caso foi notificado - apenas registramos se o caso foi notificado no mesmo ano do diagnóstico, ou no ano seguinte ao do diagnóstico, e assim por diante. Isto não é um problema do ponto de vista da modelagem, porque o modelo Poisson proposto em Brookmeyer e Damiano (1989) apenas supõe que o atraso possa ser categorizado, comportando tanto classificações quantitativas quanto qualitativas. E, sob o ponto de vista de estimação, o fato de essa variável ser qualitativa também não é um problema, porque o GLIM (Payne, 1985) está preparado para considerar essa possibilidade.

Só uma parte dos dados foi considerada na estimação da distribuição do atraso. As observações desprezadas correspondem a casos já notificados em abril de 1988, cuja data de notificação (e portanto o atraso) é desconhecida, e casos notificados em 1993. Esses últimos foram desconsiderados porque só foram obtidos os três primeiros boletins desse ano, o que inviabilizou a totalização do número de casos.

Considerando apenas os casos notificados entre maio de 1988 e dezembro de 1992, a distribuição do atraso foi estimada e os dados foram corrigidos, obtendo o que vamos chamar de Correção 1. A estimação foi baseada no método descrito na seção 2.5 e o pacote utilizado foi o GLIM (Payne, 1985).

Examinando os dados usados na estimação verifica-se que os casos diagnosticados antes de 1988 estão censurados. Por exemplo, dos casos diagnosticados em 1986 só estão sendo considerados aqueles notificados a partir de maio de 1988, ou seja, cujo atraso é maior ou igual a 2 anos, o que equivale a tomar a probabilidade de atraso menor que 2 anos igual a zero, para di-

agnósticos em 1986. Esta censura introduz um desvio artificial indesejado nas estimativas.

O problema foi contornado considerando na estimação apenas os casos com diagnóstico mais recente, a partir de 1987, obtendo o que vamos chamar de Correção 2. Note que ainda há censura sobre os casos diagnosticados em 1987 (não se conhece a data de notificação dos casos notificados antes de abril de 1988). Neste caso, a distribuição do atraso estimada é decrescente, como se esperava.

Outra forma de resolver este problema seria considerar que as probabilidades de atraso correspondentes às observações censuradas (como os casos com atraso de até 2 anos, diagnosticados em 1986) sejam iguais às das observações não-censuradas (casos diagnosticados a partir de 1988). Desta forma, apenas casos diagnosticados de 1988 em diante seriam considerados para estimar as probabilidades correspondentes a atraso de até 1 ano, por exemplo. Para estimar a probabilidade de atraso entre 1 e 2 anos podem ser considerados os casos diagnosticados a partir de 1987, e assim por diante.

Como foi dito, ainda há censura nesta situação porque os casos diagnosticados em 1987 têm atraso de pelo menos 1 ano. É feita agora a Correção 3, considerando apenas os casos com diagnóstico a partir de 1988.

Para avaliar a precisão dos resultados obtidos, a distribuição do atraso foi recalculada usando a estimativa somada e subtraída do desvio-padrão correspondente. Os resultados numéricos estão no Apêndice. Tanto em relação às distribuições estimadas quanto em relação aos dados corrigidos, a variação provocada foi pequena. A maior variação foi observada na Correção 1, onde a observação corrigida de 1992 oscilou entre 17483 e 17976, o que representa uma variação de 2,8%.

Na Tabela 3.1, podemos comparar as 3 estimativas da distribuição do atraso. As estimativas estão muito próximas. Da terceira probabilidade de atraso em diante a distribuição estimada é praticamente constante nos três ajustes. A distribuição estimada na Correção 1 não é decrescente, ao contrário do que se espera. Esse desvio pode ter sido causado pelo fato de os dados estarem censurados, já que nas correções 2 e 3 isso não se repetiu. Apesar da não monotonicidade, a variação da distribuição estimada é pequena em valores absolutos (de 0,014 para 0,027).

Analisando os dados corrigidos (Tabela 3.2), nota-se que da Correção 1 para a Correção 2 os valores corrigidos diminuíram 4% e, da Correção 2

Tabela 3.1: Distribuições de atraso estimadas

Correção 1	Correção 2	Correção 3	Interpretação do parâmetro
0,631	0,655	0,668	$P\{\text{notif. no mesmo ano que o diag.}\}$
0,231	0,239	0,249	$P\{\text{notif. no ano seguinte ao diag.}\}$
0,043	0,045	0,042	$P\{\text{notif. dois anos após o diag.}\}$
0,024	0,025	0,022	$P\{\text{notif. três anos após o diag.}\}$
0,019	0,021	0,019	$P\{\text{notif. quatro anos após o diag.}\}$
0,014	0,015	0,000	$P\{\text{notif. cinco anos após o diag.}\}$
0,027	0,000	0,000	$P\{\text{notif. seis anos após o diag.}\}$
0,011	0,000	0,000	$P\{\text{notif. sete ou mais anos após diag.}\}$

para a Correção 3, a redução foi de 2%. Além disso, a redução relativa foi aproximadamente a mesma em todos os anos. Essa redução era esperada, já que de uma correção para outra cada vez um número menor de anos é considerado, e os anos desprezados correspondem aos casos com os maiores atrasos.

A Correção 1, talvez por ser originada de dados com atrasos maiores em média, produz um acréscimo maior sobre os dados observados que as outras duas.

Com as 3 correções, o número corrigido de casos diagnosticados em 1992 é aproximadamente 3 vezes maior que o número observado. Pode estar havendo superestimação do número de casos atrasados. A magnitude das observações envolvidas neste trabalho é bem maior que a de outros trabalhos, talvez pelo fato dos dados serem anuais, ao invés de mensais, trimestrais ou quadrimestrais, como na maior parte dos estudos. Portanto, um vício nos estimadores fica mais evidente nesse caso. Além disso, se houver de fato um vício na escala do logaritmo dos dados, ele tende a ser muito maior na escala original.

Tabela 3.2: Dados observados e dados corrigidos

Ano	1985	1986	1987	1988	1989	1990	1991	1992
Observado	65	99	438	2524	2964	4165	5399	5586
Correção 1	65	102	456	2691	3235	4717	7236	17713
Correção 2			441	2588	3113	4542	6967	17049
Correção 3				2549	3058	4443	6817	16734

Para aplicação do método back-calculation foi usada a Correção 3, por só considerar dados não-censurados (sob o ponto de vista de ano de diagnóstico e mês de notificação). Daqui por diante, referências feitas aos dados corrigidos correspondem à Correção 3. As observações correspondentes aos anos de 1977 até 1987 não sofreram correção. Só foram considerados os casos notificados até dezembro de 1992.

3.2 Escolha da distribuição do tempo de incubação

Vários artigos na literatura epidemiológica e estatística apresentam estimativas e métodos de estimação da distribuição do tempo de incubação. Muitas dessas estimativas estão baseadas em estudos de coortes de grupos de risco, tais como hemofílicos, homens homossexuais ou bissexuais. O problema em muitos estudos de coorte é determinar a data de contaminação de cada participante que, geralmente, é desconhecida. Além disso, a amostra é truncada no sentido de que um indivíduo que poderia ter participado do estudo mas desenvolveu a doença antes do recrutamento é excluído da amostra. Desta forma, o critério de seleção vicia o estudo em relação a períodos de incubação curtos (Brookmeyer et al., 1987). Muitos autores usaram casos de AIDS associados a transfusões para estimar a distribuição do tempo de incubação (veja por exemplo Lui et al., 1986, Medley et al., 1987 e Kalbfleisch e Lawless, 1989), pois neste caso a data de contaminação é conhecida. Entretanto, como os conjuntos de dados excluem os indivíduos que não desenvolveram a doença até o fim do estudo, é mais fácil ter uma idéia da forma da densidade do tempo de incubação que da função de distribuição acumulada. Isto porque um histograma dos tempos de incubação observados pode dar

uma idéia sobre a forma da densidade, mesmo que haja censura sobre as observações extremas mas, no caso da função de distribuição acumulada, sua forma pode se modificar bastante, dependendo da probabilidade associada à observações extremas. Além disso, a exclusão das observações maiores implica subestimação dos momentos da distribuição.

