

Elisenda Rentería Pérez

**Estimativas de mortalidade adulta feminina
por nível de escolaridade no Brasil**

Belo Horizonte, MG
UFMG/Cedeplar
2010

Elisenda Rentería Pérez

Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil

_____ apresentada ao curso de _____ em _____ do Centro de Desenvolvimento e Planejamento Regional da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do Título de _____ em _____.

Orientador: Prof. Cássio Maldonado Turra

Co-orientadora: Prof^a. Cristine Campos de Xavier Pinto

Belo Horizonte, MG
Centro de Desenvolvimento e Planejamento Regional
Faculdade de Ciências Econômicas - UFMG
2010

Folha de Aprovação

A mis padres, por ser como son

Ao Tatu, por estar sempre

*À minha avó, que morreu antes de eu
nascer, e nunca estudou*

AGRADECIMENTOS

Escrever os agradecimentos significa ter chegado ao fim de uma etapa longa e dura, mas que, felizmente, posso chamar de muito satisfatória. Etapa que não teria sido possível sem a ajuda recebida por todos os que caminharam do meu lado em todos esses anos. Por isso, chegou a hora do meu reconhecimento a todos.

Agradeço ao Professor Cássio Maldonado Turra, orientador desta Tese e de muitas outras pesquisas que realizei na demografia, seu critério, suas críticas e sugestões sempre pertinentes, sua exigência, sua generosidade e a confiança absoluta que sempre depositou em mim. Mas, sobretudo, agradeço tudo o que aprendi dele. O Cássio é o meu Mestre em demografia e um exemplo a seguir na pesquisa universitária, e espero sempre poder contar com o conselho dele, tão valioso para mim nesses anos todos. Obrigada pela dedicação, mas, sobretudo, por me ensinar a ser uma pesquisadora em demografia.

Agradeço à Cristine Campos de Xavier Pinto, co-orientadora, pela imensa ajuda oferecida durante os últimos meses. Os desafios metodológicos da Tese que enfrentei durante a última etapa foram superados graças à ajuda dela. Aprendi dela o rigor que é necessário para realizar um trabalho de qualidade.

Aos professores do Cedeplar, devo o que aprendi na área de demografia. Obrigada a cada um deles, por todos os ensinamentos, que sempre estimei de altíssima qualidade, e que às vezes extrapolavam as horas de aula. Obrigada, sobretudo, pela paciência em acolher uma estudante estrangeira, que mais de uma vez acabou gerando demandas extra-curriculares. Em especial, devo um agradecimento diferenciado a alguns professores que foram de especial ajuda neste doutorado. Em primeiro lugar à professora Diana Sawyer, orientadora inicial do meu projeto de doutorado, e de quem sempre recebi sugestões enriquecedoras sobre os meus trabalhos, que me ajudaram a melhorar minha pesquisa. Ao professor José Alberto Magno de Carvalho, pela dedicação, pelo exemplo e pela generosidade que sempre mostrou, mas também, por toda ajuda que sempre recebi dele em diversos dos meus trabalhos. À professora Simone Wajnman, que foi minha orientadora no mestrado e continuou sendo um exemplo a seguir, além de me oferecer mais a oportunidade de participar em algumas pesquisas que foram cruciais para minha formação. Finalmente, ao professor Bernardo Lanza Queiroz, que também foi meu orientador em outras pesquisas da universidade e de quem aprendi muito nestes últimos anos no Cedeplar. Além disso, tenho

um agradecimento especial à ajuda crucial do Mauricio Lima, que me ajudou tão generosamente com a elaboração dos programas que precisei usar para realizar vários dos cálculos desta Tese.

Aos membros da banca de defesa, Iúri da Costa Leite, Jerônimo Muniz, Simone Wajnman e José Alberto Carvalho, e da banca de qualificação, Bernardo Queiroz e Laura Wong, agradeço pelas críticas feitas a este trabalho e pelas sugestões dadas, que foram fundamentais no aprimoramento deste trabalho.

A todos os funcionários do Cedeplar, da Biblioteca e da FACE, devo agradecer a disponibilidade deles e a ajuda recebida em diversos momentos difíceis. Obrigada por estar sempre dispostos a resolver todas as questões administrativas, apesar do curto tempo com que as vezes precisei de algum serviço. Quero agradecer em especial aos funcionários da Secretaria de pós-graduação, Cecília, Sebastião, Andréa e Cleusa, aos funcionários da Biblioteca, aos responsáveis da área de computação e à Adriana, responsável da copiadora da FACE.

Devo agradecer aos meus colegas da Population Division, nas Nações Unidas, as valiosas sugestões que recebi deles durante minha curta estada lá, que enriqueceram ainda mais o tempo que passei lá.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), agradeço pelo suporte financeiro durante esses anos de estudo e pesquisa que me permitiu realizar minha pesquisa no doutorado com dedicação exclusiva e assistir a diversos congressos nacionais e internacionais.

Aos colegas do Cedeplar, agradeço a companhia nessa longa viagem e as alegrias e tristezas compartilhadas. Em especial, um agradecimento à Marisa e à Claudia, companheiras incansáveis de doutorado, com as que caminhei durante todos esses anos de demografia, obrigada pela amizade e pelo carinho! Também quero deixar um agradecimento especial a aqueles companheiros com os que compartilhei mais do que aulas, e que no trabalho e fora dele, sempre mostraram seu apoio incondicional quando precisei deles. Obrigada Maira, Daylin, Julio, Marisol, Almada, Laeticia, Izabel, Harley, Marden, Clarissa, Cristina, Denise e Everton.

Também devo um agradecimento aos amigos não ceceplarianos. Em particular a um grupo de amigas muito especial, com as que sempre pude contar durante esses anos, para compartilhar histerias, festas, alegrias, e também mágoas e decepções. Sem elas, esses anos de doutorado teriam sido mais tristes, mais cinzas. Elas fazem parte de minha nova família brasileira. Obrigada Nina, Bruna, Nadja, Gabi e Jana. Vocês foram demais.

À minha família, eu devo o apoio e o amor incondicional, mesmo que na distância, sem o que não teria conseguido terminar essa Tese. Aos meus pais, Feliciano y Maria Angeles, que sempre foram uma referência para mim na vida e no trabalho, obrigada por me inspirar, por me animar, e por sempre acreditar em mim. À minha irmã, Nerea, agradeço seu carinho e sua compreensão nos momentos mais difíceis do doutorado, e por me ajudar sempre a descontraír e desestressar.

