

CUSTOS DAS MORTES POR CAUSAS EXTERNAS NO BRASIL

Alexandre Xavier de Carvalho YWATA¹
Daniel R. C. CERQUEIRA²
Rute I. RODRIGUES²
Waldir J. A. LOBÃO³

- **RESUMO:** A violência no Brasil é reconhecidamente um dos maiores problemas atualmente enfrentados pela sociedade. Entre as conseqüências daí originadas, a perda de vidas humanas representa custos substanciais. Cada vítima fatal da violência, do ponto de vista econômico, representa enorme perda de investimentos em capital humano e, portanto, de capacidade produtiva. Qual é o custo social dessas mortes violentas no Brasil? Neste artigo, apresenta-se uma metodologia para estimar a perda de capital humano devido a mortes violentas no país, a partir da base de dados de renda dos trabalhadores do IBGE e da base de dados de óbitos do Ministério da Saúde. Para combinar informações dessas duas bases de dados, inicialmente foram aplicados procedimentos de regressão não-paramétrica para estimar curvas médias de rendimento anual dos trabalhadores. Essas curvas são então utilizadas para estimar a perda de produção para cada indivíduo morto prematuramente, vítima da violência. Incluem-se nessa análise ajustes pela tábua de sobrevivência da população. Estimou-se que em 2001, o custo por perda de produção foi de R\$ 9,1 bilhões devido aos homicídios, de R\$ 5,4 bilhões devido aos acidentes de transporte e de R\$ 1,3 bilhão devido aos suicídios. O custo total resultante das mortes por causas externas foi de R\$ 20,1 bilhões. Estimou-se, ainda, o total de anos de vida perdidos: no agregado das causas externas, esse total foi de 4,96 milhões de anos, sendo 2,15 milhões devido aos homicídios e 1,24 milhão devido aos acidentes de transporte. Esses números devem-se principalmente ao alto percentual de vítimas jovens.
- **PALAVRAS-CHAVE:** Custos de produção; homicídios; acidentes de transporte; métodos não-paramétricos; superfícies de renda.

1 Introdução

As mortes por causas violentas compõem um quadro absolutamente dramático no Brasil e, conseqüentemente, representam uma questão de primeira grandeza para as políticas públicas. Apenas em 2001, mais de 120 mil pessoas foram vitimadas. Tal questão é particularmente grave em relação à população de jovens entre 15 e 29 anos, cujos óbitos por causas violentas representaram 50,9% do total das mortes nessa faixa

¹ Diretoria de Estudos Regionais, Urbanos e Fiscais ; Instituto de Pesquisa Econômica Aplicada – IPEA; SBS – Quadra 1 - Bloco J - Ed. BNDES ; 70076-900 - Brasília - DF – Brasil. E-mail: alexandre.ywata@ipea.gov.br

² Diretoria de Estudos Macroeconômicos; Instituto de Pesquisa Econômica Aplicada – IPEA; Métodos Quantitativos; SBS - Quadra 1 - Bloco J - Ed. BNDES ; 70076-900 - Brasília - DF – Brasil.

³ Escola Nacional de Ciências Estatísticas – Ence/IBGE; Brasília - DF – Brasil.

etária. Entender com maior profundidade esta questão é elemento crucial para fundamentar políticas efetivas.

Com relação aos homicídios, nos últimos 25 anos houve aumento médio anual de 5,6% no número de registros, o que posicionou o país entre os mais violentos do planeta, com taxa de 28 homicídios para cada 100 mil habitantes. Nesse período, ocorreram 794 mil assassinatos. Diante dessa marcha acelerada da violência letal no país desde 1980, não cabe afirmar que se trata de explosão súbita de criminalidade, mas sim de tragédia anunciada, cujos incidentes evoluem com regularidade estatística espantosa, em verdadeiro processo endêmico, tendo em vista a sua generalização espacial e temporal, bem como a presença de um conjunto de fatores estruturais e locais que alimenta esta dinâmica criminal. Os acidentes de transporte também possuem grande prevalência, constituindo-se como o segundo fator de mortes entre aquelas ocasionadas por causas externas, de acordo com os dados do Ministério da Saúde. Os suicídios representam outro fator de mortalidade importante que, todavia, parece se tratar de uma questão invisível socialmente no Brasil.

A Figura 1 descreve as trajetórias desses incidentes desde 1980. Até o final desta década os acidentes de trânsito é que mais matavam no Brasil. Contudo, com o crescimento paulatino dos homicídios, já na virada da década estes assumiram a primeira posição no número de mortes, dentro do grupo das causas externas. É interessante ainda observar a mudança de tendência de crescimento dos acidentes de trânsito, que se anunciava na primeira metade dos anos noventa, quando a partir de 1997 essas mortes passam a diminuir. Possivelmente, tal fato deve ter correlação com a introdução do Código de Trânsito Brasileiro, Lei no 9.503, de 23/09/1997, e das muitas campanhas para a diminuição dos acidentes nas estradas.

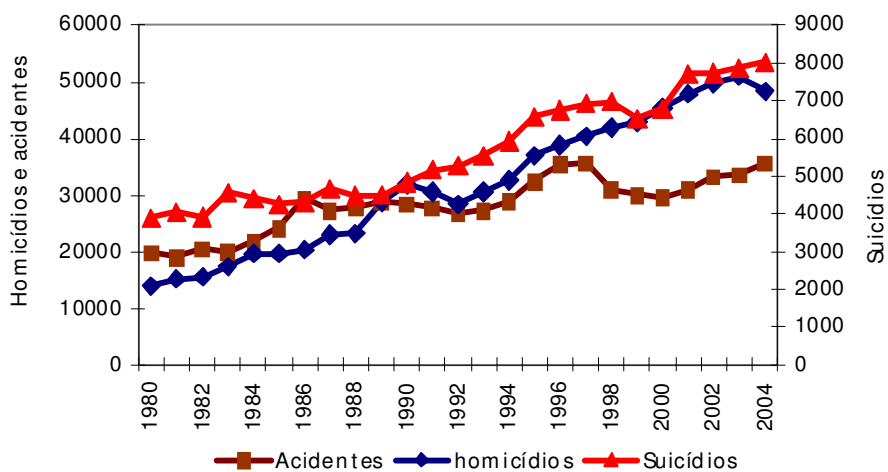


Figura 1 - Homicídios, acidentes de transporte e suicídios no Brasil.

Fonte: SIM/DATASUS. Grupo de Estudo de Violência, Ipea.

Este artigo apresenta estimativas para o custo social com a perda de capital humano decorrente das mortes por causas externas no Brasil. Para enriquecer a análise, além dos resultados para as causas externas em geral, apresentam-se também os resultados para homicídios, acidentes de trânsito e suicídios separadamente. Na mensuração de perdas de produção, calculou-se o valor presente total dos rendimentos que os indivíduos deixaram de receber em consequência dos óbitos ocasionados pela violência. A metodologia construída utilizou-se das bases de dados da Pesquisa Nacional por Amostras de Domicílios – Pnad, do IBGE, e do Sistema de Informações de Mortalidade SIM, do Ministério da Saúde.

Inicialmente, mapeou-se o perfil das pessoas que sofreram óbitos por causas violentas, a partir das informações socioeconômicas constantes no SIM (a idade do indivíduo, o gênero, o município de residência e o grau de escolaridade). Com base nos microdados da Pnad, obtiveram-se curvas médias de rendimentos para cada perfil identificado. Por fim, essas duas informações foram cruzadas, imputando-se, para cada indivíduo morto, a renda de trabalho média que esse indivíduo teria caso permanecesse vivo em todo o seu período produtivo (foram adotadas as idades de 15 a 65, inclusive). Adicionalmente, levou-se em consideração a tábua de sobrevivência do IBGE, tendo em vista que os indivíduos vítimas de mortes violentas não necessariamente viveriam até os 65 anos. Para estimar essas curvas médias de rendimentos, foram utilizadas subamostras separadas, de acordo com o sexo e com a área geográfica do país, a partir da junção dos microdados da Pnad, para diferentes anos. A divisão em várias subamostras foi possível graças ao grande número de observações existente nessa base do IBGE. A metodologia empregada permitiu que fossem construídas tabelas com as estimativas contendo o valor do rendimento médio versus a idade do indivíduo, o gênero e o estado de residência da vítima.

A partir da metodologia empregada, obteve-se uma estimativa para o custo total de perda de produção no Brasil em torno de R\$ 20,1 bilhões, para o ano de 2001. Os homicídios são responsáveis por R\$ 9,1 bilhões desse custo total. Acidentes de transporte vêm em segundo lugar, com estimativa de custo total em torno de R\$ 5,4 bilhões. Suicídios apresentam perda total de R\$ 1,3 bilhão. Em termos de valores médios de perda de produção, os homicídios apresentam as maiores médias: R\$ 189,5 mil para o ano de 2001, contra R\$ 172 mil, para os acidentes de transporte, e R\$ 163 mil por vítima, ao ano, para suicídios.

O método empregado neste artigo faz parte de um conjunto de metodologias para calcular valores de referência para a vida humana, que possui dois arcabouços teóricos distintos. Em uma primeira abordagem, do “valor da vida”, procura-se calcular o valor que os indivíduos creditam à violência ou, em particular, às mortes violentas. Em uma segunda abordagem, busca-se estimar o valor das perdas efetivamente ocorridas por conta daquelas mortes.