A partir dos trabalhos que estimam a distribuição do tempo de incubação foram escolhidas duas distribuições, pertencentes à família Weibull (α, β) , cuja densidade é dada por

$$f(t) = \alpha \beta t^{\beta-1} e^{-\alpha t^\beta}, \quad t > 0, \quad \alpha > 0, \quad \beta > 0.$$

A função de distribuição acumulada é:

$$F(t) = 1 - e^{-\alpha t^\beta}, \quad t > 0.$$

Supor que a distribuição do tempo de incubação é Weibull equivale a supor que alguma potência do tempo de incubação tem distribuição exponencial. A potência é o parâmetro de forma da distribuição Weibull, representado aqui por β .

O método “back-calculation” foi aplicado usando cada uma das distribuições escolhidas e os resultados foram comparados.

A primeira distribuição usada foi a distribuição Weibull (0,1077 ; 2,396) (que será referida como Incubação 1), estimada em Medley et al. (1987) usando casos de indivíduos residentes nos EUA contaminados por transfusão entre abril de 1978 e fevereiro de 1986, e com diagnósticos entre janeiro de 1982 e junho de 1986. De um total de 494 casos só foram considerados 297, que correspondiam a indivíduos que receberam uma única transfusão (ou algumas num curto período) de sangue (ou derivados) contaminado. A data de transfusão (ou transfusões) forneceu uma aproximação do instante de contaminação. Usaram tanto a família de distribuições gama quanto a distribuição Weibull, obtendo ajustes igualmente bons nos dois casos.

A outra distribuição usada é a distribuição Weibull (0,0243 ; 2,286) (Incubação 2), estimada por Lui et al. (1986) e também usada em Brookmeyer e Gail (1986). Esta distribuição também foi obtida a partir de uma análise de casos de AIDS dos EUA associados a transfusão.

Gráfico 1: Funções densidade de incubação sugeridas por outros trabalhos:
 — Incubação 1 ; — Incubação 2.

No Gráfico 1 pode-se avaliar conjuntamente as duas distribuições. Note que a Incubação 1 tem menor mediana (igual a 2,2 anos) e é mais concentrada. Segundo esta distribuição, existe uma probabilidade muito pequena (0,6%) de que o período de incubação dure mais que 5 anos. Já a mediana da Incubação 2 é de 4,3 anos, e a probabilidade de que o tempo de incubação dure mais que 5 anos é de 38%. Porém, mesmo a Incubação 2 prevê uma chance pequena de que o tempo de incubação dure mais que 10 anos (0,9%).

Usar mais de uma distribuição para o tempo de incubação é importante para avaliar a sensibilidade do método à escolha dos valores dos parâmetros.

3.3 Densidade de contaminação

Foi usada uma função escada para descrever a contaminação. A vantagem de usar esta forma paramétrica é a sua flexibilidade, pois qualquer função contínua pode ser aproximada por uma função escada. Quanto maior for o número de patamares, melhor será a aproximação. Por outro lado, aumentar o número de parâmetros implica aumento do erro-padrão das estimativas. Além disso, alguns autores (veja por exemplo Rude, 1993) têm notado que

os resultados são muito sensíveis à particular escolha dos pontos de salto, particularmente no passado recente. Assim, é preciso avaliar mais de uma divisão do período observado, a fim de avaliar o impacto da divisão usada sobre as projeções.

Dividir o período observado, de 1977 a 1993, em 7 ou 8 intervalos parece ser razoável. Brookmeyer e Gail (1988) usam 4 intervalos para o período que vai de 1977 a 1987, e o comprimento dos intervalos variou de 1,5 anos até 4 anos.

Acredita-se que as contaminações tenham ocorrido a taxas maiores nos últimos anos que nos primeiros anos da epidemia. Por isso, para os primeiros anos usaremos intervalos maiores que para os últimos anos.

O período em estudo foi dividido em intervalos de três maneiras diferentes, variando tanto o número de intervalos quanto seus comprimentos. Na primeira divisão são usados 7 intervalos, ficando a densidade de contaminação parametrizada como se segue:

$$I_1(s) = \begin{cases} \theta_1, & \text{se } s \in [1977, 1982) \\ \theta_2, & \text{se } s \in [1982, 1984) \\ \theta_3, & \text{se } s \in [1984, 1986) \\ \theta_4, & \text{se } s \in [1986, 1988) \\ \theta_5, & \text{se } s \in [1988, 1990) \\ \theta_6, & \text{se } s \in [1990, 1992) \\ \theta_7, & \text{se } s \in [1992, 1993) \\ 0, & \text{se } s < 1977 \text{ ou } s \geq 1993 \end{cases} \quad (3.1)$$

A segunda divisão consiste também em 7 intervalos e se diferencia da primeira porque os dois últimos intervalos têm o mesmo comprimento, 1 ano e 6 meses:

$$I_2(s) = \begin{cases} \theta_1, & \text{se } s \in [1977, 1982) \\ \theta_2, & \text{se } s \in [1982, 1984) \\ \theta_3, & \text{se } s \in [1984, 1986) \\ \theta_4, & \text{se } s \in [1986, 1988) \\ \theta_5, & \text{se } s \in [1988, 1990) \\ \theta_6, & \text{se } s \in [1990, \text{junho}/1991] \\ \theta_7, & \text{se } s \in [\text{julho}/1991, 1993) \\ 0, & \text{se } s < 1977 \text{ ou } s \geq 1993 \end{cases} \quad (3.2)$$

A intenção de usar essas duas parametrizações é avaliar o efeito da escolha dos pontos de corte nos anos mais recentes. Para avaliar o efeito da divisão num passado menos recente, foi feita uma terceira divisão, onde são tomados 8 intervalos no período, onde os 4 últimos têm 1 ano e 6 meses de comprimento:

$$I_3(s) = \begin{cases} \theta_1, & \text{se } s \in [1977, 1981) \\ \theta_2, & \text{se } s \in [1981, 1983) \\ \theta_3, & \text{se } s \in [1983, 1985) \\ \theta_4, & \text{se } s \in [1985, 1987) \\ \theta_5, & \text{se } s \in [1987, \text{junho}/1988] \\ \theta_6, & \text{se } s \in [\text{julho}/1988, 1990) \\ \theta_7, & \text{se } s \in [1990, \text{junho}/1991] \\ \theta_8, & \text{se } s \in [\text{julho}/1991, 1993) \\ 0, & \text{se } s < 1977 \text{ ou } s \geq 1993 \end{cases} \quad (3.3)$$

3.4 Resultados

Usando as duas distribuições do tempo de incubação selecionadas e as três parametrizações diferentes da densidade de contaminação, o método “back-calculation” foi aplicado. Com permissão do autor, a aplicação do método foi feita usando programa elaborado por R. Brookmeyer em linguagem Fortran 77. Demais resultados, além dos analisados no texto, se encontram no Apêndice 1 (tabelas A.9-A.14).

Para cada uma das 3 formas paramétricas usadas, as estimativas dos parâmetros da curva de contaminação variaram pouco quando a distribuição do tempo de incubação foi modificada. Na Tabela 3.3, são apresentados os resultados obtidos usando a densidade de contaminação I_1 .