Finalmente, o agradecimento mais especial. Ao Tatu, que me acompanhou em cada segundo do doutorado, onde teve momentos bons e não tão bons, mas ele nunca desistiu de mim. Por sempre estar lá me admirando e permitir que eu admirasse ele. Por ter a paciência de estar ao meu lado e estar sempre disposto a deixar as imagens e vídeos dele para aprender um pouco de demografia e bases de dados. Pela solidariedade, pelo amor, pelo carinho, pela convivência, pela calma. Por me fazer saber que podia contar com você sempre. Obrigada, cariño!

LISTA DE ABREVIATURAS E SIGLAS

PNAD – Pesquisa Nacional por Amostra de Domicílios

PPV – Pesquisa de Padrões de Vida

SUMÁRIO

1. INTRODUÇÃO	1
2. A LITERATURA RECENTE SOBRE OS DIFERENCIAIS SOCIOECONÔMICOS NA MORTALIDADE ADULTA	5
2.1. DEFINIÇÕES DE STATUS SOCIOECONÔMICO E MECANISMOS DE ASSOCIAÇÃO COM A MORTALIDADE	5
2.2. EVIDÊNCIAS EMPÍRICAS E QUESTÕES METODOLÓGICAS	11
2.2.1. Diferenciais socioeconômicos na mortalidade nos países desenvolvidos: magnitudes e tendências	11
2.2.2. Diferenciais socioeconômicos na mortalidade em países em desenvolvimento.....	17
2.2.3. Diferenciais socioeconômicos na mortalidade no Brasil: estratégias metodológicas e evidências empíricas	19
3. MÉTODOS INDIRETOS PARA ESTIMAR A MORTALIDADE ADULTA A PARTIR DE INFORMAÇÕES RETROSPECTIVAS DE SOBREVIVÊNCIA	24
3.1. A ORIGEM: LOTKA, HENRY E O MÉTODO PARA ESTIMAR O NÚMERO DE PARENTES SOBREVIVENTES.....	24
3.2. O MÉTODO DE ORFANDEDE.....	25
3.2.1. Avaliação da confiabilidade do método de orfandade	29
3.2.2. Aplicação do método de orfandade para estimar mortalidade adulta para subgrupos de população	31
3.2.3. Alocação no tempo das estimativas de mortalidade.....	33
4. MÉTODO ESTOCÁSTICO PARA ESTIMAR MORTALIDADE ADULTA.....	36
4.1. UMA VISÃO GERAL DO MÉTODO	36
4.2. ESTIMAR OS PARÂMETROS DO MÉTODO ESTOCÁSTICO	39
4.2.1. Idade de início do período de exposição	39
4.2.2. Idade final ao período de exposição	40
4.2.3. Período de exposição ao risco de morte	41

4.3. ANÁLISE DE SOBREVIVÊNCIA COM MODELOS DE REGRESSÃO	42
4.3.1. Coeficientes <i>bootstrap</i> das regressões Poisson.....	45
4.3.2. Alocação no tempo das estimativas de mortalidade.....	46
4.4. VANTAGENS E LIMITAÇÕES DO MÉTODO ESTOCÁSTICO.....	47
5. BASES DE DADOS	49
5.1. PESQUISA SOBRE PADRÕES DE VIDA.....	49
5.2. PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS	50
5.3. DADOS FALTANTES E ESTRATÉGIAS PARA O SEU TRATAMENTO.....	51
5.3.1. Tipos de valores faltantes	52
5.3.2. Métodos para tratar os valores faltantes	53
5.3.2.1. Métodos baseados em dados completos	54
5.3.2.2. Variável <i>dummy</i> para os dados faltantes.....	54
5.3.2.3. Métodos de Imputação.....	55
A. Imputação pela média.....	55
B. Imputação por regressão	56
C. Imputação por regressão estocástica	56
D. Imputação múltipla (IM)	57
E. Imputação por <i>hot deck</i>	58
5.3.2.4. O algoritmo de Expectativa e Maximização (EM).....	59
5.4. APLICAÇÃO DOS MÉTODOS DE TRATAMENTO DE DADOS FALTANTES DA VARIÁVEL DE ESCOLARIDADE DAS MÃES NA PPV E NA PNAD	61
5.4.1. Análise da informação sobre escolaridade das mães na PPV	61
5.4.2. Análise da informação sobre a escolaridade das mães na PNAD.....	64
5.4.3. Modelos de tratamento dos dados faltantes da variável escolaridade das mães	67
5.4.4. Resultados do tratamento dos dados faltantes da variável escolaridade das mães	71
6. ESTIMATIVAS DE MORTALIDADE ADULTA COM INFORMAÇÕES RETROSPECTIVAS DE SOBREVIVÊNCIA	76

6.1. ESTIMATIVAS DE MORTALIDADE ADULTA COM BASE NO MÉTODO DE ORFANDADE.....	76
6.1.1. Estimativas de mortalidade adulta por idade	76
6.1.2. Estimativas de mortalidade por nível de escolaridade	79
6.2. ESTIMATIVAS DE MORTALIDADE ADULTA COM BASE NO MÉTODO ESTOCÁSTICO	82
6.2.1. Estimativas de mortalidade adulta por idade	86
6.2.1. Estimativas de mortalidade adulta por nível de escolaridade.....	89
6.2.3. Estimativas de mortalidade por coorte de nascimento	94
6.3. Comparação dos resultados do método de orfandade com os do método estocástico ..	97
7. CONCLUSÃO	100
REFERÊNCIAS BIBLIOGRÁFICAS.....	106
ANEXOS.....	106