No primeiro método, o valor da vida é obtido por meio de estimações da disposição a pagar (willingness to pay) dos indivíduos para evitar a violência, ou a probabilidade de ser vitimado. Ainda nesta primeira abordagem, várias metodologias são empregadas, entre as quais a dos preços hedônicos e valoração contingente^[1]. Uma análise aprofundada de trabalhos estimando o valor estatístico da vida, e avaliando o trade-off entre valores monetários e riscos de fatalidades, é encontrada em Viscusi e Aldy (2003). Nesse

[1] Vide Cerqueira *et al.* (2007) e Soby e Ball (1995) para maiores detalhes.

trabalho, os autores apresentam extensa revisão com mais de 60 artigos sobre valores monetários do risco de mortalidade (mortality risk premium) e aproximadamente 40 estudos sobre valores monetários do risco de danos físicos (injury risk premium). Outros trabalhos de interesse que utilizam este enfoque são devidos a Usher (1973), Rosen (1988), Christie e Soguel (1995), Cohen et al. (2004), Atkinson et al. (2005), Nagin e Piquero (2006), Zarkin, Cates e Bala (2000), Cook e Ludwig (2000) e Ludwig e Cook (2001).

A segunda abordagem não trata do valor que as pessoas creditam à vida, ou como os indivíduos valoram a probabilidade da vitimização, mas calcula o valor efetivo que as mortes por causas violentas gera para a sociedade em termos de perda de rendimento destes indivíduos. Este valor é calculado com base na estimação de curvas de rendimentos dos indivíduos, em que o diferencial de rendimentos entre eles é baseado no nível de escolaridade, gênero e experiência profissional do indivíduo. Nessa abordagem, desenvolvida primeiro por Mincer (1974), as diferenças salariais são interpretadas como diferenciais de produtividade baseados em modelos de investimento em capital humano.

Segundo Soby e Ball (1995), uma das críticas à utilização da segunda abordagem – empregada neste texto – é que ela subestima a real disposição a pagar da sociedade para evitar mortes prematuras. Contudo, a grande limitação dos modelos que empregam a disposição a pagar segue justamente pelo motivo inverso, ou pelo fato de essas metodologias superestimarem o valor da vida, tendo em vista: i) a alta (e óbvia) correlação entre renda e disposição a pagar; e ii) o fato de que, normalmente, a maior vitimização ocorre entre as classes de menor rendimento. Para o Brasil o único trabalho encontrado em que se estimou a disposição a pagar para evitar as mortes por homicídios foi de Soares (2004).

Os custos de produção estimados neste trabalho correspondem apenas a uma das parcelas da perda total decorrente do falecimento da vítima. Obviamente, há muitos outros custos, cuja possibilidade de mensuração em termos monetários é até mesmo questionável. Abstraindo-se de uma análise totalmente monetária em termos de perda de produção, foi calculado, alternativamente, o total de anos perdidos devido às mortes por homicídios ou acidentes de transporte. A idéia é calcular a esperança matemática dos anos vividos por cada uma das vítimas, caso ela não tivesse falecido devido ao evento registrado nos dados de óbitos. No agregado, as mortes por causas externas ocasionaram, em 2001, um total de 4,96 milhões de anos perdidos.

Além desta seção introdutória, o artigo está dividido em mais três seções. A Seção 2 descreve os procedimentos seguidos em cada passo da metodologia empregada. Uma discussão extensa sobre a abordagem utilizada para a estimação das curvas de renda do trabalho é apresentada. Na Seção 3, apresentam-se os principais resultados para as estimativas de perda de produção e discutem-se os motivos de alguns dos resultados encontrados. Conclusões e comentários finais são apresentados na Seção 4.

2 Metodologia

Nesta seção, descreve-se a metodologia empregada para estimar as perdas de produção no Brasil nos anos recentes. O trabalho combina diferentes bases de dados governamentais, contendo informações socioeconômicas (idade, gênero, escolaridade e salários anuais) e informações de óbitos, decorrentes de diferentes causas externas. As

bases utilizadas nas estimativas de perda de produção decorrente das mortes por causas externas foram a Pesquisa Nacional por Amostra de Domicílios – Pnad, do IBGE, a tábua de sobrevivência do IBGE e a base de registros de óbitos do SIM, do Ministério da Saúde. Conforme será discutido mais adiante, para aumentar a precisão das diversas estimativas econométricas, foram combinadas bases da Pnad para os anos de 2001, 2002 e 2003. As bases do SIM utilizadas correspondem aos anos de 2000 e 2001. Na base do SIM foram utilizados os dados constantes na 10ª versão da Classificação Internacional de Doenças (CID-10), capítulo XX (Causas Externas de Morbidade e de Mortalidade), que correspondem aos códigos V01-Y98. No capítulo XX, além dos incidentes relacionados aos homicídios, suicídios e acidentes de transporte, existe uma série de outras causas específicas de mortalidade, como afogamento, quedas, envenenamentos não intencionais etc. Neste trabalho, optou-se por agrupar as informações pelas categorias homicídios, suicídios, acidentes de transporte e pelo total das causas externas.

Em termos gerais, a estimação das perdas de produção no Brasil, devido a homicídios e acidentes de transporte, segue os seguintes passos:

- 1) Estimação econométrica de superfícies que relacionam a renda média anual do trabalho, exclusivamente para as pessoas entre 15 e 65 anos, que estavam trabalhando na semana de referência da Pnad, às características pessoais, em termos de idade, anos de escolaridade, localização geográfica onde o indivíduo reside e gênero. Essas estimações utilizam-se das informações da Pnad para os anos de 2001 a 2003. Com isso, obtém-se para cada combinação entre sexo, idade, escolaridade e localização geográfica a renda média do trabalho.
- 2) A partir das equações econométricas que relacionam renda média do trabalho a variáveis socioeconômicas, é possível ter idéia do fluxo de renda médio que o trabalhador geraria, caso não tivesse sido vítima de mortes por causas externas. Na base de dados do SIM, para cada vítima de causas externas registrada, estão disponibilizadas informações sobre as quatro variáveis socioeconômicas utilizadas como variáveis explicativas nas estimações econométricas (idade, escolaridade, localização geográfica e sexo)^[1]. Portanto, a partir das curvas de renda do trabalho, item 1 acima, é possível estimar o fluxo médio de renda para cada vítima no SIM e trazer esse fluxo a valor presente, obtendo estimativa da perda de produtividade para esse indivíduo específico. Os fluxos futuros foram ajustados para as probabilidades de falecimento futuro, a partir da tábua de sobrevivência descrita em Caetano (2006).
- 3) Finalmente, a partir dos valores presentes calculados para as vítimas registradas no SIM, podem-se somar esses valores por subgrupo populacional, ou subgrupo por tipo de causas externas (homicídios, acidentes de transporte ou suicídios), obtendo-se estimativas para as perdas agregadas de produção. Por exemplo, é possível se obter a perda de produção, devido a acidentes de transporte, nas áreas urbanas na região Sudeste do Brasil.

[1] A escolha das variáveis idade, escolaridade, localização geográfica e sexo, como variáveis explicativas nos modelos econométricos estimados a partir da Pnad, deveu-se ao fato de essas quatro variáveis explicativas também possuírem informações disponibilizadas no SIM. Caso fossem utilizadas outras covariáveis na equação estimada com dados da Pnad, não haveria como utilizar tal equação para predição da renda salarial média para as vítimas registradas no SIM.

Nas seções a seguir, apresentam-se em mais detalhes os diversos passos empregados na estimação das perdas de produção em decorrência de causas externas. Inicialmente, discutem-se conceitualmente os motivos pelos quais optou-se por uma metodologia econométrica mais flexível, ao invés de abordagens paramétricas comumente encontradas na literatura. Em seguida, são detalhadas as técnicas não-paramétricas empregadas para estimação das superfícies de renda do trabalho *versus* idade e escolaridade. Finalmente, discute-se a estimação das perdas de produção, a partir do valor presente dos fluxos de renda futura.

2.1 Divisão das amostras em subgrupos para as estimações econométricas

Para capturar as marcantes diferenças regionais no Brasil, os dados na base de dados da Pnad foram divididos em 20 domínios geográficos. A idéia foi estimar curvas de salários médios diferentes em cada subgrupo. A composição dos 20 subgrupos selecionados, bem como a população e a renda mensal média do trabalho, estão apresentadas na Tabela 1. Essa composição foi escolhida de acordo com proximidade geográfica e de forma a se ter número razoável de indivíduos na amostra da Pnad em cada subgrupo, para estimar apropriadamente diferentes funções de salário *versus* idade e escolaridade.

Tabela 1 - Composição dos subgrupos geográficos escolhidos

Grupos geográficos	Descrição da área geográfica	População (mil hab.)	Renda média (R\$) ^[1]
Áreas Urbanas	Região Norte, exceto Pará e Tocantins	4,57	475,64
	Pará	4,78	347,21
	Tocantins, Maranhão e Piauí	6,78	329,20
	Ceará	5,97	307,50
	Rio Grande do Norte, Paraíba, Sergipe e Alagoas	8,36	347,73
	Pernambuco	6,19	353,75
	Bahia	9,02	368,15
	Minas Gerais	15,92	461,42
	Rio de Janeiro e Espírito Santo	17,24	634,61
	São Paulo	37,12	730,37
	Paraná e Santa Catarina	13,10	600,11
	Rio Grande do Sul	8,56	603,89
	Mato Grosso e Mato Grosso do Sul	3,94	549,67
	Goiás	4,75	501,33
	Distrito Federal	2,24	997,98
	Áreas Rurais	Norte	3,39
Nordeste		11,50	105,83
Sudeste		6,21	281,40
Sul		4,70	291,60
Centro-Oeste		1,63	268,25

Fonte: Pnad/2003. Elaboração dos autores.