Analisando conjuntamente as estimativas obtidas das 3 densidades de contaminação, destaca-se o fato de que todas elas têm sua moda no último intervalo. No caso da densidade de contaminação I_1 , a probabilidade de contaminação no ano de 1992 foi estimada em 0,82 para a Incubação 1 e 0,86 para a Incubação 2. Para I_2 e I_3 , a probabilidade estimada de contaminação no período julho/91-dezembro/92 foi bem menor, ficando em aproximadamente 0,50, para ambas as distribuições de incubação. Talvez esse salto desproporcional no último intervalo de I_1 seja devido à proporção de casos de AIDS

Tabela 3.3: Densidade de contaminação I_1 estimada

Incubação		Interpretação do parâmetro
1	2	
0,0000	0,0000	$P\{\text{contam. entre 01/01/1977 e 01/01/1982}\}$
0,0004	0,0004	$P\{\text{contam. entre 01/01/1982 e 01/01/1984}\}$
0,0030	0,0029	$P\{\text{contam. entre 01/01/1984 e 01/01/1986}\}$
0,0127	0,0093	$P\{\text{contam. entre 01/01/1986 e 01/01/1988}\}$
0,0171	0,0031	$P\{\text{contam. entre 01/01/1988 e 01/01/1990}\}$
0,0545	0,0535	$P\{\text{contam. entre 01/01/1990 e 30/06/1992}\}$
0,8246	0,8615	$P\{\text{contam. entre 01/07/1992 e 01/01/1993}\}$

correspondentes ao período, em relação ao número total de casos (16734 em 35897, que representa 47%). No caso de I_2 e I_3 , talvez o salto não tenha sido tão grande no último intervalo porque seu comprimento é maior. Isso permitiria uma acomodação melhor das probabilidades de contaminação no último intervalo, tornando menor a estimativa da densidade de contaminação no intervalo.

Observe também que a modificação da distribuição do tempo de incubação não provocou impacto significativo sobre as estimativas da curva de contaminação. Para cada parametrização usada, as estimativas apresentaram o mesmo padrão usando as duas distribuições consideradas.

Deve-se ressaltar também que a estimativa do tamanho da epidemia ($\hat{N}$) e as projeções baseadas na densidade de contaminação I_1 são substancialmente maiores (aproximadamente 80%) que as baseadas nas outras densidades (veja Tabelas 3.4 e 3.5). Esta diferença não parece ser devida ao número de intervalos da densidade de contaminação (já que I_1 e I_2 possuem ambas 7 intervalos), mas sim ao comprimento do último intervalo, que no caso de I_1 é de 1 ano e, para I_2 e I_3 , é de 1 ano e 6 meses. Já mudanças no comprimento de intervalos anteriores causam variação bem menor nas estimativas, como

Tabela 3.4: Projeção do número acumulado de casos de AIDS até o ano 2000

Distribuição do tempo de incubação	Densidade de contaminação		
	I_1	I_2	I_3
1 - Weibull (0,1077 ; 2,396)	262.805	146.963	147.727
2 - Weibull (0,0243 ; 2,286)	775.371	431.517	444.034

se pode notar quando são comparados os resultados baseados em I_2 com os baseados em I_3 , para ambas as distribuições do tempo de incubação.

De fato, tem-se muito pouca informação sobre as taxas de contaminação no passado mais recente, em comparação com o início da epidemia. Isso implica que as estimativas referentes ao passado mais recente têm mais incerteza associada que as do passado menos recente. Brookmeyer e Gail (1994) notaram que quando o comprimento do último intervalo diminui, aumenta o desvio-padrão associado às estimativas. Rosenberg, Gail e Pee (1991) sugerem usar para o último intervalo um comprimento de aproximadamente 4 anos. No caso brasileiro, entretanto, acredita-se que as taxas de contaminação estejam ainda crescendo, particularmente entre os heterossexuais. Por isso, não nos pareceu razoável usar um comprimento de 4 anos para o último intervalo.

Para cada combinação de distribuição de incubação e densidade de contaminação, foram calculadas projeções da incidência e da prevalência até o ano 2000, sem considerar novas contaminações após 01/01/1993 (veja Tabela 3.4). Desta forma, as projeções representam cotas inferiores do número acumulado de casos de AIDS. As estimativas baseadas na Incubação 1 (cuja mediana é igual a 2,2 anos) são mais “otimistas”, no sentido de que apontam para um número menor de contaminados até dezembro de 1992 e, em consequência, um número acumulado de casos até o ano 2000 menor que os estimados pela Incubação 2 (mediana igual a 4,3 anos).

O impacto da escolha da distribuição do tempo de incubação parece ser maior sobre a estimativa do tamanho da epidemia (número de pessoas contaminadas em 1993), $\hat{N}$ e, conseqüentemente, sobre as projeções do número de casos de AIDS (veja Tabela 3.4), que sobre as estimativas dos parâmetros

Tabela 3.5: Estimativas do tamanho da epidemia - $\hat{N}$

Distribuição do tempo de incubação	Densidade de contaminação		
	I_1	I_2	I_3
1 - Weibull (0,1077 ; 2,396)	262.806	146.963	147.728
2 - Weibull (0,0243 ; 2,286)	842.831	461.036	474.834

da densidade de contaminação. Dada uma certa forma paramétrica da densidade de contaminação, a estimativa do tamanho da epidemia parece ser maior quando a mediana do tempo de incubação é maior. Isto parece ser razoável. Imagine primeiro que a mediana do tempo de incubação seja de 2 anos. Neste caso, até dezembro de 1992 pelo menos a metade das pessoas contaminadas até janeiro de 1991 já devem ter desenvolvido a doença. Por outro lado, se a mediana do tempo de incubação for de 5 anos, até dezembro de 1992 pelo menos a metade dos indivíduos contaminados até janeiro de 1988 terão desenvolvido a doença, mas menos da metade dos contaminados a partir de então a terão desenvolvido. Em resumo, quanto maior for a mediana do tempo de incubação, maior tenderá a ser a porcentagem dos indivíduos contaminados até dezembro de 1992 que desenvolverão a doença nos próximos anos e, por consequência, o número total de contaminados.

Outro aspecto que parece influenciar a estimativa do número de contaminados pelo HIV são as estimativas dos parâmetros da densidade de contaminação, particularmente os valores correspondentes às probabilidades de contaminação nos intervalos mais recentes. As estimativas de I_1 correspondentes ao período janeiro-dezembro/92 são 72% maiores que as de I_2 e I_3 relativas ao período julho/91-dezembro/92, e prevêm números de contaminados de 70 a 80% maiores que I_2 e I_3 .

Este resultado parece fazer sentido. As distribuições consideradas para o tempo de incubação têm medianas de 2,2 e 4,3 anos, maiores que os intervalos usados para parametrizar a densidade de contaminação. Particularmente, o intervalo mais recente tem comprimento de 1 ano (I_1) ou 1 ano e 6 meses (I_2 e I_3). Desta forma, das pessoas contaminadas no último intervalo mais da metade só deve desenvolver a doença a partir de janeiro de 1993. Assim,

Tabela 3.6: Sensibilidade das estimativas a variações da mediana do tempo de incubação M

M	$\beta = 2,286$			$\beta = 2,396$		
	$\hat{N}$	$\hat{X}.(2000)$	$\hat{X}.(2000)/\hat{N}$	$\hat{N}$	$\hat{X}.(2000)$	$\hat{X}.(2000)/\hat{N}$
5	633.630	547.470	86,40%	707.718	619.777	87,57%
6	927.396	681.750	73,51%	1.054.798	785.942	74,51%
7	1.284.368	784.730	61,10%	1.504.499	919.348	61,11%
8				2.006.101	1.011.822	50,44%

$\hat{X}.(2000)$ = estim. do número acum. de casos diagnosticados até o ano 2000

quanto maior for o número estimado de pessoas contaminadas no último intervalo, maior será o total estimado de pessoas contaminadas ($\hat{N}$).