LISTA DE ILUSTRAÇÕES

FIGURA 1: ASSOCIAÇÕES ENTRE AS CONDIÇÕES SOCIOECONÔMICAS E A MORTALIDADE E MORBIDADE	7
FIGURA 2. DIAGRAMA DE LEXIS REPRESENTANDO A ESTIMAÇÃO DOS PARÂMETROS DO MÉTODO ESTOCÁSTICO.....	38
TABELA 1. DISTRIBUIÇÃO DA RESPOSTA E NÃO RESPOSTA DA VARIÁVEL ESCOLARIDADE DAS MÃES, SEGUNDO DIVERSAS CARACTERÍSTICAS DOS ENTREVISTADOS NA PPV 1996	64
TABELA 2. DISTRIBUIÇÃO DA RESPOSTA E NÃO RESPOSTA DA VARIÁVEL DE ESCOLARIDADE DAS MÃES, SEGUNDO DIVERSAS CARACTERÍSTICAS DOS ENTREVISTADOS NA PNAD 1996.....	67
TABELA 3. DISTRIBUIÇÃO DOS VALORES IMPUTADOS DA VARIÁVEL ESCOLARIDADE DAS MÃES, SEGUNDO O TIPO DE IMPUTAÇÃO, PPV E PNAD, 1996.....	72
TABELA 4. PROPORÇÃO DE MÃES MORTAS SEGUNDO AS CATEGORIAS DA VARIÁVEL ESCOLARIDADE DAS MÃES, POR TIPO DE TRATAMENTO DOS VALORES FALTANTES, PPV E PNAD, 1996.....	74
TABELA 5. BRASIL - ESTIMATIVAS DE PROBABILIDADES CONDICIONADAS DE SOBREVIVER DE MULHERES, ATRAVÉS DO MÉTODO DE ORFANDADE, PPV E PNAD, 1996.....	77
TABELA 6. BRASIL - ESTIMATIVAS DAS PROBABILIDADES CONDICIONAIS DE SOBREVIVER DAS MULHERES PELO MÉTODO DE ORFANDADE, ANO DE REFERÊNCIA E ESPERANÇA DE VIA AO NASCER	79
TABELA 7. BRASIL - IDADES À MATERNIDADE DAS MULHERES POR NÍVEL DE ESCOLARIDADE DAS MÃES, 1970, 1980, 1991	80
TABELA 8. BRASIL - PROBABILIDADES CONDICIONAIS DE SOBREVIVER DAS MULHERES ESTIMADAS PELO MÉTODO DE ORFANDADE POR NÍVEL DE ESCOLARIDADE, ANO DE REFERÊNCIA E ESPERANÇA DE VIA AO NASCER, PNAD E PPV, 1996.....	82

TABELA 9. BRASIL - NÚMERO DE MORTES, PESSOAS-ANO E TAXAS DE MORTALIDADE POR GRUPOS DE IDADE, PARA MULHERES DE DIVERSAS AMOSTRAS, PPV E PNAD, 1996.....	86
TABELA 10. BRASIL - COEFICIENTES DA REGRESSÃO POISSON DO NÚMERO DE MORTES DE MULHERES EM FUNÇÃO DA IDADE, PPV E PNAD 1996	87
TABELA 11. BRASIL - TAXAS DE MORTALIDADE ESPECÍFICAS DE MULHERES POR IDADE (X 1000) ESTIMADAS ATRAVÉS DO MÉTODO ESTOCÁSTICO E ESPERANÇAS DE VIDA AO NASCER E AOS 30 ANOS, PPV E PNAD, 1996	88
TABELA 12. BRASIL - NÚMERO DE MORTES, PESSOAS-ANO E TAXAS DE MORTALIDADE BRUTAS, POR GRUPOS DE IDADE E DE ESCOLARIDADE PARA MULHERES, PPV, PNAD, 1996	90
TABELA 13. BRASIL - COEFICIENTES DA REGRESSÃO POISSON DO NÚMERO DE MORTES DE MULHERES EM FUNÇÃO DA IDADE E DA ESCOLARIDADE, PPV, PNAD 1996.....	91
TABELA 14. BRASIL - TAXAS DE MORTALIDADE ESPECÍFICAS (X 1000) DE MULHERES EM FUNÇÃO DA IDADE E DA ESCOLARIDADE, PPV E PNAD, 1996.....	92
TABELA 15. BRASIL - COEFICIENTES DA REGRESSÃO POISSON DO NÚMERO DE MORTES DE MULHERES EM FUNÇÃO DA IDADE, DA ESCOLARIDADE E DA INTERAÇÃO ENTRE AS DUAS VARIÁVEIS, PPV E PNAD, 1996	94
TABELA 16. BRASIL - COEFICIENTES DA REGRESSÃO POISSON DO NÚMERO DE MORTES DE MULHERES EM FUNÇÃO DA IDADE, DA COORTE E DA ESCOLARIDADE, PPV	95
TABELA 17. BRASIL - TAXAS DE MORTALIDADE (X 1000) DE MULHERES SEGUNDO A IDADE, A COORTE E A ESCOLARIDADE, PPV	96
TABELA 18. BRASIL – ESPERANÇAS DE VIDA AO NASCER FEMININAS ESTIMADAS ATRAVÉS DO MÉTODO DE ORFANDE E DO MÉTODO	

ESTOCÁSTICO, PARA TODA A AMOSTRA E POR NÍVEL DE ESCOLARIDADE, PNAD, PPV, 1996	98
TABELA A 1. PESOS PERTENCENTES AO MÉTODO DE ORFANDADE DE BRASS (BRASS E HILL, 1973) PARA CONVERTER AS PROPORÇÕES DE ÓRFÃOS EM ESTIMATIVAS DE PROBABILIDADES CONDICIONAIS DE SOBREVIVÊNCIA DAS MÃES	121
TABELA A 2. VALORES DA FUNÇÃO $Z(M+N)$ PARA ALOCAR NO TEMPO AS ESTIMATIVAS DE MORTALIDADE ADULTA DO MÉTODO DE ORFANDADE.....	121
TABELA A 3. VARIÁVEIS PRESENTES NO QUESTIONÁRIO DA PPV 1996 UTILIZADAS NA TESE.....	122
TABELA A 4. VARIÁVEIS PRESENTES NO QUESTIONÁRIO DA PNAD 1996 UTILIZADAS NA TESE	124
TABELA A 5. DISTRIBUIÇÃO DOS VALORES FALTANTES DA VARIÁVEL DE ESCOLARIDADE IMPUTADOS EM FUNÇÃO DAS CATEGORIAS DE ESCOLARIDADE E DO TIPO DE IMPUTAÇÃO REALIZADA, PPV E PNAD 1996.....	126
TABELA A 6. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR REGRESSÃO ESTOCÁSTICA, PPV 1996.....	127
TABELA A 7. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR IMPUTAÇÃO MÚLTIPLA, PPV 1996	128
TABELA A 8. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR <i>HOTDECK</i> , PPV 1996.....	129
TABELA A 9. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA COM O ALGORITMO EM, PPV 1996.....	130