^[1] Os dados de população e renda média foram obtidos da Pnad/2003. Os valores de renda correspondem às rendas mensais médias de todos os trabalhos, para os indivíduos empregados na semana de referência. Toda a manipulação de bases de dados foi efetuada utilizando-se o software estatístico SAS.

Por meio de estimações de modelos diferentes para cada um dos 20 domínios geográficos apresentados na Tabela 1, é possível contabilizar as diferenças regionais de renda, inclusive em termos de como a variável salário se relaciona às variáveis idade e escolaridade. Por outro lado, outra variável observável na base de dados do SIM, e que também possui significância para explicar diferenças salariais, é a variável sexo do indivíduo (de acordo com estimativas paramétricas preliminares). Tradicionalmente, nas estimações envolvendo a variável gênero, utiliza-se uma variável categórica. Entretanto, dada a disponibilidade do grande número de observações na base de dados da Pnad, a utilização de variáveis categóricas simplesmente pode levar à perda de informação, devido à natureza restritiva dessa abordagem (modelos puramente aditivos, não levando em conta a presença de interações entre gênero e as demais variáveis). A alternativa utilizada neste trabalho foi dividir cada um dos 20 subgrupos regionais em duas amostras, cada qual correspondendo a indivíduos de um determinado sexo. Assim, o número total de subgrupos, depois de cruzar domínios geográficos com gênero, resultou em 40.

Para cada um desses 40 grupos, foi estimada uma função específica para a relação entre salário de todos os trabalhos *versus* idade e escolaridade. Conforme será discutido mais adiante, foram utilizadas expansões de funções base, de forma a estimar a função renda *versus* idade e escolaridade não-parametricamente. Além disso, informações das bases de dados da Pnad para os anos de 2001, 2002 e 2003 foram combinadas. A hipótese implícita nesse caso é de que a relação entre salário e renda e idade não variou muito entre os anos de 2001 a 2003, o que parece ser plausível. Além disso, o objetivo do trabalho foi de encontrar uma visão estrutural, e não conjuntural, para os custos da violência no Brasil. Os valores de salário nominal foram ajustados para o mesmo ano base, utilizando-se o IPNC geral^[1].

Uma questão importante nas estimações efetuadas neste trabalho é o que levou à escolha de modelos mais flexíveis ao invés de formas paramétricas mais tradicionais, conforme comumente encontrado na literatura. De fato, uma maneira simples de estimar as curvas de renda do trabalho *versus* idade, escolaridade e sexo, seria utilizar a especificação

$$Renda_i = \beta_0 + \beta_1 \times idade_i + \beta_2 \times Sexo_i + \beta_3 \times Escolaridade_i + \varepsilon_i.$$

O subscrito *i* indica o indivíduo na amostra, que segue a tradição de Mincer (1974). A variável sexo é igual a 1 quando o trabalhador for do sexo masculino, e 0 caso contrário. Para contabilizar as diferenças regionais, pode-se adicionar um conjunto de variáveis categóricas para os domínios geográficos. Além disso, podem-se incluir uma série de termos polinomiais de ordem maior, do tipo idade_{*i*}², idade_{*i*}³, anos escolaridade_{*i*}³ etc., para contabilizar não-linearidades. A utilização de um modelo paramétrico conforme especificado na equação acima é particularmente interessante quando o objetivo é estimar (e testar a significância de) um parâmetro específico, como o coeficiente β_3 , que indica o retorno do salarial à educação.

O objetivo específico das estimações econométricas neste artigo não é testar ou estimar parâmetro algum na relação entre renda do trabalho e as variáveis explicativas. O objetivo é justamente estimar a renda média do trabalho, para cada idade, para cada nível

^[1] Para todos os exercícios de deflacionamento/inflacionamento de valores monetários neste trabalho, utilizou-se o INPC geral, obtido do endereço www.ipeadata.gov.br – a fonte primária para esse índice é o IBGE.

de escolaridade, por sexo e por área geográfica. A escolha dessas quatro variáveis explicativas está diretamente ligada às informações que estão de fato disponíveis na base de dados do SIM. Dado que o objetivo é estimar a renda mensal média para cada combinação de valores das variáveis preditoras, a utilização de formas paramétricas restritivas, como a apresentada na especificação na equação acima, possui a desvantagem de que, ao final das estimativas, para grande parcela de combinações entre as diversas variáveis explicativas, o valor estimado para a média da renda do trabalho resulta em valor negativo, impossibilitando a utilização dessa estimativa no cálculo da perda de produtividade, conforme será descrito mais adiante. De fato, em diversas estimações paramétricas efetuadas inicialmente, encontraram-se mais de 25% dos valores preditos com sinal negativo. Obviamente, uma primeira alternativa seria utilizar o logaritmo $\log(\text{Renda}_i)$, ao invés de Renda_i , no lado esquerdo da equação acima. Essa alternativa, porém, não foi utilizada, pois foram observados problemas no ajuste dos valores de renda próximos a zero, bem como dos valores de renda mais altos.

2.2 Estimação não-paramétrica das curvas de salário anual

Nesta seção, apresenta-se a abordagem de estimação não-paramétrica^[1] para estimar a curva de salário anual médio *versus* faixa de escolaridade e idade do trabalhador. Conforme discutido anteriormente, as estimações apresentadas neste trabalho foram efetuadas por subárea geográfica (vide tabela 1) e por sexo dos trabalhadores. A combinação entre sexo e subárea geográfica incorreu em diversos subgrupos populacionais, em que diferentes curvas foram estimadas para cada subgrupo. Para facilitar a apresentação da metodologia a seguir, suprime-se o subscrito k , correspondente ao subgrupo populacional k , de forma a simplificar a notação.

A idéia da estimação não-paramétrica consiste em estimar uma função do tipo:

$$r_i = g(e_i, x_i) + \varepsilon_i, \quad (1)$$

em que r_i é o salário anual do indivíduo, e_i é a escolaridade média do indivíduo e x_i é a idade. O índice i corresponde ao indivíduo específico na base de dados e o termo ε_i é uma variável aleatória, com média zero e variância desconhecida, contabilizando os demais fatores não incluídos na parte sistemática da regressão.

A função $g(e_i, x_i)$ possui forma funcional desconhecida e deverá ser estimada a partir dos dados. Para estimá-la, emprega-se uma expansão de funções base (*basis functions*). Essa expansão baseia-se na aproximação da curva desconhecida $g(e_i, x_i)$, utilizando a forma paramétrica flexível:

$$g(e_i, x_i) \cong \sum_{m=1}^M b_m \times u_m(e_i, x_i), \quad (2)$$

em que as funções $u_m(e_i, x_i)$ são as funções base, com forma funcional conhecida. A forma funcional em (2) engloba uma série de modelos comumente encontrados na

^[1] Modelos de regressão não-paramétrica são aqueles em que a forma funcional da função resposta não é conhecida e tem de ser estimada a partir dos dados. Para mais detalhes, ver Hastie, Tibshirani e Friedman (2001). As estimativas foram obtidas através de funções escritas pelos autores na linguagem de programação matricial Ox (vide <http://www.doornik.com>).

literatura de *statistical machine learning e data mining*. Entre esses modelos encontram-se, por exemplo, regressões de redes neurais e regressões de *wavelets* (vide Hastie, Tibshirani e Friedman, 2001).

Neste trabalho, utiliza-se expansão do tipo *B-splines* de ordens q_1 e q_2 . Para descrever a expansão de *B-splines*, empregadas na estimação da função $g(e_i, x_i)$, será considerado inicialmente o caso em que há apenas uma variável preditora. Para estimar a renda anual do trabalhador como função da idade x_i apenas, tem-se

$$r_i = h(x_i) + \varepsilon_i. \quad (3)$$

Nesse caso, a expansão de funções base pode ser escrita como

$$h(x_i) \cong \sum_{l=1}^L b_l \times u_l(x_i). \quad (4)$$

Devido ao fato de as funções base $u_l(x)$ nas expansões de *B-splines* dependerem da ordem q dos *splines*, escreve-se explicitamente $u_{l,q}(x)$. Assume-se que a variável explicativa (independente) x varia no intervalo $[x_{min}, x_{max}]$. Considera-se, então, um vetor de w pontos (x_1, x_2, \dots, x_w) dividindo o intervalo $[x_{min}, x_{max}]$, em que $x_{min} < x_1 < x_2 < \dots < x_w < x_{max}$. A idéia da expansão de *B-splines* é ajustar um polinômio de grau $(q-1)$ em cada intervalo definido por pontos consecutivos (pontos no conjunto $x_{min}, x_1, x_2, \dots, x_w, x_{max}$). Em geral, utiliza-se $q = 3$ ou 4 , de forma que os polinômios utilizados possuem grau 2 ou 3 respectivamente.