Por ser uma projeção, a porcentagem das pessoas contaminadas até dezembro de 1992 que desenvolvem a doença até o ano 2000 (representado por $\hat{X}.(2000)$) também é influenciada pela escolha da distribuição do tempo de incubação. Usando a Incubação 1, as projeções indicam que praticamente todas as pessoas contaminadas antes de 1993 desenvolvem a doença até o ano 2000. Já com a Incubação 2, a previsão é de que entre 92 e 94% dos contaminados já tenham desenvolvido AIDS no ano 2000. Essa diferença é consequência do fato de a cauda da Incubação 2 ser mais pesada, ou seja, associar maior probabilidade a tempos de incubação “grandes”, que a da Incubação 1 (veja o Gráfico 1).

É importante ainda analisar a sensibilidade das estimativas a variações da mediana do tempo de incubação. Para isto, fixou-se o parâmetro de forma (β) da distribuição Weibull e o valor do outro parâmetro foi tomado tal que a mediana fosse igual a um certo valor M . Esta análise foi feita usando os valores de β estimados nos estudos de coorte (2,286 e 2,396) e tomando a forma paramétrica I_3 da densidade de contaminação. Na Tabela 3.6 são dadas as estimativas do número de pessoas contaminadas até dezembro de 1992 e a projeção do número acumulado de casos diagnosticados até o ano 2000. A porcentagem de casos diagnosticados até o ano 2000, dentre as pessoas contaminadas até dezembro de 1992, é aproximadamente a mesma para as duas distribuições do tempo de incubação consideradas, fixado o valor

M da mediana. A variação do parâmetro de forma teve maior influência sobre a estimativa do número de pessoas contaminadas e sobre a projeção do número acumulado de casos diagnosticados. Isso era esperado, já que aumentar o parâmetro de forma da distribuição Weibull mantendo fixa a mediana implica concentrar mais a distribuição.

Para β fixo igual a 2,286, aumentar a mediana de 4,3 anos para 7 anos implicou um aumento de 170% da estimativa do número de pessoas contaminadas (passou de 474.834 para 1.284.368).

Capítulo 4

Discussão e Conclusões

O método “back-calculation” foi usado para estimar o número total de contaminados com o HIV e prever a incidência de casos de AIDS num curto prazo, no estado de São Paulo. Para modelar a densidade de contaminação usou-se uma função escada e supôs-se que a duração do período de incubação tenha uma distribuição Weibull.

A distribuição do tempo de incubação e a densidade de contaminação são duas fontes de incerteza inerentes ao método. Tentou-se avaliar essa incerteza analisando a sensibilidade dos resultados a mudanças nas suposições feitas. No que se refere à distribuição do tempo de incubação, foram modificados alternadamente o parâmetro de forma e a mediana.

As modificações na mediana do tempo de incubação tiveram maior impacto sobre a estimativa do tamanho da epidemia, N , que sobre a densidade de contaminação estimada, o que parece ser intuitivo. Aumentar a mediana do tempo de incubação implica esperar mais até que as pessoas contaminadas pelo HIV desenvolvam a doença. Mantendo fixa a incidência observada de casos de AIDS, quando se aumenta a mediana diminui-se a porcentagem de contaminados pelo HIV que já desenvolveram a doença. Em relação à densidade de contaminação, o aumento da mediana implica um deslocamento de massa de probabilidade da cauda da curva para a parte central, já que os indivíduos que desenvolveram AIDS precisam ter sido contaminados mais cedo. Esse deslocamento é mais discreto que o aumento da estimativa de N .

Em relação à densidade de contaminação, foram feitas modificações nos pontos de salto e número usado de patamares. Os resultados mostraram sensibilidade a variações dos pontos de salto da curva de contaminação, sobre-

tudo nos intervalos correspondentes ao passado mais recente. Modificações feitas em períodos mais recentes tiveram impacto muito maior sobre os resultados que modificações feitas em períodos menos recentes. Este tipo de sensibilidade também foi notado por Marion e Schechter (1992). Rude (1993) notou que os resultados eram particularmente sensíveis ao último período da série de casos de AIDS. Brookmeyer e Gail (1994) associaram essa sensibilidade ao fato de se ter pouca informação sobre as taxas de contaminação no passado mais recente, quando comparado ao início da epidemia, e isso implicaria instabilidade nas estimativas caso se utilizem intervalos “curtos” no passado recente. Eles observaram que o erro padrão associado às estimativas aumenta quando diminui o tamanho do intervalo. Por outro lado, se o comprimento do último intervalo for muito grande e as taxas de contaminação estiverem, na verdade, crescendo, as estimativas produzidas seriam muito viciadas. Uma maneira de contornar esse problema seria usar intervalos mais longos (especialmente os últimos). Rosenberg, Gail e Pee (1991) sugerem que o último intervalo deveria ter comprimento de 4 anos, argumentando que esse comprimento estabeleceria uma relação razoável entre vício e erro padrão. Outra maneira de contornar esse problema seria relaxar a escolha dos pontos de salto, fixando apenas o número de saltos da função, como feito em Rosenberg, Biggar et al. (1991). Uma análise de sensibilidade mais detalhada deste aspecto está fora do escopo deste trabalho, ficando para futura pesquisa.

A existência de atrasos de notificação é mais uma fonte de incerteza, trazendo variabilidade aos resultados, além da prevista pelo modelo. O fato de as datas de diagnóstico serem codificadas pelo ano também traz imprecisão à correção do atraso. É importante avaliar ainda as origens dos atrasos, a fim de buscar modelos que eventualmente sejam mais adequados para corrigir os dados.

Um primeiro passo nessa direção seria avaliar quanto tempo um caso leva para ser notificado no estado de São Paulo dentro dos trâmites burocráticos atuais, e se esse tempo pode variar de região para região. Em contatos informais com profissionais do CRT-AIDS, foi levantada a hipótese de o atraso depender do local em que o caso é diagnosticado. Segundo estes profissionais, casos diagnosticados na rede hospitalar da cidade de São Paulo, com exceção do Hospital Emílio Ribas, levam em média 1 mês para serem notificados. Casos diagnosticados no Hospital Emílio Ribas e no CRT-AIDS têm notificação mais ágil, levando de 7 a 15 dias em média. Já casos do interior do

estado têm notificação mais lenta, levando 2 a 3 meses, em média. Atrasos um pouco maiores (6 meses a 1 ano) podem se dever à burocracia associada à notificação. Atrasos maiores que 1 ano têm sido creditados a uma demora do próprio médico para encaminhar a notificação, por exemplo, por aguardar resultados de exames que confirmem o diagnóstico. Não são raros os casos em que a notificação do diagnóstico é feita simultaneamente à notificação do óbito.

Neste trabalho, foi usado um modelo de regressão com erro Poisson para estimar a distribuição do atraso de notificação, seguindo os passos de Brookmeyer e Damiano (1989).

Pode estar havendo superestimação do número de casos atrasados, o que explicaria a correção desproporcional sofrida pela observação correspondente ao último ano. O fato disso não ter sido apontado em outras aplicações que usam o mesmo modelo para correção do atraso de notificação poderia ser explicado pela diferença entre a magnitude das observações envolvidas. Uma superestimação dessa natureza implicaria superestimação do tamanho da epidemia ($\hat{N}$) e, conseqüentemente, das projeções da incidência e da prevalência. A avaliação quantitativa desse vício e possíveis modificações no modelo são deixados para estudos futuros.