TABELA A 10. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR REGRESSÃO ESTOCÁSTICA, PNAD 1996.....	131
TABELA A 11. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR IMPUTAÇÃO MÚLTIPLA, PNAD 1996.....	132
TABELA A 12. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA POR <i>HOTDECK</i> , PNAD 1996	133
TABELA A 13. PROPORÇÕES DE ÓRFÃOS E PROBABILIDADES CONDICIONAIS, PARA A VARIÁVEL DE ESCOLARIDADE IMPUTADA COM O ALGORITMO EM, PNAD 1996	134
TABELA A 14. PESOS DE KARUP-KING PARA OBTER PROBABILIDADES DE SOBREVIVÊNCIA DE IDADES INTERMÉDIAS DE GRUPOS QÜINQÜENAS	135
TABELA A 15. TABELA DE SOBREVIVÊNCIA PADRÃO DE HIMES ET AL., (1994) PARA ESTENDER UMA TABELA DE SOBREVIVÊNCIA ATÉ OS 100 ANOS	136
TABELA A 16. BRASIL – TAXAS DE FECUNDIDADE POR PERÍODO E TAXAS DE FECUNDIDADE PONDERADAS PELA MORTALIDADE, 1933 – 1988	137
TABELA A 17. BRASIL – PROBABILIDADES DE SOBREVIVÊNCIA, 1965, 1975, 1985	138
TABELA A 18. BRASIL – ÓBITOS ACUMULADOS, 1965.....	139
TABELA A 19. BRASIL – ÓBITOS ACUMULADOS, 1975.....	140
TABELA A 20. BRASIL – ÓBITOS ACUMULADOS, 1985.....	141
TABELA A 21. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE E DA ESCOLARIDADE POR TIPO DE IMPUTAÇÃO, DA VARIÁVEL DE ESCOLARIDADE PPV 1996 (ENTREVISTADOS COM 15 ANOS E MAIS).....	142

TABELA A 22. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE E DA ESCOLARIDADE POR TIPO DE IMPUTAÇÃO DA VARIÁVEL DE ESCOLARIDADE, PNAD 1996 (ENTREVISTADOS COM 40 ANOS E MAIS).....	142
TABELA A 23. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE E DA ESCOLARIDADE POR TIPO DE IMPUTAÇÃO DA VARIÁVEL DE ESCOLARIDADE, PPV 1996 (ENTREVISTADOS COM 40 ANOS E MAIS).....	143
TABELA A 24. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE, DA ESCOLARIDADE E DA INTERAÇÃO ENTRE IDADE E ESCOLARIDADE, POR TIPO DE IMPUTAÇÃO DA VARIÁVEL DE ESCOLARIDADE, PPV 1996 (ENTREVISTADOS COM 15 ANOS E MAIS).....	144
TABELA A 25. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE, DA ESCOLARIDADE E DA INTERAÇÃO ENTRE IDADE E ESCOLARIDADE, POR TIPO DE IMPUTAÇÃO DA VARIÁVEL DE ESCOLARIDADE, PNAD 1996 (ENTREVISTADOS COM 40 ANOS E MAIS) ...	145
TABELA A 26. COEFICIENTES DA REGRESSÃO POISSON EM FUNÇÃO DA IDADE, DA ESCOLARIDADE E DA INTERAÇÃO ENTRE IDADE E ESCOLARIDADE, POR TIPO DE IMPUTAÇÃO DA VARIÁVEL DE ESCOLARIDADE, PPV 1996 (ENTREVISTADOS COM 40 ANOS E MAIS).....	146

RESUMO

Embora exista uma vasta literatura sobre o estudo dos diferenciais na mortalidade adulta por status socioeconômico nos países desenvolvidos, a produção sobre este tema é ainda escassa nos países em desenvolvimento. O Brasil não é exceção. Apesar de ser um dos países com maiores níveis de desigualdade social, e da longa tradição de pesquisa nesta área, ainda sabe-se muito pouco sobre como as disparidades socioeconômicas afetam a mortalidade. A grande dificuldade reside na obtenção de dados confiáveis para a geração de estimativas consistentes.

O objetivo principal desta tese é desenvolver uma metodologia que permita utilizar dados individuais de pesquisas domiciliares para estimar taxas de mortalidade por nível socioeconômico. Esta metodologia, aqui chamada de método estocástico é inspirada no método de orfandade de Brass, já que combina informações retrospectivas sobre a sobrevivência e sobre o status socioeconômico dos progenitores dos entrevistados pela pesquisa domiciliar. Este método estocástico se diferencia, principalmente, do método clássico de orfandade pela utilização de microdados, ao invés de dados agregados, e pela aplicação de diversas técnicas estatísticas. Desta forma é possível associar a mortalidade com outras variáveis coletadas pela pesquisa, medindo a significância estatística dos diferenciais de mortalidade na população em estudo.

Outro objetivo da tese é aplicar o método estocástico ao caso brasileiro. Para isto, são utilizadas duas bases de dados: a Pesquisa de Padrões de Vida (PPV) e a Pesquisa Nacional por Amostra de Domicílios (PNAD), ambas realizadas em 1996. Na aplicação desta nova metodologia, são estimadas medidas de mortalidade por idade e escolaridade, representando o nível socioeconômico, unicamente para as mulheres adultas (mães dos entrevistados). Além disso, foram estimadas, também, medidas de mortalidade adulta (por idade e escolaridade) através do método de orfandade a serem comparadas com os resultados do método estocástico, para dar maior confiabilidade aos resultados.

As estimativas por idade encontradas através do método de orfandade e do método estocástico foram semelhantes e concordaram com estimativas realizadas por outros estudos, indicando que os resultados são robustos. A aplicação do método estocástico revelou a existência de grandes diferenças de mortalidade por nível de escolaridade entre

as mulheres adultas no Brasil. Os resultados com base nos dados da PPV indicaram que mulheres sem escolaridade apresentam taxas de mortalidade 1,4 vezes maiores do que aquelas de mulheres com 1 a 8 anos de estudo e 2,2 vezes maiores do que as taxas das mulheres com 9 anos de estudo ou mais.