Considere-se agora o vetor de nós $x_{min}, \dots, x_{min}, x_1, x_2, \dots, x_w, x_{max}, \dots, x_{max}$, em que os valores x_{min} e x_{max} nos extremos são repetidos um número q de vezes. Para facilitar a discussão a seguir, escreve-se o vetor $(x_{min}, \dots, x_{min}, x_1, x_2, \dots, x_w, x_{max}, \dots, x_{max})$ na forma $(t_1, t_2, \dots, t_{w+2q})$. Portanto, $t_1 = x_{min}, \dots, t_q = x_{min}$. A partir do vetor de nós $(t_1, t_2, \dots, t_{w+2q})$ e da ordem q , as funções base $u_{l,q}(x)$ podem ser construídas recursivamente, como segue:

$$u_{l,1}(x) = \begin{cases} 1, & t_l \leq x < t_{l+1} \\ 0, & \text{caso contrário,} \end{cases} \quad (5)$$

$$u_{l,p}(x) = \frac{x - t_l}{t_{l+p-1} - t_l} u_{l,p-1}(x) + \frac{t_{l+p} - x}{t_{l+p} - t_{l+1}} u_{l+1,p-1}(x), \text{ para } p = 2, \dots, q. \quad (6)$$

Observe-se que no denominador do segundo termo na expressão (6) aparece o valor t_{l+p} . Portanto, para *B-splines* de ordem q , o número total de funções base será igual ao número de nós menos o valor q . Ou seja, o número L de funções base é dado por $L = w + q$ e as funções base serão $u_{1,q}(x), \dots, u_{L,q}(x)$. Pode-se, então, reescrever a expressão (4), especificamente para a expansão de *B-splines*, como

$$h(x_i) \cong \sum_{l=1}^L b_l \times u_{l,q}(x_i). \quad (7)$$

O grau de flexibilidade da forma funcional em (7) é regulado pelo número L de funções base. Esse número está diretamente relacionado ao número de pontos divisórios

w , bem como à ordem q . Quanto maior o número L , maior a flexibilidade da expansão semiparamétrica. Uma vez fixado o número de funções base L para estimar a função desconhecida $h(x_i)$, a estimação dos parâmetros b_l , $l = 0, 1, 2, \dots, L$ pode ser efetuada via estimadores paramétricos tradicionais. Podem-se utilizar estimadores de mínimos quadrados ordinários, ou algum estimador do tipo mínimos quadrados ponderados, com correção para heteroscedasticidade, ou para a presença de observações discrepantes (*outliers*). Para mais detalhes, vide, por exemplo, Silveira *et al.* (2007). Observe-se agora que o problema de ajuste da curva $h(x_i)$ consiste na estimação do modelo de regressão (linear nos parâmetros):

$$r_i = \sum_{l=1}^L b_l \times u_{l,q}(x_i) + \varepsilon_i. \quad (8)$$

A escolha do número L de funções base pode ser efetuada utilizando-se algum critério de seleção de modelos, como por exemplo, o critério de informação de Akaike – AIC ou o critério de informação Bayesiano – BIC^[1] (regras de escolha *in-sample*), ou utilizando-se alguma regra de validação cruzada (*cross-validation*), que correspondem a regras de escolha *out-of-sample*. Esses procedimentos têm por objetivo evitar um sobreajuste (*overfitting*) do modelo não-paramétrico. De fato, caso se aumente o número de funções base indefinidamente, obtém-se modelo com ajuste perfeito dentro da amostra, mas com um poder preditivo questionável para observações fora da massa de dados. Por outro lado, para um número L muito baixo, obtém-se baixa flexibilidade da expansão de funções base, o que também incorre na perda de poder preditivo do modelo. Os diversos critérios de seleção de modelos auxiliam na escolha do número L de forma a maximizar a relação de compromisso entre flexibilidade do modelo e número excessivo de parâmetros desconhecidos^[2]. Para mais detalhes em escolhas do grau de flexibilização em modelos semiparamétricos e não-paramétricos, vide Hastie, Tibshirani e Friedman (2001) e Alpaydin (2004).

O procedimento univariado discutido acima pode ser estendido para incluir a variável escolaridade e_i , de forma que a curva estimada passe a ser a função bivariada $g(e_i, x_i)$. Nesse caso, é preciso encontrar nova expansão de funções base para o caso de duas variáveis predictoras. Um modelo geral, que contabiliza para possíveis interações entre idade e escolaridade, é a expansão em que a matriz modelo (*design matrix*) possui as funções base de cada variável individualmente, além dos termos cruzados (resultantes da multiplicação) entre as funções base das duas variáveis (vide Hastie, Tibshirani e Friedman, 2001, para mais detalhes sobre *B-splines* multidimensionais, a partir do produto tensorial de *B-splines* unidimensionais). Portanto, a forma mais geral para a expansão de funções base no caso bivariado tem expressão:

$$g(e_i, x_i) \cong \sum_{m=1}^{L_1} d_m \times v_{m,q_1}(e_i) + \sum_{l=1}^{L_2} b_l \times u_{l,q_2}(x_i) + \sum_{m=1}^{L_1} \sum_{l=1}^{L_2} c_{m,l} \times [v_{m,q_1}(e_i) \times u_{l,q_2}(x_i)]. \quad (9)$$

Para cada subgrupo populacional, estimaram-se modelos utilizando-se diferentes expansões de funções base, em que os coeficientes desconhecidos d_m , b_l , e $c_{m,l}$ foram

^[1] Para mais detalhes sobre o AIC e o BIC, ver Burnhan e Anderson (1998).

^[2] Na literatura em estatística, esse problema é comumente conhecido como *trade-off* viés-variância (*bias-variance*).

estimados via mínimos quadrados ordinários. Para cada subgrupo, as ordens q_1 e q_2 dos *splines*, bem como os números de nós w_1 e w_2 foram escolhidos utilizando os critérios de informação AIC e BIC. A partir das estimativas para os coeficientes desconhecidos, foram estimados os salários médios para indivíduos em diferentes faixas de escolaridade, com idades variando entre 15 a 65 anos. A escolha desse intervalo de idade aproxima-se de padrões internacionais (alguns autores utilizam 64 anos), correspondendo à idade produtiva dos indivíduos. Esses valores de salários médios foram, então, empregados para estimar a perda de produção em decorrência das mortes por causas externas.

2.3 Estimação dos custos da violência com a perda de capital humano

Uma vez estimadas as funções de renda média anual do trabalho *versus* idade, escolaridade, sexo e domínio geográfico, o próximo passo foi aplicar essas curvas para obter a perda de produção, devido às mortes por causas externas. Para cada registro de óbito no SIM, decorrente de causas externas, estimou-se o fluxo médio de renda do trabalho que o indivíduo teria caso continuasse vivo nos próximos anos da sua vida produtiva (dos 15 aos 65 anos). O custo de produtividade referente a esse indivíduo específico é dado pelo valor presente desse fluxo de renda do trabalho. O valor presente foi ajustado de forma a levar em conta as diferentes probabilidades de sobrevivência que o indivíduo teria, caso não tivesse morrido prematuramente vítima da violência. Por meio do somatório dos custos de produtividade de cada registro individualmente, foi possível estimar o custo de produtividade para diferentes tipos de óbitos, em diferentes unidades da federação, ou em diferentes cidades.

A expressão geral para o valor presente VP_k da perda de produção, para o indivíduo k , é dada por

$$VP_k = \sum_{x=D_k}^T \frac{1}{(1+t_d)^{(x-D_k)}} \times \Pr(F_k > x/F_k \geq D_k) \times \hat{g}_{\text{subgrupo de } k}(e_k, x), \quad (10)$$

em que t_d é a taxa de desconto anual, $\hat{g}_{\text{subgrupo de } k}(e_k, x)$ corresponde à estimativa, via *B-splines*, da função renda média anual do trabalho, específica para o subgrupo (um dos 40 subgrupos, de acordo com as figuras A1 a A10) ao qual pertence o indivíduo k , na base do SIM. O número T corresponde ao teto de idade escolhido – no caso, 65 anos. A probabilidade $\Pr(F_k > x/F_k > D_k - 1)$ corresponde à probabilidade de o indivíduo estar vivo com a idade x , dado que ele não faleceu aos D_k anos de idade (idade registrada no SIM). A variável F_k corresponde a uma variável aleatória indicando a idade de falecimento^[1] do indivíduo k .

^[1] No cômputo das probabilidades de sobrevivência condicionais, para ponderar as parcelas no cálculo do valor presente, considerou-se a possibilidade de se o indivíduo não tivesse morrido pela causa externa estudada, ele poderia ter morrido por outro motivo, no mesmo ano. Esse é o motivo de se usar a expressão $\Pr(F_k > x/F_k \geq D_k)$. Note que, dado que o falecimento ocorreu quando o indivíduo tinha D_k anos, a hipótese assumida como condicionante na probabilidade condicional é que o indivíduo atingiu uma idade acima dos $(D_k - 1)$ anos. Uma alternativa seria assumir que, caso o indivíduo não tivesse falecido devido àquele evento específico, ele não teria morrido por qualquer outra causa aos D_k anos de idade e a probabilidade condicional teria expressão $\Pr(F_k > x/F_k > D_k)$. Diversos dos resultados foram recalculados considerando-se essa última expressão para a probabilidade condicional e os números obtidos foram muito similares aos valores apresentados neste artigo.