O fator de correção dos dados, cuja última parcela (representada por p_m em 2.7) aparece dividida por 2, não foi explicitamente justificado pelos autores na referência original. Aparentemente, os autores supuseram que a probabilidade de atraso é constante no intervalo a que se refere e, supondo que os atrasos se distribuam uniformemente sobre o intervalo (dado que foram classificados naquele intervalo), a probabilidade média de atraso corresponde à probabilidade de atraso naquele intervalo dividida por 2. Nos parece, entretanto, mais intuitivo considerar a última parcela integralmente, como foi feito em Brookmeyer e Liao (1990), já que

$$\begin{aligned} \text{n. casos diag. ano } j \text{ c/ atraso } i &= \text{n. casos diag. ano } j \times \text{prob. atraso } i \\ \Rightarrow \frac{\text{n. casos diag. ano } j \text{ c/ atraso } i}{\text{prob. atraso } i} &= \text{n. casos diag. ano } j \end{aligned}$$

onde se observa o número de casos com atraso e a probabilidade de atraso é estimada. Se for considerado como “atraso i ” o evento “notificação no mesmo ano do diagnóstico”, por exemplo, parece ser razoável usar na correção dos dados a probabilidade de um caso ser notificado no mesmo ano do diagnóstico

integralmente.

A comparação da distribuição do atraso estimada neste trabalho com as de outros estudos deve ser feita com cautela, pois aqui estima-se a probabilidade da notificação ser feita no ano $i + j$, dado que o diagnóstico foi feito no ano j , $i = 0, 1, 2, \dots, I$. Desta forma, um caso diagnosticado em dezembro de 1990 e notificado em fevereiro de 1991 tem atraso classificado na categoria 1, enquanto um caso diagnosticado em janeiro de 1992 e notificado em dezembro do mesmo ano tem atraso classificado na categoria 0. Este problema ocorre porque os diagnósticos são classificados anualmente, não sendo possível calcular o atraso e corrigir os dados com erro menor.

Note que a categorização do atraso usada é qualitativa e não quantitativa. Brookmeyer e Damiano (1989) usaram originalmente o modelo Poisson para corrigir o atraso de notificação classificado quantitativamente. Mas este modelo comporta perfeitamente classificações qualitativas, desde que no processo de estimação isto seja considerado.

Em Rude (1993), o método “back-calculation” é aplicado ao caso da França, usando uma função escada para a curva de contaminação e uma distribuição Weibull com mediana de aproximadamente 10 anos para o tempo de incubação. O impacto da correção do atraso de notificação sobre os últimos períodos foi contornado considerando apenas os casos diagnosticados até junho de 1990, notificados até 1993, na aplicação do “back-calculation”. Este truncamento também atenua a imprecisão da estimativa do valor da curva de contaminação no último degrau.

O programa usado neste trabalho gerou estimativas negativas para trechos da densidade de contaminação, em algumas combinações de distribuição do tempo de incubação e forma paramétrica da densidade de contaminação. Em Brookmeyer e Gail (1994), comenta-se que uma forma paramétrica da densidade de contaminação com um número de parâmetros grande demais pode dar origem a instabilidade nos resultados, podendo até gerar esse tipo de resultado. Notamos ainda que quanto maior for a mediana do tempo de incubação, maior é a chance de obter estimativas negativas. A avaliação da estabilidade do método nesse sentido não foi feita neste trabalho, sendo deixada como tema para estudos futuros.

Como já havia sido comentado, as estimativas dos parâmetros da curva de contaminação variaram pouco quando a distribuição do tempo de incubação foi modificada, mostrando pouca sensibilidade a mudanças da duração do período de incubação.

Diversos trabalhos têm confirmado a influência na duração da incubação de fatores como sexo (veja Medley et al., 1987 e Murphy et al., 1991) e idade (além dos já citados, veja também Brookmeyer e Goedert, 1989, Biggar, 1990 e Mariotto et al., 1992). Parece haver efeito do modo de transmissão, mas ele não foi estatisticamente significativo nos estudos de Jason et al., 1989, Darby et al., 1990 e Mariotto et al., 1992. Uma maneira de tentar melhorar a qualidade dos resultados seria considerar esses efeitos.

O uso de tratamento também influencia o tempo de incubação, mas essa informação não estava disponível. Uma maneira de contornar esse problema seria considerar uma certa distribuição do tempo de incubação para contaminações ocorridas até um certo instante t , e outra para as contaminações ocorridas após t . O instante t ideal seria aquele tal que os indivíduos contaminados até t não tenham recebido tratamento antes do diagnóstico de AIDS, e que os contaminados após t o tenham recebido. O serviço público de saúde brasileiro não realiza rotineiramente testes para determinar se portadores do HIV assintomáticos têm indicação para uso de tratamento (AZT), sendo a droga geralmente administrada apenas após o diagnóstico de AIDS. Por isso, a influência do tratamento sobre o período de incubação no Brasil pode ser desprezada.

Acredita-se que o período de incubação de pessoas que residam em países do primeiro mundo tenha distribuição diferente do de pessoas residentes em países do terceiro mundo. A soroprevalência de doenças infecto-contagiosas de cada região e o comportamento da população local seriam algumas componentes do fator geográfico. Em decorrência destes fatores, é necessário haver cautela na utilização de distribuições de incubação estimadas em outras populações.

No Brasil, até 1994 não estavam disponíveis estudos de acompanhamento de coortes que avaliassem a duração do período de incubação em brasileiros aqui residentes. Por isso, neste trabalho foram usadas distribuições para o tempo de incubação estimadas em outras populações. Pelo fato de os resultados serem fortemente influenciados pela particular distribuição escolhida, a qualidade das estimativas depende do quão próximo da distribuição verdadeira está a distribuição tomada.

Para avaliar a adequação dessas distribuições ao nosso caso, truncamos a série (incidência) observada, aplicamos o método e comparamos as projeções com o que foi observado. Isso foi feito usando o ano de 1991 como ponto de corte, ou seja, desprezando a última observação. Duas divisões do período

observado foram usadas, que serão referenciadas por *Divisão 1* e *Divisão 2*, a fim de avaliar o efeito da divisão sobre os resultados. Usando a Incubação 1, as projeções para 1993 calculadas foram 10073 (*Divisão 1*) e 10838 (*Divisão 2*). Com a Incubação 2, as projeções foram 11403 (*Divisão 1*) e 11211 (*Divisão 2*). Portanto, o efeito da divisão foi pequeno (aumento de 7,6% quando foi usada a Incubação 1 e queda de 1,7% quando foi usada a Incubação 2). Fixando agora a divisão e variando a distribuição do tempo de incubação, nota-se que os resultados também variaram pouco (foram 13% maiores na *Divisão 1* e apenas 3% maiores na Incubação 2). Assim, todas as projeções são parecidas e menores que o valor observado (corrigido para o atraso de notificação), 16734. Isso pode ser um indício de que o valor da mediana do tempo de incubação é maior que os supostos, 2,2 e 4,3 anos. Para dar continuidade a essa análise, esse procedimento poderia ser repetido truncando a série em outros anos (por exemplo, 1989 e 1990). Note, porém, que a análise do ajuste da distribuição do tempo de incubação está baseada na estimativa obtida da densidade de contaminação. Portanto, seria importante fazer essa análise considerando diversas estimativas dessa curva. Essa análise também foi deixada para futura pesquisa.

Ainda analisando os resultados obtidos com o truncamento da série, é importante comparar os números estimados de contaminados com os números de casos observados. Baseado na Incubação 1, o número estimado de pessoas contaminadas até dezembro de 1991 ficou entre 45.000 e 52.000. Usando a Incubação 2, este número ficou entre 100.000 e 130.000 pessoas. Se for levado em consideração que até dezembro de 1992 foram diagnosticados 35.897 casos, um total em torno de 50.000 pessoas contaminadas parece ser pequeno, o que significa que os resultados baseados na Incubação 2 parecem ser mais próximos da série observada.