Este tipo de estudo representa uma contribuição importante para entender as conseqüências que a desigualdade social tem sobre o bem-estar da população no Brasil. Além disso, permite elaborar tabelas de sobrevivência e esperanças de vida por nível de escolaridade que podem ser de grande utilidade nos estudos sobre previdência social. Finalmente, espera-se que o desenvolvimento desta metodologia contribua para o avanço no conhecimento sobre diferenciais socioeconômicos em mortalidade adulta em outros países que enfrentam problemas de dados similares aos do Brasil.

ABSTRACT

Differentials in mortality by socioeconomic status have been widely studied in developed countries; however, there is a lack of studies of this kind among developing countries. Brazil is not an exception. Although it is a country where social and income inequality is very high and persistent over time, and with a long tradition of studies in this area, we know very little about how the income and social inequalities translate into adult mortality disparities there. The greatest difficulty is to find reliable data to generate robust estimates.

The main objective of this dissertation is to develop a methodology to estimate mortality rates by socioeconomic status using individual data from a household survey. This methodology is called here stochastic method and it is inspired in the Brass orphanhood method because it combines information on the survival and education of mothers, collected from the participants of a household survey. The main difference between the classic orphanhood method and the stochastic method is that the new methodology uses micro data, instead of aggregate data, and it applies several statistical techniques. Therefore, it is possible to link mortality estimates to other variables present at the survey, measuring the statistical significance of mortality differentials in this population.

Another purpose of this dissertation is to implement the stochastic method to the Brazilian case. The datasets used for this procedure are the Living Standards Measurement Study – PPV, and the National Household Survey (PNAD), both collected in 1996. Mortality estimates are obtained by age and educational status, only for adult women (respondents' mothers). In addition, it is used the orphanhood method to estimate adult mortality by age and education and compare them with stochastic estimates.

Mortality estimates by age from both methodologies were very similar, and agreed with other estimates, meaning that the results are robust. Estimates from the stochastic method revealed great differences of mortality by level of education among adult women in Brazil. The results based on PPV data indicate that death rates from women without education are 1,4 times greater than those from women with a to 8 years of school and 2,2 times higher than death rates from women with more than 9 years of school.

This kind of study represents an important contribution to understand the implications that social inequality has on Brazilian's population well being. Besides, based on this estimates it is possible to build life tables and obtain life expectancies by level of education that may be useful in social security and retirement studies. Finally, the development of this stochastic method to estimate mortality can be helpful to improve the knowledge about socioeconomic differentials in adult mortality in other countries facing similar data problems as Brazil.

Keywords: Adult mortality, Socioeconomic differentials in mortality, Indirect methods.

1. INTRODUÇÃO

Um grande número de estudos em todo o mundo tem demonstrado que os indivíduos de status socioeconômico mais baixo (normalmente definido através da educação, renda ou ocupação) experimentam maiores níveis de mortalidade e de morbidade do que os indivíduos com maior status socioeconômico (Preston e Taubman, 1994; Hummer *et al.*, 1998; Wilmoth e Dennis, 2006; Goldman, 2001; Cutler *et al.*, 2005). Esta associação é válida para toda a distribuição do atributo socioeconômico, mesmo dentro das classes sociais mais elevadas, definindo o que os pesquisadores chamam de gradiente social em saúde e mortalidade (Adler *et al.* 1994).

Como Krieger *et al.* (1997) argumentam, o exame da associação entre o status socioeconômico e a mortalidade adulta é de máxima importância para se entender como a desigualdade social se transforma em uma estratificação social da morte. Segundo Preston e Taubman (1994), o estudo da desigualdade social é importante em si mesmo, já que as sociedades não estão preocupadas apenas com os níveis médios de bem-estar dos indivíduos, mas também em conhecer sua distribuição na população. Além disso, a partir de estudos de diferenciais socioeconômicos em mortalidade, pode se conhecer melhor a origem e causas de certas doenças (Hummer *et al.*, 1998). Finalmente, o estudo do gradiente social em saúde e mortalidade permite identificar quais são os grupos mais vulneráveis, e definir políticas públicas de saúde mais focadas e, portanto, mais eficientes (Preston e Taubman, 1994).

Embora exista uma vasta literatura sobre o estudo dos diferenciais na mortalidade adulta por status socioeconômico nos países desenvolvidos, a produção sobre este tema é ainda escassa nos países em desenvolvimento. O Brasil não é exceção. Apesar de ser um dos países com maiores níveis de desigualdade social, e da longa tradição de pesquisa nesta área, principalmente na avaliação da desigualdade de renda (Barros, Foguel e Ulyseia 2007), ainda se sabe muito pouco sobre as disparidades socioeconômicas na mortalidade adulta. A grande dificuldade reside na obtenção de dados confiáveis para a geração de estimativas robustas. Não obstante, a análise da desigualdade social na mortalidade adulta já foi abordada a partir de metodologias indiretas, com base em estimativas de mortalidade infantil (Carvalho e Wood, 1978; Wood e Carvalho, 1988) ou se valendo de informações agregadas por regiões

ou áreas geográficas (Abreu e Rodrigues, 2000; Barros e Ramos, 2006; Barros, 1984; Cerqueira e Paes, 1998; Ishitani *et al.*, 2006; Paes, 2001; Abreu *et al.*, 2007; Bassanessi *et al.*, 2008; Drummond & Barros, 1999; Macedo *et al.*, 2001; Paes-Souza, 2002; Paim e Costa, 1986; Silva *et al.*, 1999; Szwarcwald *et al.*, 1999). Há, apenas, dois trabalhos que utilizam dados individuais para o estudo dos diferenciais socioeconômicos em mortalidade, porém, estes são limitados a pequenas áreas do país (Duncan *et al.*, 1994; Cordeiro e Silva, 2001). Todos estes trabalhos estimaram diferenciais socioeconômicos em mortalidade expressivos. Por exemplo, em Wood e Carvalho (1988), as famílias com maior renda domiciliar per capita apresentavam até 12 anos a mais de esperança de vida do que as famílias do estrato de renda mais baixo.