Essa abordagem baseia-se implicitamente em uma série de hipóteses:

- 1) Para trazer o fluxo de renda do trabalho para valor presente, é necessário especificar taxa de desconto, escolhida arbitrariamente. Análise de sensibilidade foi conduzida para testar a sensibilidade dos resultados a diferentes taxas de desconto. A Subseção 3.2 apresenta discussão sobre o valor utilizado para a taxa de desconto.
- 2) As tábuas de sobrevivência utilizadas foram discutidas em Caetano (2006) e correspondem a toda a população, sem haver divisão por subgrupos. Na prática, não necessariamente essas probabilidades seriam iguais – dada a restrição de disponibilidade de informações, utilizaram-se as probabilidades indiscriminadamente, assumindo que os resultados não serão demasiadamente afetados.
- 3) Os fluxos esperados de renda do trabalho para os próximos anos do período produtivo de cada indivíduo registrado no SIM baseiam-se nas curvas de renda estimadas a partir de dados estáticos para o período de 2001 a 2003, correspondendo, portanto, à situação econômica desses anos. Não necessariamente essa configuração de salários será a observada nas próximas décadas. Porém, para simplificar as estimativas, assume-se que esse cenário não mudará demasiadamente, de forma que as estimativas aqui apresentadas não estarão muito destoantes da realidade.
- 4) Na estimação do fluxo futuro de renda do trabalho de um determinado indivíduo ao longo de sua vida, considerou-se apenas a evolução média da renda devido ao indivíduo estar em diferentes faixas de idade. Não está sendo considerada aqui a evolução na escolaridade dos cidadãos, o que implicaria aumento da renda do trabalho, de acordo com as superfícies apresentadas na figura 1. Portanto, as estimativas de renda futura estão de certa forma subestimadas, principalmente para as vítimas mais jovens, e portanto com maior probabilidade de migrar para níveis educacionais mais elevados. Esse fato implica subestimação das perdas de produção.
- 5) Na tabela 1, os primeiros 15 domínios geográficos correspondem a áreas urbanas, de acordo com os dados da Pnad. Foi preciso, então, definir correspondentemente as áreas geográficas urbanas para os dados do SIM. Assumiu-se que áreas urbanas no SIM corresponderiam aos municípios compondo as aglomerações urbanas definidas em estudo sobre a questão, desenvolvido pelo Ipea, IBGE e Unicamp (2002). Essa definição de aglomerações urbanas vem sendo utilizada em diversos trabalhos recentes, como, por exemplo, Da Mata *et al.* (2005a e 2005b) e Carvalho, Timmins e Lall (2006).

3 Resultados

Nesta seção, discutem-se os principais resultados para a estimação das perdas de produção em decorrência das mortes por homicídios, acidentes de transporte, suicídios e causas externas em geral. Inicialmente, apresentam-se os resultados para as estimativas das superfícies de renda anual média de todos os trabalhos *versus* idade e nível de escolaridade. Em seguida, são apresentadas as principais estimativas de perda de produção, a partir do valor presente do fluxo esperado de renda do trabalho, para todos os indivíduos registrados no SIM, em determinados anos, para determinadas categorias de óbitos.

3.1 Superfícies de renda *versus* idade e escolaridade

A Figura 2 apresenta as superfícies de renda do trabalho anual *versus* idade e escolaridade, para trabalhadores do sexo masculino, nas áreas urbanas nos estados do Rio de Janeiro, Espírito Santo, São Paulo (excluindo região metropolitana), Paraná, Santa Catarina e Rio Grande do Sul. Os valores de renda do trabalho anual estão em R\$ de setembro de 2001 (mês de referência da Pnad/2001). Essas superfícies correspondem à renda média de todos os trabalhos, exclusivamente para as pessoas que exerciam alguma atividade na semana de referência da Pnad. Figuras simulares foram geradas para as demais subamostras.

Diversas regressões preliminares, utilizando modelos paramétricos, comprovaram a significância estatística das variáveis sexo e escolaridade para explicar a renda salarial média. Por outro lado, graças à utilização de uma estimação via expansão de *B-splines*, é possível observar a interação existente entre essas duas covariáveis. Além disso, a estrutura de interação varia conforme o domínio geográfico e o sexo do trabalhador. Nas superfícies apresentadas na Figura 2, nota-se que a renda não necessariamente aumenta com a idade para todas as faixas de escolaridade. Para Santa Catarina e Paraná, nas áreas urbanas, para faixas de escolaridade mais baixas, o salário médio do trabalhador parece ter um comportamento crescente até uma certa idade, sofrendo uma redução para idades mais elevadas. Essa observação pode estar relacionada ao fato de que, para faixas de escolaridade mais baixas, os empregos são basicamente braçais, nos quais a juventude do trabalhador faz diferença. Para faixas de escolaridade mais altas, as atividades são menos braçais, e mais intelectuais, de forma que pessoas de idade mais elevadas não sofrem perdas salariais. Os padrões observados nas superfícies na Figura 2 são encontrados, com menos ou mais intensidade, nos gráficos para as demais subamostras.

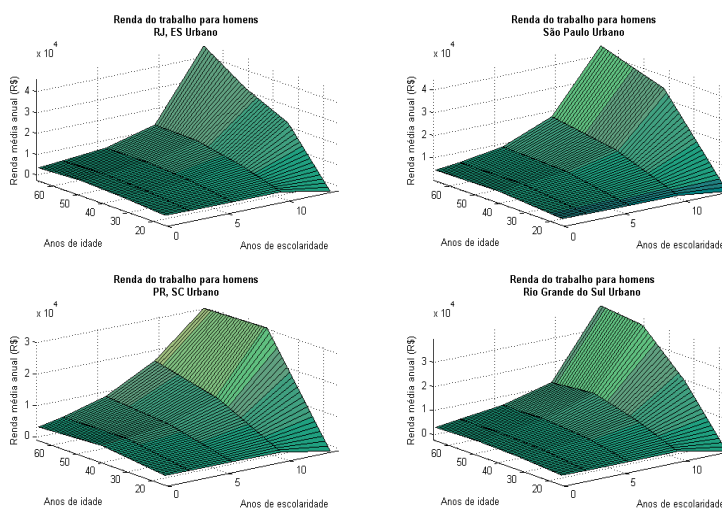


Figura 2 - Superfícies de renda anual do trabalho *versus* idade e escolaridade para trabalhadores do sexo masculino.

3.2 Estimativas dos custos de produção

As Tabelas A1 a A4, no Anexo, apresentam os principais resultados da estimação da perda de produção devido a mortes por causas externas, e especificamente para homicídios, suicídios e acidentes de transporte. Os resultados nessas tabelas levam em conta uma taxa^[1] de desconto anual de 3%. Além das estimativas de perda de produção agregadamente para o Brasil, as tabelas também apresentam os valores desagregados por unidade da federação. As colunas 2 e 4 dessas tabelas apresentam o número de vítima registradas em cada categoria de óbito, por unidade da federação de residência, nos anos 2000 e 2001.

Conforme discutido anteriormente, a estimação das perdas de produção baseia-se em um conjunto de co-variáveis, com base nas quais atribuímos curvas de renda do trabalho às vítimas registradas no SIM. Essas covariáveis são sexo, nível de escolaridade, idade e local de residência. Para as variáveis sexo, idade e local de residência, a proporção de observações, em que essas variáveis estão ausentes (*missing values*) é praticamente nula. Por outro lado, para a variável grau de escolaridade da vítima, em uma parcela muito grande das observações essa informação não está disponível. Conforme observado na Figura 2, existe uma relação positiva entre escolaridade e nível de renda (foram estimadas regressões lineares da renda versus diversas covariáveis, e o coeficiente da variável escolaridade resultou positivo e estatisticamente significativo).

Para contornar o problema da informação ausente para a variável educação, foram adotadas duas abordagens complementares. A primeira abordagem visa a estimar um limite inferior para a perda total de produção, seguindo as idéias em Manski (1999 e 2000). Dado que menos escolaridade implica salários menores, um limite inferior para a estimativa do custo de capital pode ser obtido pela utilização do valor escolaridade igual a zero para todas as observações (vítimas registradas) em que a educação não está disponível na base. Esses valores estão apresentados nas colunas 3 e 6 das tabelas A1 a A4.

A segunda abordagem utilizada para contornar o problema da ausência de informação para a variável educação foi imputar o valor presente do fluxo de salários, de acordo com as demais variáveis disponíveis. Por exemplo, considere-se uma vítima registrada no SIM, do sexo feminino, residente da região Norte urbana, falecida aos 21 anos de idade, da qual não conhecemos o grau de escolaridade. O valor presente do fluxo das rendas do trabalho imputado para essa vítima é dado pela média dos valores presentes para as vítimas do mesmo sexo, residindo na mesma área geográfica, falecidas com a mesma idade (das quais a informação de nível educacional existe, de forma que foi possível estimar o valor presente do fluxo de rendas futuras). Os resultados estão apresentados nas colunas 4 e 7 das tabelas A1 a A4. Assim como no caso das estimativas de limites inferiores para a perda total de produção, foram calculados valores tanto agregados quanto valores por unidade da federação.

Para ter idéia do erro obtido nessa imputação, utilizaram-se simulações de Monte Carlo. Ao invés da imputação direta pela média estimada para os indivíduos da base em que se conhecem os níveis educacionais, foram simulados valores aleatórios a partir de uma distribuição com média igual à média da imputação e variância igual à variância

^[1] Para resultados similares considerando-se taxas anuais de desconto de 6% e de 12%, vide Carvalho e Cerqueira (2008).

amostral para os indivíduos do mesmo sexo, na mesma área geográfica, com a mesma idade ao falecer. A distribuição paramétrica escolhida para gerar os valores simulados foi distribuição *gamma* (vide Roussas, 1997), que garante que todos os valores simulados sejam não-negativos. Os resultados estão apresentados na tabela 2.

Observando os valores nas tabelas 2 e A1 a A4, nota-se que os valores médios de acordo com as simulações são praticamente os mesmos dos valores médios imputados diretamente, e apresentados nas tabelas A1 a A4, nas colunas 4 e 7. Isso sugere a validade dos exercícios de simulação. Por outro lado, observando os valores nas colunas 3 e 4 na Tabela 2, nota-se que o intervalo entre os percentis 5% e 95% é relativamente estreito, indicando baixo grau de imprecisão na imputação pela média. Dado que a imprecisão é relativamente baixa, quando se olham os números agregados de perda de produção para o Brasil, isso sugere, em princípio, certa confiança nos valores das médias imputadas.