Modelos alternativos à distribuição Weibull têm sido sugeridos na literatura. Marion e Schechter (1993) comentam que a função de falha ("hazard function") da distribuição empírica do tempo de incubação tem crescimento pelo menos linear (como a de uma distribuição Weibull) até um certo instante t e, a partir daí se mantém constante. Baseado nisso, introduzem o modelo "Dampened Hazard" (DH) para estimar a distribuição do tempo de incubação, que é equivalente a um modelo em 2 estágios onde o tempo de visita ao primeiro tem distribuição Weibull e o tempo de visita ao segundo tem distribuição Exponencial. O método "back-calculation" é aplicado usando dados canadenses e o ajuste baseado no modelo DH foi melhor que o que

usa a distribuição Weibull.

Modelos em dois estágios onde o tempo de visita a cada estágio tem distribuição Weibull já haviam sido usados antes. Em Brookmeyer e Liao (1990) e Mariotto et al. (1992) supõe-se que a distribuição do tempo de visita a cada estágio seja Weibull e seus parâmetros são estimados. Em ambos os trabalhos, as estimativas correspondentes ao segundo estágio sugerem que a distribuição exponencial poderia ter sido usada, em lugar da distribuição Weibull. Em Brookmeyer (1991) são usadas a distribuição Weibull para o primeiro estágio e a distribuição exponencial para o segundo, gerando resultados consistentes com os de Marion e Schechter (1993), que usou o modelo DH, e de Rosenberg et al. (1992), que usou o modelo TSI.

A incerteza envolvida na escolha da distribuição do tempo de incubação poderia ser bastante reduzida se fosse feito um estudo de coorte no Brasil com o objetivo de estimar essa distribuição na nossa população.

As estimativas do número de pessoas contaminadas pelo HIV produzidas com o método "back-calculation" são maiores que as fornecidas pela ONU, quando se toma um valor pelo menos igual a 4,3 anos para a mediana do tempo de incubação. Estas estimativas apontam para um número entre 280 mil e 420 mil contaminados no Brasil, na faixa de 15 a 49 anos, em janeiro de 1992. Note que o erro associado a essas estimativas, obtidas utilizando o programa "Epimodel" do Programa Global de AIDS da OMS, é bem maior que o erro associado às estimativas obtidas com o back-calculation (entre 30 e 40 mil). Mantendo-se a relação atual dos casos diagnosticados em São Paulo com o total do Brasil, o número de pessoas contaminadas em São Paulo estaria entre 170000 e 250000. As estimativas obtidas com o "back-calculation" que supõem uma mediana para o tempo de incubação igual a 4,3 anos apontam para valores maiores que esse (entre 460.000 e 840.000).

Das três parametrizações usadas da densidade de contaminação, duas (I_2 e I_3) forneceram resultados bastante semelhantes. No que se refere à distribuição do tempo de incubação, a Incubação 2 apontou para resultados mais próximos do que foi observado. Por isso, para concluir, seria razoável considerar os resultados baseados em I_2 e I_3 e na Incubação 2.

Bibliografia

- Abramowitz, M., Stegun, I. (1970). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Applied Mathematics Series), Washington, DC: National Bureau of Standards.
- Aitken, M., Anderson, D., Francis, B., Hinde, J. (1989). *Statistical Modelling in GLIM*. Oxford: Clarendon Press.
- Anderson, R. M., Medley, G. F., Blythe, S. P., Johnson, A. M. (1987). Is it possible to predict the minimum size of the AIDS epidemic in the United Kingdom?. *Lancet*, **ii**, 1073-1075.
- Arcà, M., Perucci, C. A., Spadea, T. (1992). The epidemics dynamics of HIV-1 in Italy: modelling the interaction between intravenous drug users and heterosexual population. *Statistics in Medicine*, **11**, 1657-1684.
- Bacchetti, P., Segal, M. R., Hessol, N. A., Jewell, N. P. (1993). Impact of uncertain AIDS incubation period on reconstruction of HIV epidemics. *Proceedings of the National Academy of Sciences of the USA*, **90**, 2194-2196.
- Bailey, N. J. J. (1975). *The mathematical theory of infectious diseases and its applications*. 2^a ed. London: Charles Griffin.
- Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
- Biggar, J. (1990). AIDS incubation in 1891 HIV seroconverters from different exposure groups. *AIDS*, **4**, 1059-1066.
- Bishop, Y. M., Fienberg, S. E., Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. Cambridge, MA: M. I. T. Press.
- Brookmeyer, R. (1991). Reconstruction and future trends of the AIDS epidemic in the United States. *Science*, **253**, 37-42.
- Brookmeyer, R., Damiano, A. (1989). Statistical methods for short-term projections of AIDS incidence. *Statistics in Medicine*, **8**, 23-34.

- Brookmeyer, R., Gail, M. H. (1986). Minimum size of the AIDS epidemic in the United States. *Lancet*, **ii**, 1320-1322.
- Brookmeyer, R., Gail, M. H. (1987). Biases in prevalent cohorts. *Biometrics*, **43**, 739-749.
- Brookmeyer, R., Gail, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association*, **83**, 301-308.
- Brookmeyer, R., Gail, M. H. (1994). *AIDS Epidemiology: A Quantitative Approach*, New York, NY: Oxford University Press.
- Brookmeyer, R., Gail, M. H., Polk, B. F. (1987). The prevalent cohort study and the acquired immunodeficiency syndrome. *American Journal of Epidemiology*, **126**, 14-24.
- Brookmeyer, R., Goedert, J. J. (1989). Censoring in an epidemic with an application to hemophilia-associated AIDS. *Biometrics*, **45**, 325-335.
- Brookmeyer, R., Liao, J. (1990). Statistical modelling of the AIDS epidemic for forecasting health care needs. *Biometrics*, **46**, 1151-1163.
- Centro de Vigilância Epidemiológica/SP (1994). *Boletim Epidemiológico AIDS*. Secretaria de Estado da Saúde/SP, julho-1994.
- Cooper, D. A., Gold, J., Maclean, P., Donovan, B., Finlayson, R., Barnes, T. G., Michelmores, H. M., Brooke, P., Penny, P. R. (1985). Acute AIDS retrovirus infection. *Lancet*, **i**, 537-540.
- Curran, J. W., Morgan, M. W., Hardy, A. M., Jaffe, H. W., Darow, W. W., Dowdle, W. R. (1985). The epidemiology of AIDS: current status and future prospects. *Science*, **229**, 1352-1357.
- Darby, S. C., Doll, R., Thakrar, B., et al. (1990). Time from infection with HIV onset of AIDS in patients with *Hæmophilus* in the U. K. *Statistics in Medicine*, **9**, 681-689.
- Downs, A. M., Ancelle, R. A., Jager, Han J. C., Brant, J. (1987). AIDS Europe: current trends and short-term predictions estimated from surveillance data, January 1981-June 1986. *AIDS*, **1**, 53-57.