O objetivo principal desta tese é desenvolver uma metodologia que permita estimar medidas de mortalidade por status socioeconômico, utilizando dados individuais de pesquisas domiciliares que sejam representativos da população em estudo. Esta metodologia é inspirada no método de orfandade de Brass, já que combina informações retrospectivas sobre a sobrevivência e sobre o status socioeconômico dos progenitores dos entrevistados pela pesquisa domiciliar. A principal diferença da metodologia desenvolvida nesta Tese com o método clássico de orfandade é que a primeira define o período de exposição ao risco de morte das mães através de técnicas estocásticas, de forma que o início e o fim do período de observação são o resultado de um processo aleatório. É por isso, que a partir deste ponto, esta metodologia vai ser denominada método estocástico, diferenciando-la do método de orfandade. Além disso, o método estocástico utiliza microdados, ao invés de dados agregados. Desta forma é possível aplicar técnicas de regressão para estimar as medidas de mortalidade e associá-las com outras variáveis coletadas pela pesquisa, medindo a significância estatística dos diferenciais de mortalidade na população em estudo.

Outro objetivo da tese é aplicar o método estocástico ao caso brasileiro e posteriormente estimar taxas de mortalidade por nível de escolaridade. Para isto, são utilizadas duas pesquisas domiciliares do IBGE: a Pesquisa de Padrões de Vida (PPV) e a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizadas em 1996. A aplicação do método, nesta tese, é restrita à mortalidade feminina (mães dos entrevistados), pelas mesmas razões que são apresentadas no caso do método de orfandade, isto é, pela dificuldade em se definir o período reprodutivo dos homens e pela maior imprecisão da informação sobre paternidade. As variáveis de interesse no estudo são a idade e a escolaridade, representando o nível

socioeconômico das mulheres. A escolha da escolaridade se deve ao fato de se tratar de uma variável relativamente estável no ciclo de vida (a partir das idades adultas), e utilizada usualmente para medir status socioeconômico entre as mulheres. Além disso, é de fácil coleta e comparável ao longo do tempo, e entre populações. Apesar das inúmeras vantagens em se medir status socioeconômico a partir da educação, constatou-se, no caso dos dados da PPV e da PNAD, a presença de dados faltantes para a escolaridade das mães. A proporção de não repostas nesta variável é significativa (3,4% na PPV, e 16% na PNAD), e sua distribuição não é aleatória, já que depende em grande medida do status de sobrevivência das mães. Diante disso, a tese apresenta um estudo detalhado da distribuição de dados faltantes nas diversas amostras e do tipo de tratamento que estes dados poderiam receber. São aplicadas várias destas técnicas para o tratamento dos dados faltantes, e comparam-se os seus resultados, evitando, assim, possíveis vieses nas estimativas.

O uso de duas bases de dados na aplicação do método estocástico para o Brasil (PPV e PNAD) tem como objetivo permitir uma avaliação da robustez do novo método. Neste mesmo sentido, as estimativas derivadas de dados do método estocástico também são comparadas a estimativas de mortalidade (por idade e por escolaridade) calculadas com base no método clássico de orfandade, bem como a outras estimativas publicadas na literatura. As estimativas por idade encontradas através do método de orfandade e do método estocástico foram semelhantes e concordaram com estimativas realizadas por outros estudos, indicando que os resultados são robustos. Por exemplo, os resultados das taxas de mortalidade por idade utilizando os dados da PPV indicam que a esperança de vida das mulheres de 30 anos em 1979 estaria entre os 41,2 e os 49,7 anos, intervalo que inclui a esperança de vida de 42,2 anos estimada por Carvalho (1976) para a década de 1970/80. Além disso, a aplicação do método estocástico revelou a existência de grandes diferenças de mortalidade por nível de escolaridade entre as mulheres adultas no Brasil. Os resultados com base nos dados da PPV indicam que mulheres sem escolaridade apresentam taxas de mortalidade 1,4 vezes maiores às das mulheres com 1 a 8 anos de estudo e 2,2 vezes maiores às taxas das mulheres com 9 anos de estudo ou mais. No caso da PNAD, o diferencial encontrado é menor, porém ainda expressivo, igual a 1,5 vezes entre os grupos de menor e maior escolaridade. Apesar de que não foi possível conhecer a evolução dos diferenciais de escolaridade em função da idade, por falta de significância estatística nos modelos de regressão, os coeficientes indicam que parece existir um maior diferencial de escolaridade na mortalidade entre as idades mais jovens.

A Tese está organizada em sete capítulos, incluindo esta introdução. O segundo capítulo apresenta uma revisão da literatura sobre os diferenciais socioeconômicos em mortalidade, tanto em países desenvolvidos como em desenvolvimento, com ênfase nos trabalhos feitos para o Brasil. O terceiro capítulo revisa o método clássico de orfandade para estimar mortalidade adulta. O quarto capítulo, central na Tese, foca no desenvolvimento do método estocástico. O quinto capítulo apresenta e discute as bases de dados, dando ênfase ao tratamento dos valores faltantes na informação sobre escolaridade das mães. Os resultados da aplicação do método estocástico para o Brasil são discutidos no capítulo seis. O último capítulo conclui este trabalho.

2. A LITERATURA RECENTE SOBRE OS DIFERENCIAIS SOCIOECONÔMICOS NA MORTALIDADE ADULTA

Estudos realizados para diferentes populações e em diversos períodos indicam que a mortalidade é menor entre os subgrupos de população com maior status socioeconômico, normalmente definido por uma melhor educação, maior renda ou por melhor status ocupacional (Preston e Taubman, 1994; Hummer *et al.*, 1998; Wilmoth e Dennis, 2006; Goldman, 2001; Cutler *et al.*, 2005). A literatura sobre esta questão provém de áreas de pesquisa muito diversas, como a demografia, a economia, a sociologia, a epidemiologia, a psicologia ou a biologia evolutiva.

A primeira parte deste capítulo irá rever as principais definições e as medidas de status socioeconômico comumente utilizadas na literatura, tentando ressaltar as vantagens e as limitações de cada uma delas. A segunda parte discutirá os resultados empíricos de trabalhos feitos para países desenvolvidos e em desenvolvimento. Uma última seção é dedicada aos resultados encontrados para o Brasil e as dificuldades metodológicas que enfrentam os pesquisadores desta área neste país.