O potencial problema em se utilizar os valores médios de custo de capital dos indivíduos similares, para imputar o valor de custo de capital para as vítimas registradas sem o grau de escolaridade, é que esse procedimento parte de hipótese que não pode ser testada a partir das informações disponíveis. Implicitamente, a imputação pela média assume que o processo estocástico, que determina se um indivíduo terá o seu óbito com ou sem registro de escolaridade, não depende de outras covariáveis além das observadas na base de dados (ou seja, sexo, idade e grupo geográfico). Caso haja outros fatores determinando se uma vítima será registrada sem informação de escolaridade, e esse outro valor não é razoavelmente previsto pelas variáveis sexo, escolaridade e área geográfica, a estimativa do valor do custo de capital via imputação pela média pode estar levando a estimativas viesadas para os valores de perda de produção. Por esse motivo, implementaram-se neste trabalho as estimações de limites inferiores para o custo de capital, apresentadas nas colunas 3 e 6 das tabelas A1 a A4.

Tabela 2 - Resultados das simulações para inferir a incerteza devido aos valores ausentes para a variável grau de escolaridade

Taxa de desconto de 3% ao ano	Resultados simulações (bilhões R\$) ^[1]		
	Média	Percentil 5%	Percentil 95%
Ano 2000			
Acidentes de Trânsito	5,04	5,01	5,07
Homicídios	8,80	8,77	8,83
Suicídios	1,07	1,06	1,08
Causas Externas	19,72	19,67	19,77
Ano 2001			
Acidentes de Trânsito	5,40	5,37	5,43
Homicídios	9,13	9,10	9,16
Suicídios	1,28	1,26	1,29
Causas Externas	20,15	20,10	20,20

Elaboração dos autores

^[1] Valores monetários em R\$ de agosto de 2006.

As estimativas foram utilizadas com base nos registros de óbitos do SIM para os anos de 2000 e 2001. Uma das grandes vantagens em se utilizar dois anos consecutivos é que podem-se comparar os resultados agregados e conferir se eles estão próximos – o que é o esperado, dado que não se espera que haja mudanças muito bruscas na distribuição de vítimas de um ano para o outro. De fato, os resultados nas tabelas 2 e A1 a A4 indicam que os valores para as perdas estimadas de produção estão muito próximos entre 2000 e 2001. À época do início desta pesquisa, esses eram os últimos dois anos para os quais os registros dos microdados do Datusus pareceram mais confiáveis. À medida que bases de dados anuais mais recentes forem sendo disponibilizadas, a metodologia descrita aqui pode ser empregada para gerar números mais atualizados. Em todo caso, comparando-se os números entre 2000 e 2001, observa-se que não há diferenças significativas, levando a acreditar que as estimativas mais atualizadas não estarão muito distantes, pelo menos em ordem de grandeza, das estimativas apresentadas neste artigo.

Para uma taxa de desconto anual de 3%, obteve-se estimativa para a perda total de produção para o ano de 2001, somando-se homicídios, acidentes de transporte e demais causas externas, em torno de R\$ 20,1 bilhões. Desagregando-se por categorias de causas externas, os homicídios foram responsáveis por R\$ 9,1 bilhões desse custo total. Acidentes de transporte resultaram em estimativa de perda total para o Brasil em torno de R\$ 5,4 bilhões, enquanto suicídios resultaram em estimativa de R\$ 1,3 bilhão. Todos os valores monetários estão em R\$ de agosto de 2006. Em termos de valores médios de perda de produção, os homicídios apresentam as maiores médias: R\$ 189,5 mil para o ano de 2001. Acidentes de transporte possuem média em torno de R\$ 172 mil em 2001, e suicídios, em torno de R\$ 163 milhares por vítima. As tabelas A1 a A4 indicam que essas médias de fato podem mudar de ano para ano, dependendo da composição da escolaridade, idade, sexo e local de residência das vítimas. Em todo caso, para os diferentes anos e taxas de desconto, a média da perda de produção decorrente dos homicídios aparenta ser maior do que a média decorrente dos acidentes de transporte. A partir da tabela A4, a perda média de produção por vítima para causas externas em geral é de R\$ 165,2 milhares por ano. Homicídios, acidentes de transporte e suicídios contabilizam 87,4 mil (72%) dos 121,9 mil registros para causas externas. Portanto, os demais tipos de causas externas devem estar puxando a média geral para baixo.

Os valores discutidos acima se baseiam em taxa de desconto anual de 3% ao ano. Mota (2006) apresenta discussão extensa sobre escolha de taxa de desconto social: taxa considerada socialmente desejável para substituir o consumo no presente pelo consumo no futuro. Apesar de a discussão em Mota (2006) aplicar-se diretamente a valoração de projetos ambientais, a principal mensagem é que o problema de escolha da taxa de desconto é bastante controverso, com diferentes estudiosos sugerindo abordagens diferentes. Por exemplo, o Departamento do Tesouro Britânico recomenda taxa de desconto de 6% para muitas propostas de projetos. No único estudo encontrado que mede o custo de bem-estar da violência para 73 países, devido a Soares (2004), a taxa de desconto utilizada foi de 3%. Por conta disso, os resultados apresentados neste artigo levam em conta também taxa de 3% ao ano.

Finalmente, abstraindo-se de uma análise totalmente monetária em termos de perda de produção, foi calculado, alternativamente, o total de anos perdidos devido a mortes por homicídios ou acidentes de transporte. A idéia é calcular qual seria o valor esperado de anos vividos EA_k por cada uma das vítimas, caso ela não tivesse falecido devido ao evento

registrado no SIM. Nesse caso, pode-se empregar a fórmula a seguir, muito semelhante à expressão (10),

$$EA_k = \sum_{x=D_k}^T \Pr(F_k > x / F_k \geq D_k). \quad (11)$$

Os resultados estão apresentados na tabela 3 a seguir. Os homicídios e os acidentes de transporte ocasionam total de anos perdidos em torno de 3,4 milhões, utilizando os registros no SIM de 2001. Os homicídios são responsáveis por 2,15 milhões de anos perdidos, enquanto os acidentes de transporte são responsáveis por 1,24 milhão de anos perdidos das vítimas no Brasil. No agregado, as causas externas são responsáveis por quase cinco milhões de anos perdidos, em 2001 apenas.

Tabela 3 - Número de anos perdidos devido às mortes por causas externas

Unidade da federação	Total de anos perdidos (milhares)	
	2000	2001
Acidentes de transporte	1198,6	1240,9
Homicídios	2040,1	2148,1
Suicídios	249,5	289,7
Causas externas	4920,2	4964,9

Elaboração dos autores.

4 Comentários finais

Este trabalho procurou estimar a perda de produção em decorrência das mortes por causas externas, a partir da estimação do fluxo de renda do trabalho esperada que a vítima obteria, caso não tivesse falecido. Para estimar o valor futuro do fluxo de renda, foram utilizados microdados da Pesquisa Nacional por Amostras de Domicílios – Pnad. Para se ampliar o conjunto de informação utilizado para estimar o custo de perda de produção, combinaram-se as bases da Pnad para os anos de 2001, 2002 e 2003^[1]. Os microdados da Pnad foram utilizados para estimar funções de renda média do trabalho *versus* idade, grau de escolaridade, sexo e área geográfica. Essas funções foram, então, utilizadas para gerar o fluxo de renda média do trabalho esperado para as vítimas na base do SIM, do Ministério da Saúde.

Para a estimação das funções médias de renda do trabalho *versus* idade, sexo, área geográfica e nível de escolaridade, utilizou-se abordagem diferente da abordagem tradicionalmente utilizada na literatura de estimação de funções de salário. Ao invés da utilização de especificação puramente linear paramétrica, optou-se por metodologia mais flexível, combinando divisão da amostra total em subgrupos com modelos não-paramétricos. A utilização da metodologia empregada deveu-se, entre outros motivos, à grande quantidade de observações disponíveis nas Pnads, possibilitando estimar as formas

^[1] Escolheram-se essas três Pnads por motivo de compatibilização temporal com os dados do SIM disponíveis à época do estudo. Os autores pretendem refazer as estimativas após dez anos para checar a ocorrência de alterações estruturais nos custos da violência no Brasil.

funcionais desconhecidas, além de capturar interações entre as diversas covariáveis. Em diversas estimações preliminares, ficou comprovada a significância estatística das quatro covariáveis utilizadas (idade, sexo, localização e escolaridade) e da interação existente entre elas. Além disso, essas são variáveis reconhecidamente importantes para explicar renda, de acordo com a literatura de mercado de trabalho (vide Ashenfelter e Card, 1999).

Primeiramente dividiu-se a amostra total de dados (com os três anos da Pnad empilhados) em 40 subgrupos, de acordo com o sexo e com a área geográfica do trabalhador. Em seguida, em cada uma dessas 40 subamostras, foram estimadas superfícies de renda do trabalho *versus* idade e nível de escolaridade, utilizando-se técnicas de estimação não-paramétrica. Isso possibilitou evitar a previsão de valores negativos para a renda do trabalho, para alguns determinados grupos populacionais (combinações de sexo, área geográfica, idade e escolaridade). Estimativas preliminares, utilizando modelos lineares paramétricos tradicionais, incorreram em até mais de 25% de valores negativos para a previsão de renda do trabalho nos indivíduos da própria amostra utilizada para previsão.