- Gail, M., Brookmeyer, R. (1988). Methods for projecting the course of the AIDS epidemic. *Journal of the National Cancer Institute*, **80**, 900-911.
- Gail, M. H., Rosenberg, P. S. (1992). Perspectives on using backcalculation to estimate HIV prevalence and project AIDS incidence. In: Dietz, K., Farewell, V. e Jewell, N. (eds.), *AIDS Epidemiology: Methodological Issues*. Boston: Birkhauser-Boston.
- Gamerman, D., Migon, H. S. (1991). Forecasting the number of AIDS cases in Brazil. *The Statistician*, **40**, 427-442.
- Goedert, J. J., Biggar, R. J., Melbye, M. et al. (1987). Effect of T4 count and co-factors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus. *Journal of the American Medical Society*, **257**, 331-334.
- Harris, J. E. (1987). Delay in reporting acquired immunodeficiency syndrome. Working Paper Number 2278, National Bureau of Economic Research.
- Heisterkamp, S. H., De Haan, B. J., Jager, J. C., Van Druten, J. A. M., Hendriks, J. C. M. (1992). Short and medium term projections of the AIDS/HIV epidemic by a dinamic model with an application to the risk group of homo/bisexual men in Amsterdam. *Statistics in Medicine*, **11**, 1425-1441.
- Isham, V. (1988). Mathematical modelling of the transmission dynamics of HIV infection and AIDS: a review. *Journal of the Royal Statistical Society, Series A*, **151**, 5-30.
- Jason, J., Lui, K. j., Ragni, M. et al. (1989). Risk of developing AIDS in HIV-infected cohorts of hæmophilic and homosexual men. *Journal of the American Medical Association*, **261**, 725-727.
- Kalbfleisch, J. D., Lawless, J. F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association*, **84**, 360-372.
- Kuo, J.-M., Taylor, J. M. G., Detels, R. (1991). Estimating the AIDS incubation period from a prevalent cohort. *American Journal of Epidemiology*, **133**, 1050-1057.
- Levy, J. A. (1993). Pathogenesis of human immunodeficiency virus infection. *Microbiological Reviews*, **57**, 183-289.

- Longini, I. M., Byers, R. H., Hessel, N. A., Tan, W. Y. (1992). Estimating the stage-specific numbers of HIV infections using a Markov model and back-calculation. *Statistics in Medicine*, **11**, 831-843.
- Longini, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F., Hethcote, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, **8**, 831-843.
- Longini, I. M., Clark, W. S., Haber, M. et al. (1990). The stages of HIV infection: waiting times and infection transmission probabilities. In: *Mathematical and Statistical Approaches to AIDS Epidemiology*. Ed. C. Castillo-Chaves. Lecture Notes in Biomathematics, **83**, Springer-Verlag, Berlin.
- Lui, K.-J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. W., Bregman, D. J. (1986). A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the USA*, **83**, 3051-3055.
- Mann, J., Tarantola, D. J. M., Netter, T. W., Sanecki, K., Way, P. O. (1993). In: Mann, J., Tarantola, D. J. M., Netter, T. W. (eds.), *A AIDS no mundo*. Rio de Janeiro: Relume Dumará.
- Marion, S. A. e Schechter, M. T. (1993). Use of backcalculation for estimation of the probability of progression from HIV infection to AIDS. *Statistics in Medicine*, **12**, 617-631.
- Mariotto, A. B., Mariotti, S., Pezzotti, P., Rezza, G., Verdecchia, A. (1992). Estimation of the AIDS incubation period in intravenous drug users: a comparison with male homosexuals. *American Journal of Epidemiology*, **135**, 428-437.
- Massad, E. (1987). *Sobre a biomatemática das moléstias infecciosas: um estudo teórico exemplificado por dois modelos originais*. São Paulo, Faculdade de Medicina - USP. (Tese - Livre Docência).
- McEvoy, M., Tillett, H. E. (1985). Some problems in the prediction of future numbers of cases of the acquired immunodeficiency syndrome in the U. K. *Lancet*, **ii**, 541-542.
- Medley, G. F., Anderson, R. M., Cox, D. R., Billard, L. (1987). Incubation period of AIDS in patients infected via blood transfusions. *Nature*, **328**, 719-721.

- Mode, C. J., Gollwitzer, H. E., Herrmann, N. (1988). A methodological study of a stochastic model of an AIDS epidemic. *Mathematical Biosciences*, **92**, 201-229.
- Murphy, E. L., Figueroa, J. P., Gibbs, W. N., Holding-Cobham, M., Cranston, B., Malley, K., Bodner, A. J., Alexander, S. S., Blattner, W. A. (1991). HTLV-I seroprevalence in Jamaica. I. Demographic determinants. *American Journal of Epidemiology*, **133**, 1114-1124.
- OMS (1991). In point of fact. The global HIV/AIDS situation: WHO projects 40 million HIV infections by the year 2000. **74**.
- Payne, C. D. (1985). *The generalized linear interactive modeling system*, Release 3.77, Manual Numerical Algorithms Group, Oxford.
- Programa Nacional de DST/AIDS (1994). *Boletim Epidemiológico - AIDS*. Ministério da Saúde, Ano VII (8), agosto-94.
- Rosenberg, P. S., Biggar, R. J., Goedert, J. J., Gail, M. H. (1991). Backcalculation of the number with human immunodeficiency virus infection in the United States. *American Journal of Epidemiology*, **133**, 276-285.
- Rosenberg, P. S., Gail, M. H. (1990). Uncertainty in estimates of HIV prevalence derived by backcalculation. *Annals of Epidemiology*, **1**, 105-115.
- Rosenberg, P. S., Gail, M. H., Carrol, R. J. (1992). Estimating HIV prevalence and projecting AIDS incidence in the United States: a model that accounts for therapy and changes in the surveillance definition of AIDS. *Statistics in Medicine*, **11**, 1633-1655.
- Rosenberg, P. S., Gail, M. H., Pee, D. (1991). Mean square error of estimates of HIV prevalence and short-term AIDS projections derived by backcalculation. *Statistics in Medicine*, **10**, 1167-1180.
- Rude, N. (1993). *Reconstruction et tendances futures de l'épidémie de l'infection a VIH: approche par la methode de retrocalcul*. Paris, Université Denis Diderot - Paris 7. (Tese - Doutorado)
- Sanathanan, L. (1972). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, **43**, 142-152.
- Simmons, P., Lainson, F. A. L., Cuthbert, R., Steel, C. M., Peutherer, J. F., Ludlam, C. A. (1988). HIV antigen and antibody detection: variable responses to infection in the Edinburgh hæmophilic cohort. *British Medical Journal*, **296**, 593-598.

Zeger, S. L., See, L.-C., Diggle, P. J. (1989). Statistical methods for monitoring the AIDS epidemic. *Statistics in Medicine*, 8, 3-21.

Apêndice

Dados e Resultados

Tabela A.1: Casos de AIDS no Estado de São Paulo

Notificação no intervalo	Ano do Diagnóstico										
	82	83	84	85	86	87	88	89	90	91	92
Até abr/1988	5	17	73	238	444	892	258				
Mai-dez/1988	1	0	4	10	28	204	1296				
1989	0	4	2	14	17	107	732	1872			
1990	0	1	0	7	9	59	103	836	2470		
1991	0	3	0	23	10	31	56	143	1410	3624	
1992	0	1	2	11	35	37	79	113	285	1775	5586
1993	0	0	0	2	0	5	24	32	47	109	1334
Total parcial*	6	26	81	303	543	1330	2524	2964	4165	5399	5586
Total geral**	6	26	81	305	543	1335	2548	2996	4212	5508	6920

* - casos notificados até dezembro de 1992 (os anos 88-92 foram usados para estimar a distribuição do atraso de notificação e, posteriormente, sofreram correção do atraso).

** - total de casos notificados até maio de 1993.

Fonte: Centro de Referência e Tratamento de AIDS do Estado de São Paulo (CRT-SP).