2.1. DEFINIÇÕES DE STATUS SOCIOECONÔMICO E MECANISMOS DE ASSOCIAÇÃO COM A MORTALIDADE

Antes de analisar a associação entre o status socioeconômico e a mortalidade, é importante definir o que se entende por *classe social* ou *status socioeconômico*. Entre os teóricos, não existe um consenso e utilizam-se diferentes definições. Segundo Weber (1991), a sociedade pode ser estratificada em três dimensões: a dimensão econômica, que define as pessoas por suas capacidades materiais; a dimensão do prestígio, definida pela posição onde uma pessoa é vista pelos outros; e por último, a dimensão do poder, referente à capacidade de um grupo em mobilizar recursos para o interesse próprio. As três dimensões, em conjunto, definem o status socioeconômico de cada indivíduo.. Esta é uma estratificação que resulta inevitavelmente na desigualdade social e econômica, por ser uma construção relativa, que sempre vai implicar a existência de hierarquização (Krieger *et al.*, 1997).

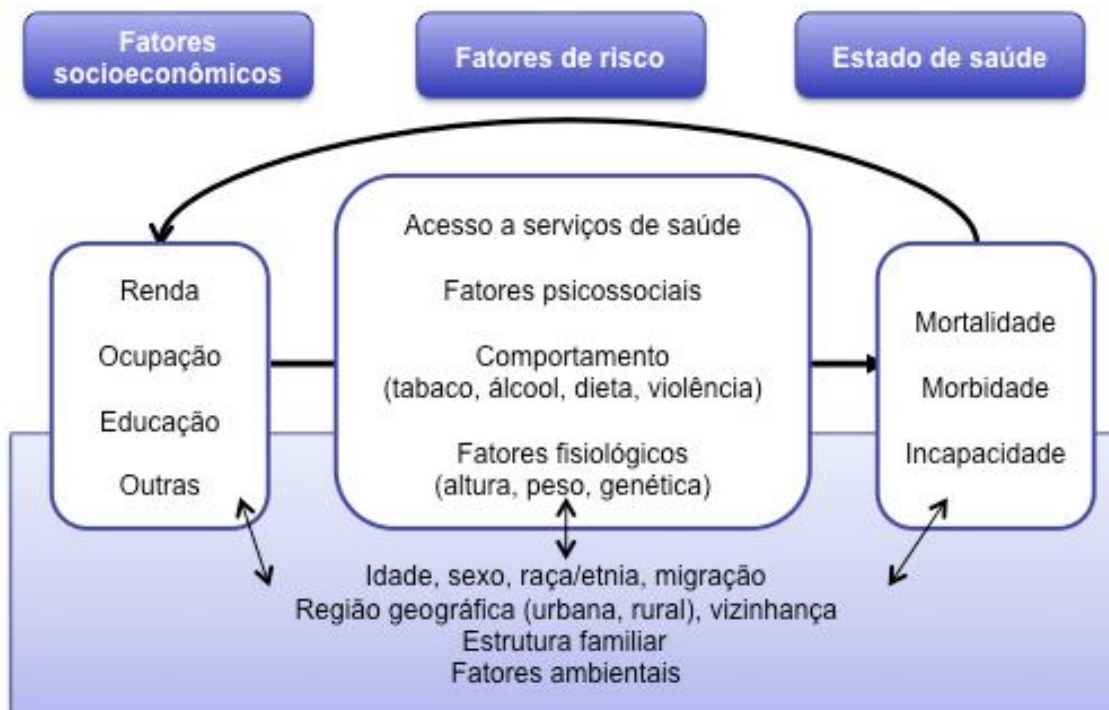
A associação entre o status socioeconômico e a mortalidade é constatada desde antigamente e, na maioria dos casos, na mesma direção: quanto melhores são as condições socioeconômicas, melhores são os indicadores de morbidade e menor é o risco de morrer (Preston e Taubman, 1994; Hummer *et al.*, 1998; Wilmoth e Dennis, 2006; Goldman, 2001; Cutler *et al.*, 2005). Recentemente, tem-se descoberto que esta relação é gradual e se aplica a todos os níveis socioeconômicos, mesmo entre os mais elevados, e para todas as idades, recebendo o nome de *gradiente na saúde* (Adler *et al.*, 1994). Além disso, estes determinantes socioeconômicos podem atuar em diversos níveis: no país como um todo, nas condições da região, da vizinhança ou comunidade, no domicílio ou no entorno familiar e, por último, no próprio indivíduo (Krieger *et al.*, 1997; Behm, 1992). Todas essas condições socioeconômicas são finalmente incorporadas biologicamente nos indivíduos, reproduzindo a desigualdade na própria saúde e na mortalidade (Krieger, 2001).

Neste trabalho, o foco principal recai sobre o estudo das características socioeconômicas individuais e como estas se correlacionam com a morbidade e os riscos de morte. Entre elas, as variáveis que melhor definem o status socioeconômico e que têm sido mais amplamente utilizadas nos trabalhos sobre mortalidade e saúde são: a educação, a renda e a posição na ocupação (Preston e Taubman, 1994). No entanto, existem outras variáveis socioeconômicas como a riqueza ou a participação no mercado de trabalho (Rogers *et al.*, 2002), além de outras variáveis sócio-demográficas, como a raça ou a origem étnica (Kington e Nickens, 2001; Hummer *et al.*, 1998; Sorlie *et al.*, 1992); o gênero (McDonough *et al.*, 1999); o estado conjugal (Hu e Goldman, 1990); o status migratório (Elo *et al.*, 2004; Palloni e Arias, 2004); o capital social (Kawachi *et al.*, 1997); ou a religião (O'Reilly e Rosato, 2008, Hummer *et al.*, 1999) que contribuem na determinação do status socioeconômico e sua correlação com a mortalidade. As diversas variáveis econômicas, sociais ou demográficas podem estar correlacionadas entre si e podem afetar conjuntamente o estado de saúde e, finalmente, a probabilidade de morrer.

Como pode ser observado no diagrama apresentado na FIG. 1, baseado nos exemplos discutidos por Blakely *et al.* (2004) e Rogers *et al.* (2002), cada uma dessas características socioeconômicas e demográficas se correlaciona com o estado de saúde, direta ou indiretamente. Essas variáveis explicam as possibilidades de acesso a bens e serviços que proporcionam bem-estar aos indivíduos, o tipo de relações sociais que se estabelecem com o

entorno, a capacidade de tomar decisões e o grau de autonomia, influenciando o comportamento individual e, finalmente, o estado de saúde.