Estimou-se um custo total de perda de produção em torno de R\$ 20,1 bilhões (em valores de agosto de 2006), utilizando dados do SIM de 2001. Desagregando-se por categorias de causas externas, os homicídios são responsáveis por R\$ 9,1 bilhões desse custo total. Acidentes de transporte vêm em segundo lugar, com estimativa de custo total para o Brasil em torno de R\$ 5,4 bilhões. Em termos de valores médios de perda de produção, os homicídios apresentam as maiores médias: R\$ 189,5 mil para o ano de 2001. Já acidentes de transporte possuem perda de produção média por vítima em torno de R\$ 172 mil. Os maiores valores médios para as mortes por homicídios devem-se ao fato de as vítimas serem predominantemente do sexo masculino (com rendimentos de trabalho mais altos) e os óbitos terem maior incidência nas vítimas de faixa etária mais baixa. Além disso, maior proporção das vítimas de homicídios reside nas áreas urbanas, onde os rendimentos são maiores. Finalmente, contabilizando o número total de anos perdidos devido a homicídios e acidentes de transporte, o total em 2001 chega a 3,4 milhões. Homicídios apenas são responsáveis por 2,15 milhões de horas perdidas para as vítimas no Brasil.

Além dos custos decorrentes de perda de capital humano, outros custos sociais decorrentes da violência no Brasil precisam ser estimados. A oferta de bens e serviços não é apenas impactada pela perda de capital humano, mas pelo aumento dos custos marginais, em face da necessidade de as empresas despenderem recursos para a autoproteção. Do ponto de vista da demanda, o medo, a dor e o sofrimento mudam o comportamento, muitas fazem diminuir a demanda por determinados bens e serviços e geram perdas patrimoniais nos mercados imobiliários localizados em regiões com maior prevalência de incidentes violentos. Entender mais precisamente as conseqüências e custos da violência no Brasil é, portanto, uma questão crucial para o balizamento de políticas públicas de saúde e de segurança que imponham uma lógica de racionalidade ao uso dos recursos públicos. Este artigo faz parte de uma série de estudos^[1], em que os autores procuram estimar outros custos que derivam da violência no Brasil.

[1] Vide Cerqueira *et al.* (2007) e Rodrigues *et al.* (2008).

Agradecimentos:

Este trabalho foi financiado pelo Projeto de Pesquisa “Análise de Custos Econômicos e Sociais da Violência no Brasil”, com recursos do Ministério da Saúde e do CNPq. Os autores agradecem às assistentes de pesquisa Vivian Almeida e Michelle Guimarães pela dedicação e a Carlos Octávio Ocké Reis, Marcelo Abi-Ramia Caetano, Marcelo Medeiros e José Aroudo Mota por valiosos comentários. Todos os erros remanescentes no trabalho são de inteira responsabilidade dos autores.

CARVALHO, A. X. Y.; CERQUEIRA, D. R. C.; RODRIGUES, R. I.; LOBÃO, W. J. A. Production costs of deaths by external causes in Brazil. *Rev. Bras. Biom.*, São Paulo, v.26, n.3, p.7-21, 2008.

- **ABSTRACT:** *Violence in Brazil is well known to be one of the biggest issues nowadays affecting society. This problem incurs in several economic costs, in addition to all kinds of immeasurable costs due to life losses. For the economic costs, one can mention medical treatment costs, police maintenance costs, production costs. This paper focuses on the latest cost type, and a methodology to estimate the production loss due to deaths from external causes in Brazil is presented, using government databases on wages and registered obits. To combine different sources of data, non-parametric regression methods are initially employed to estimate average surfaces of annual wages. These surfaces are then used to estimate the production loss for specific categories of registered obits. The analysis is complemented by adjustments for survival rates in the Brazilian population. For 2001, the total estimated cost, due to external causes, was US\$ 10 billion. Homicides alone accounted for US\$ 4.5 billion, while traffic accidents accounted for US\$ 2.6 billion and suicides incurred in a total loss of US\$ 0.6 billion. Estimations for total loss in terms of life expectancy are also presented: the total number was almost 5 million years lost, only for the victims in 2001. One of the main reasons for these figures is the high percentage of young victims.*
- **KEY-WORDS:** *Production costs; homicides; traffic accidents; non-parametric methods; wage surfaces.*

Referências

- ALPAYDIN, E. *Introduction to machine learning*. Cambridge: The MIT Press, 2004. 445p.
- ASHENFELTER, O.; CARD, D. *Handbook of labor economics*. Amsterdam: North-Holland, 1999. v.3, 928p.
- ATKINSON, G.; HEALEY, A.; MOURATO, S. Valuing the costs of violent crime: a stated preference approach. *Oxford Econ. Papers*, Oxford, v. 57, n.4, p.559-585, 2005.
- BURNHAN, K.; ANDERSON, D. *Model selection and inference: a practical information-theoretic approach*. New York: Springer, 1998. 353p.
- CAETANO, M. A. *Subsídios cruzados na previdência social brasileira*. Brasília: IPEA. 2006. 25p. (Texto para Discussão, 1.211).

- CARVALHO, A. X.; CERQUEIRA, D. Estimação da perda de produção devido a mortes por causas externas nas cidades brasileiras. In: CARVALHO, A. X.; ALBUQUERQUE, C. W.; MOTA, J. A.; PIANCASTELLI, M. (ORG.) *Ensaio de economia regional e urbana* organizadores. Brasília: IPEA, 2008. 460p.
- CARVALHO, A. X.; TIMMINS, C.; LALL, S. V. *Regional subsidies and industrial prospects for lagged regions*. Washington: Banco Mundial, 2006. 48p. (Texto para Discussão, 3.843).
- CERQUEIRA, D.; CARVALHO, A. X.; LOBÃO, W.; RODRIGUES, R. I. *Análise dos custos e conseqüências da violência no Brasil*. Brasília: IPEA, 2007. 61p. (Texto para Discussão, 1284).
- COHEN, M. A.; RUST, R. T.; STEEN, S.; TIDD, S. T. Willingness to pay for crime control programs. *Criminology*, Jerusalém, v.42, n 1, p. 89-109, 2004.
- COOK, P. J.; LUDWIG, J. *Gun violence: the real costs*. Oxford: Oxford University Press. 2000. 242p.
- CHRISTE, N. G. S.; SOGUEL, N. C. *Contingent valuation, transport safety and the value of life*. studies in risk and uncertainty. New York: Springer, 1995. 208p.
- DA MATA, D.; DEICHMANN, U.; VERNON HENDERSON, J.; LALL, S. V.; WANG, H. G. *Determinants of city growth in Brazil*. Brasília: IPEA. 2005. 39p. (Discussion Paper, 1.112).
- _____. *Examining the growth patterns of Brazilian cities*. Brasília: IPEA. 2005. 37p. (Discussion Paper, 1.113).
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*. data mining inference and prediction. New York: Springer, 2001. 552p.
- INSTITUTO DE PESQUISA ECONÔMICA APLICADA. *Configuração atual e tendências da rede urbana*. Brasília: IPEA. 2002. 396p. (Série Configuração Atual e Tendências da Rede Urbana).
- LUDWIG, J.; COOK, P. The benefits of reducing gun violence: evidence from contingent-valuation survey data. *J. Risk Uncertainty*, New York, v.22, n.3, p.207-26, 2001.
- MANSKI, C. *Identification problems in the social sciences*. Harvard: Harvard University Press, 1999. 194p.
- _____. Monotone instrumental variables: with an application to the returns to schooling. *Ver. Econ.*, Rio de Janeiro, v.68, n.4, p.997-1.010, 2000.
- MINCER, J. *Schooling, experience, and earnings*. New York: NBER Press, 1974. 167p.
- MOTA, J. A. *O valor da natureza*. 2. ed. Rio de Janeiro: Garamond, 2001. 167p.
- NAGIN, D. S.; PIQUERO, A. R. *Public preferences for rehabilitation versus incarceration of juvenile offenders: evidence from a contingent valuation survey*. University of Virginia Law School. The John M. Olin Program in Law and Economics, 2006, 41p. (Working Paper Series, n. 28).

- RODRIGUES, R. I.; CERQUEIRA, D.; LOBÃO, W.; CARVALHO, A. X. Os custos da violência para o sistema público de saúde no Brasil: informações disponíveis e possibilidades de estimação. *Cadernos de Saúde Pública*, Rio de Janeiro, 2008. No prelo.
- ROSEN, S. The value of changes in life expectancy. *J. Risk. Uncertainty*, New York, v.1, p.285-304, 1988.
- ROUSSAS, G. G. *A course in mathematical statistics*. 2. ed. San Diego: Academic Press, 1997. 572p.
- SILVEIRA, F. G.; CARVALHO, A. X.; AZZONI, C. R.; CAMPOLINA, B.; IBARRA, A. *Dimensão, magnitude e localização das populações pobres do Brasil*. Brasília: IPEA, 2007. 67p. (Texto para Discussão).
- SOARES, R. R. *The welfare cost of violence*. Department of Economics, University of Maryland. College Park, 2004. 41p. Mimeografado.
- SOBY, B. A.; BALL, D. J. Valuing consumer safety. *Int. J. Injury Control Safety Promot.* London, v.2, n.3, p.117-131, 1995.
- USHER, D. An imputation of the measure of economic growth for changes in life expectancy. Studies in income and wealth performance. In: CONFERENCE ON RESEARCH INCOME AND WEALTH, 38., 1973, New York. *Abstracts...* Columbia, University Press, p.193-225, 1973.
- VISCUSI, W. K.; ALDY, J. E. *The value of a statistical life: a critical review of market estimates throughout the world*. *J. Risk. Uncertainty*, New York, v.27, p.5-76, 2003.
- ZARKIN, G. A.; CATES, S. C.; BALA, M. V. Estimating the willingness to pay for drug abuse treatment. *J. Substance Abuse Treat.*, New York, v.18, p.149-159, 2000.