Tabela A.2: Estimativas do efeito do atraso, obtidas ajustando o modelo $\log(x_{ij}) = \alpha_i + \beta_j + \epsilon_{ij}$, onde x_{ij} é o número de casos diagnosticados no ano j e notificados no ano $i + j$, $\hat{\sigma}_i$ é o desvio-padrão associado à estimativa $\hat{\alpha}_i$ e ϵ_{ij} tem distribuição Poisson, $i = 0, 1, \dots, I$, $j = 1982, 1983, \dots, 1992$.

i	Correção 1		Correção 2		Correção 3	
	$\hat{\alpha}_i$	$\hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\sigma}_i$
1	7,77	0,014	6,03	0,051	6,74	0,021
2	6,76	0,019	5,02	0,049	5,75	0,024
3	5,08	0,042	3,35	0,061	3,97	0,047
4	4,48	0,067	2,77	0,079	3,33	0,078
5	4,28	0,091	2,53	0,103	3,21	0,113
6	3,97	0,141	2,22	0,165	-4,69	6,338
7	4,61	0,150	-4,69	6,338		
8	3,76	0,281				

Observações:

- Correção 1: são considerados todos os casos notificados entre maio de 1988 e dezembro de 1992.
- Correção 2: são considerados caso notificados entre maio de 1988 e dezembro de 1992, cujo diagnóstico foi dado a partir de 1987.
- Correção 3: são considerados casos notificados entre maio de 1988 e dezembro de 1992, cujo diagnóstico foi dado a partir de 1988.

Tabelas A.3, A.4 e A.5: Distribuições de atraso calculadas a partir da própria estimativa $\hat{\alpha}_i$ e desta somada e subtraída do erro-padrão correspondente ($\hat{\alpha}_i - \hat{\sigma}_i$, $\hat{\alpha}_i + \hat{\sigma}_i$), $i = 0$ (notificação no mesmo ano do diagnóstico), 1 (notificação no ano seguinte ao do diagnóstico), ..., I .

Tabela A.3: Correção 1

i	Distribuição calculada usando		
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
0	0,639	0,631	0,621
1	0,232	0,231	0,229
2	0,042	0,043	0,044
3	0,023	0,024	0,024
4	0,018	0,019	0,021
5	0,013	0,014	0,016
6	0,024	0,027	0,030
7	0,009	0,011	0,015

Tabela A.4: Correção 2

i	Distribuição calculada usando		
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
0	0,657	0,655	0,648
1	0,240	0,239	0,236
2	0,045	0,045	0,045
3	0,025	0,025	0,026
4	0,020	0,021	0,021
5	0,013	0,015	0,016
6	0,000	0,000	0,008

Tabela A.5: Correção 3

i	Distribuição	calculada	usando
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
0	0,671	0,668	0,662
1	0,249	0,249	0,247
2	0,041	0,042	0,043
3	0,021	0,022	0,023
4	0,018	0,019	0,021
5	0,000	0,000	0,004

Tabelas A.6, A.7 e A.8: Correções obtidas a partir das distribuições estimadas

Tabela A.6: Correção 1

Ano	Dados	corrigidos	usando
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
1985	65	65	65
1986	101	101	102
1987	456	459	462
1988	2669	2691	2718
1989	3203	3235	3271
1990	4666	4717	4776
1991	7148	7236	7337
1992	17483	17713	17976

Tabela A.7: Correção 2

Ano	Dados corrigidos		usando
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
1987	441	441	445
1988	2583	2588	2614
1989	3104	3113	3147
1990	4526	4542	4594
1991	6942	6967	7048
1992	16992	17049	17242

Tabela A.8: Correção 3

Ano	Dados corrigidos		usando
	$\hat{\alpha}_i - \hat{\sigma}_i$	$\hat{\alpha}_i$	$\hat{\alpha}_i + \hat{\sigma}_i$
1988	2547	2549	2562
1989	3051	3058	3078
1990	4428	4443	4477
1991	6787	6817	6875
1992	16654	16734	16885

Tabelas A.9, A.10 e A.11: Resultados do “back-calculation” obtidos com a distribuição de incubação sugerida por Brookmeyer e Gail (1988), Weibull (0,0243; 2,286). O desvio-padrão correspondente à estimativa é representado por “d. p.”.

Tabela A.9: Densidade de contaminação 1

Intervalos	Dens. estimada		$\hat{N}$	Casos acumulados	
	θ_i	d. p.	(d.p.)	até o ano 2.000	
77.0-82.0	0,0000	0,0000			
82.0-84.0	0,0004	0,0001	842.831	S/ novas	C/ novas
84.0-86.0	0,0029	0,0004	(46.538)	contaminações	contaminações
86.0-88.0	0,0093	0,0010		775.371	2.699.157
88.0-90.0	0,0031	0,0022			
90.0-92.0	0,0535	0,0075			
92.0-93.0	0,8615	0,0608			

Tabela A.10: Densidade de contaminação 2

Intervalos	Dens. estimada		$\hat{N}$	Casos acumulados	
	θ_i	d. p.	(d.p.)	até o ano 2.000	
77.0-82.0	0,0000	0,0000			
82.0-84.0	0,0008	0,0002	461.036	S/ novas	C/ novas
84.0-86.0	0,0054	0,0007	(16.442)	contaminações	contaminações
86.0-88.0	0,0170	0,0018		431.517	1.086.934
88.0-90.0	0,0059	0,0041			
90.0-91.5	0,0913	0,0111			
91.5-93.0	0,5365	0,0433			

Tabela A.11: Densidade de contaminação 3

Intervalos	Dens. estimada		$\hat{N}$ (d.p.)	Casos acumulados até o ano 2.000	
	θ_i	d. p.			
77.0-81.0	0,0000	0,0000			
81.0-83.0	0,0001	0,0001	474.834	S/ novas	C/ novas
83.0-85.0	0,0023	0,0004	(17.579)	contaminações	contaminações
85.0-87.0	0,0112	0,0014		444.034	1.132.764
87.0-88.5	0,0133	0,0034			
88.5-90.0	0,0100	0,0067			
90.0-91.5	0,0777	0,0137			
91.5-93.0	0,5474	0,0450			

Tabelas A.12, A.13 e A.14: Resultados do “back-calculation” obtidos com a distribuição de incubação sugerida por Medley et al. (1987), Weibull (0,1077; 2,396).

Tabela A.12: Densidade de contaminação 1

Intervalos	Dens. estimada		$\hat{N}$ (d.p.)	Casos acumulados até o ano 2.000	
	θ_i	d. p.			
77.0-82.0	0,0000	0,0000			
82.0-84.0	0,0004	0,0001	262.806	S/ novas	C/ novas
84.0-86.0	0,0030	0,0003	(8.497)	contaminações	contaminações
86.0-88.0	0,0127	0,0006		262.805	1.292.843
88.0-90.0	0,0171	0,0011			
90.0-92.0	0,0545	0,0038			
92.0-93.0	0,8246	0,0351			

Tabela A.13: Densidade de contaminação 2

Intervalos	Dens. estimada		$\hat{N}$	Casos acumulados	
	θ_i	d. p.	(d.p.)	até o ano 2.000	
77.0-82.0	0,0000	0,0000			
82.0-84.0	0,0007	0,0002	146.963	S/ novas	C/ novas
84.0-86.0	0,0053	0,0005	(3.005)	contaminações	contaminações
86.0-88.0	0,0227	0,0011		146.963	492.758
88.0-90.0	0,0308	0,0020			
90.0-91.5	0,0922	0,0054			
91.5-93.0	0,4950	0,0241			

Tabela A.14: Densidade de contaminação 3

Intervalos	Dens. estimada		$\hat{N}$	Casos acumulados	
	θ_i	d. p.	(d. p.)	até o ano 2.000	
77.0-81.0	0,0000	0,0000			
81.0-83.0	0,0001	0,0001	147.728	S/ novas	C/ novas
83.0-85.0	0,0021	0,0003	(3.352)	contaminações	contaminações
85.0-87.0	0,0128	0,0008		147.727	496.508
87.0-88.5	0,0286	0,0017			
88.5-90.0	0,0306	0,0032			
90.0-91.5	0,0908	0,0070			
91.5-93.0	0,4967	0,0273			