Recebido em 09.05.2008.

Aprovado após revisão 29.08.2008.

ANEXO

Estimativas de Custos de Capital Humano devido a Mortes por Causas Externas

Tabela A1 - Custos de produção devido a acidentes de transporte

Unidade da federação	2000			2001		
	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)
Rondônia	328	41.239	54.460	323	43.833	55.892
Acre	92	11.165	12.957	102	9.834	13.627
Amazonas	347	53.905	56.718	274	39.273	42.081
Roraima	134	27.785	29.272	128	26.811	27.291
Pará	707	61.218	73.324	811	84.674	103.716
Amapá	99	19.402	20.478	113	28.904	29.245
Tocantins	293	26.230	32.829	287	23.688	29.172
Maranhão	484	43.858	53.391	564	56.943	65.803
Piauí	800	80.096	90.130	810	91.087	101.015
Ceará	1.266	88.009	120.451	1.366	101.995	148.895
Rio Grande do Norte	479	36.357	44.076	425	34.423	41.173
Paraíba	442	30.187	38.168	445	28.960	37.344
Pernambuco	1.483	104.707	178.317	1.345	89.752	148.362
Alagoas	549	29.503	50.898	553	28.188	49.190
Sergipe	349	19.304	33.154	347	23.811	32.961
Bahia	1.206	101.890	131.378	1.250	108.634	142.870
Minas Gerais	2.535	290.093	388.437	2.734	323.696	425.494
Espírito Santo	826	93.609	152.885	817	94.323	147.329
Rio de Janeiro	2.582	400.884	490.132	2.719	410.274	496.080
São Paulo	5.975	871.118	1.220.708	6.972	1.111.388	1.498.056
Paraná	2.568	408.703	467.956	2.543	412.516	458.926
Santa Catarina	1.503	207.987	287.015	1.561	212.626	310.196
Rio Grande do Sul	1.931	243.324	371.466	1.864	234.045	355.269
Mato Grosso do Sul	415	56.365	79.516	503	67.269	92.214
Mato Grosso	718	130.405	153.985	713	128.143	152.286
Goiás	1.406	151.140	246.408	1.351	147.409	233.273
Distrito Federal	523	130.906	161.830	490	138.024	165.751
BRASIL	30.040	3.759.388	5.040.339	31.410	4.100.526	5.403.511
Média por vítima	---	125,1	167,8	---	130,5	172,0

Obs.: Valores monetários em milhares de R\$ de agosto de 2006.

Tabela A2 - Custos de produção devido a homicídios

Unidade da federação	2000			2001		
	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)
Rondônia	466	59.162	79.457	573	74.988	101.858
Acre	106	12.489	13.881	121	14.911	19.413
Amazonas	553	85.600	89.196	480	73.593	77.608
Roraima	130	22.601	22.946	108	18.113	19.080
Pará	807	85.944	101.410	965	99.094	116.265
Amapá	156	38.637	39.407	182	44.708	44.826
Tocantins	173	14.677	17.219	209	20.806	25.062
Maranhão	351	41.168	45.689	562	68.307	78.873
Piauí	460	42.768	48.838	524	56.156	63.324
Ceará	1.232	93.852	127.816	1.284	100.157	137.261
Rio Grande do Norte	257	21.779	25.875	322	26.238	30.955
Paraíba	507	40.941	48.535	484	36.472	45.078
Pernambuco	4.276	308.799	456.769	4.697	337.308	497.195
Alagoas	727	44.097	66.165	830	50.715	77.546
Sergipe	409	26.979	40.346	517	40.451	53.130
Bahia	1.242	120.294	142.966	1.616	161.745	187.607
Minas Gerais	2.106	282.302	324.219	2.365	334.324	381.882
Espírito Santo	1.432	169.367	285.233	1.452	173.859	281.672
Rio de Janeiro	7.328	1.352.026	1.625.225	7.349	1.282.049	1.546.275
São Paulo	15.581	3.264.143	3.765.134	15.731	3.270.363	3.819.227
Paraná	1.779	292.186	323.463	2.034	349.629	380.544
Santa Catarina	424	56.705	75.108	466	59.799	83.859
Rio Grande do Sul	1.664	220.628	312.263	1.853	240.592	330.919
Mato Grosso do Sul	650	91.350	122.995	621	93.262	124.158
Mato Grosso	990	170.547	199.616	973	173.339	196.561
Goiás	1.080	138.382	199.924	1.151	144.703	202.117
Distrito Federal	687	172.288	202.988	692	184.072	206.333
BRASIL	45.573	7.269.711	8.802.684	48.161	7.529.755	9.128.628
Média por vítima	---	159,5	193,2	---	156,3	189,5

Obs.: Valores monetários em milhares de R\$ de agosto de 2006.

Tabela A3 - custos de produção devido a suicídios

Unidade da federação	2000			2001		
	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)
Rondônia	76	9.344	11.188	76	9.819	11.563
Acre	27	3.050	3.262	29	3.137	4.811
Amazonas	77	10.314	10.666	94	13.799	14.499
Roraima	22	3.713	3.753	17	2.898	2.898
Pará	116	12.045	14.396	146	13.318	17.536
Amapá	16	2.906	2.933	30	7.641	7.847
Tocantins	38	3.567	5.053	47	3.396	4.367
Maranhão	71	6.945	8.680	109	14.823	17.338
Piauí	148	13.848	15.655	228	27.241	29.800
Ceará	269	16.402	21.929	376	29.726	41.677
Rio Grande do Norte	85	5.440	7.032	116	7.487	9.022
Paraíba	37	2.575	3.310	45	2.759	3.466
Pernambuco	273	16.304	25.994	268	17.222	27.369
Alagoas	72	3.424	6.637	96	4.824	9.479
Sergipe	48	2.925	5.029	69	4.482	5.783
Bahia	202	12.519	14.533	240	16.832	20.562
Minas Gerais	603	71.838	92.707	800	97.815	133.285
Espírito Santo	106	11.062	19.959	116	13.055	21.889
Rio de Janeiro	392	66.909	82.760	457	82.377	100.691
São Paulo	1.413	232.646	302.147	1.637	255.132	339.386
Paraná	586	83.311	92.073	664	100.305	108.266
Santa Catarina	426	46.893	65.816	460	48.830	68.190
Rio Grande do Sul	1.022	101.416	136.578	1.035	112.870	162.511
Mato Grosso do Sul	173	21.354	25.803	138	17.572	21.721
Mato Grosso	142	20.521	25.039	144	21.283	24.785
Goiás	326	30.456	41.669	327	31.857	44.236
Distrito Federal	86	18.489	22.256	79	22.152	25.516
BRASIL	6.852	830.216	1.066.855	7.843	982.652	1.278.492
Média por vítima	---	121,2	155,7	---	125,3	163,0

Obs.: Valores monetários em milhares de R\$ de agosto de 2006.

Tabela A4 - Custos de produção devido a causas externas em geral

Unidade da federação	2000			2001		
	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)	Núm. registros	Lim. inferior (mil R\$)	Média estimada (mil R\$)
Rondônia	1.221	149.063	195.894	1.358	174.054	228.872
Acre	302	33.918	38.845	321	34.107	46.071
Amazonas	1.352	190.517	202.174	1.243	169.964	180.494
Roraima	348	61.305	63.406	328	58.754	61.265
Pará	2.234	203.241	244.059	2.632	253.337	305.087
Amapá	339	71.032	73.372	389	92.930	93.909
Tocantins	652	53.706	66.395	705	59.229	71.132
Maranhão	1.611	158.680	186.368	1.832	192.431	220.647
Piauí	2.070	186.207	214.002	2.204	219.175	249.230
Ceará	3.941	266.407	364.117	4.073	286.778	403.035
Rio Grande do Norte	1.530	119.450	145.668	1.452	105.991	128.833
Paraíba	1.391	96.523	121.376	1.242	81.039	103.283
Pernambuco	7.423	504.750	793.785	7.575	508.462	782.357
Alagoas	1.671	91.669	152.873	1.781	97.203	160.577
Sergipe	1.181	68.027	112.825	1.249	87.219	119.387
Bahia	6.299	535.748	639.788	6.483	578.073	692.880
Minas Gerais	8.204	905.698	1.131.771	8.867	1.011.791	1.259.676
Espírito Santo	2.936	323.891	523.697	2.972	328.942	515.024
Rio de Janeiro	14.857	2.327.727	2.833.263	15.024	2.279.979	2.742.353
São Paulo	33.896	5.858.462	7.108.351	34.009	5.839.453	7.145.153
Paraná	6.812	1.007.152	1.128.183	6.950	1.069.720	1.177.922
Santa Catarina	3.312	416.563	557.165	3.380	417.584	581.712
Rio Grande do Sul	6.555	744.643	1.049.533	6.602	763.707	1.092.600
Mato Grosso do Sul	1.645	208.107	277.703	1.653	216.832	288.161
Mato Grosso	2.376	388.557	454.624	2.338	393.752	450.866
Goiás	3.647	396.313	587.736	3.692	406.281	588.892
Distrito Federal	1.596	372.839	450.081	1.567	396.512	454.066
BRASIL	119.401	15.740.194	19.717.052	121.921	16.123.299	20.143.484
Média por vítima	---	131,8	165,1	---	132,2	165,2

Obs.: Valores monetários em milhares de R\$ de agosto de 2006.