FIGURA 1: Associações entre as condições socioeconômicas e a mortalidade e morbidade



Como está exposto no diagrama, o efeito que as condições socioeconômicas exercem sobre a saúde não é direto, sendo mediado pelos fatores de risco. Esses fatores de risco são adquiridos ao longo da vida do indivíduo, e sua relação com a mortalidade ou a morbidade muda em função da idade, do sexo, da condição domiciliar (urbana ou rural), de fatores familiares, ambientais, entre outros.

A partir das variáveis de renda, ocupação e educação, tomadas como medidas representativas do status socioeconômico, apresentam-se, a seguir, alguns dos mecanismos de associação entre essas condições socioeconômicas e o estado de saúde ou o risco de morte.

Uma das teorias mais utilizadas como referência nos estudos econômicos, para explicar o impacto da renda e da educação sobre o estado de saúde, é o modelo de produção de saúde de

Grossman (1972), inspirado no modelo de capital humano de Becker (1965). Nesse modelo, os indivíduos valorizam positivamente a própria saúde, um vez que uma melhor saúde implica uma maior produtividade do indivíduo e, portanto, maior disponibilidade de tempo para dedicar ao lazer. Assim, os indivíduos com mais renda ou maior escolaridade fazem mais investimentos em bens e serviços que melhoram sua saúde. A renda determina a quantidade de bens e serviços, e a escolaridade define sua escolha e, portanto, sua eficiência.

Portanto, a renda determina as condições básicas de sobrevivência e define as relações sociais e a capacidade de controle sobre a própria vida (Marmot, 2002). Por outro lado, uma maior escolaridade se correlaciona com menores chances de fumar, maiores chances de realizar exercício físico, ter hábitos mais saudáveis na alimentação, menor consumo de álcool, e maior disponibilidade para realizar testes preventivos de saúde (Adler *et al.*, 1994; Pamuk *et al.*, 1998; Goldman e Smith, 2005). Ou seja, a escolaridade proporciona conhecimentos e habilidades que levam a comportamentos mais positivos referentes à saúde, além de capacitar as pessoas a incorporar mais rapidamente a inovação tecnológica aplicada à saúde. O estudo de Glied e Lleras-Muney (2003) confirma este último argumento, pois mostra que os maiores diferenciais na mortalidade por grau de escolaridade se concentram naquelas doenças onde o conhecimento científico avançou mais rapidamente.

Porém, em relação ao efeito positivo da renda individual na sobrevivência do indivíduo, existem algumas críticas. Se a renda melhora a saúde porque permite ao indivíduo ter acesso a mais bens e serviços de saúde, a implementação de sistemas universais de saúde deveria reduzir os diferenciais na mortalidade por renda, o que não é respaldado pelas evidências empíricas, já que, apesar da existência deles, os diferenciais continuam aumentando (Townsend *et al.*, 1982). Além disso, uma maior renda também pode afetar negativamente a saúde, já que permite o acesso a comportamentos nocivos, como o consumo de cigarros e de bebida, ou a dietas também nocivas à saúde (Rosero-Bisby e Dow, 2009).

A própria desigualdade socioeconômica pode ter um efeito negativo sobre a saúde dos indivíduos (Wilkinson, 1997; Marmot e Wilkinson, 2001; Kawachi *et al.*, 1997). Por exemplo, em uma sociedade com uma grande desigualdade de renda, a classe social que contribui com maior quantidade de impostos e, portanto, possui um maior poder político, é a menos interessada em investir em bens sociais como educação e serviços de saúde, provocando um menor acesso a este tipo de serviços entre as camadas mais pobres e impedindo melhorar sua saúde. Hsieh e Pugh (1993) ainda acrescentam que a desigualdade

social reduz a coesão social, levando a uma maior prevalência de violência. Desde a perspectiva psicossocial, a desigualdade socioeconômica afeta o estado de saúde através da frustração do indivíduo consciente da sua própria posição inferior na hierarquia social, e da inevitável comparação com as camadas superiores. Acompanhando este argumento, uma das conclusões principais dos estudos chamados “Whitehall Studies” (Ferrie, 2004) é que os postos de trabalhos onde os trabalhadores não têm controle sobre suas atividades geram um maior stress. Segundo Cassel (1976), tanto as frustrações, como o stress afetam diretamente as funções neuroendócrinas dos indivíduos e aumentam a susceptibilidade de contrair doenças e de morrer por doenças do coração (Dressler, 1996). Os mecanismos biológicos que explicam a relação entre stress e doenças crônicas são denominados “*carga alostática*”, termo apresentado pela primeira vez por McEwen (1998; 2000). O nome se refere à resposta que o corpo humano cria para enfrentar as situações de stress e restaurar o equilíbrio metabólico. Estas respostas consistem na produção de energia por parte de uma série de hormônios, que ajudam a equilibrar o corpo no curto prazo, mas que, no longo prazo, provocam carências em outros órgãos vitais para a sobrevivência no futuro.

O stress gerado entre as pessoas situadas em uma baixa posição dentro da hierarquia social também está associado a comportamentos nocivos à saúde. Vários estudos (Ferrie, 2004; Waldron e Lye, 1989) concluem que uma posição mais baixa na hierarquia social pode estar correlacionada a um maior consumo de tabaco, a uma dieta pior e a uma menor probabilidade de realizar exercício físico. Finalmente, o tipo de ocupação também define as condições físicas do lugar de trabalho, que podem afetar diretamente o estado de saúde e aumentar o risco de morte.

Contudo, a associação entre posição socioeconômica e saúde não é de direção única, e, como já foi apresentado na FIG. 1, o estado de saúde também determina o status socioeconômico (Smith, 1999). Este fenômeno é chamado na literatura de *causalidade reversa* (Goldman, 2001). Há evidências claras de que os indivíduos com uma saúde pior têm maiores riscos de serem desempregados ou inativos e de trabalharem menos horas, recebendo menores rendimentos (Chirikos e Nestel, 1985; Case e Deaton, 2003; Smith, 1999). Renda e ocupação são variáveis mais suscetíveis aos problemas de causalidade reversa, por variarem ao longo de todo o ciclo de vida. Por outro lado, o nível de escolaridade, por ser definido em um estágio inicial da vida, não sofre os efeitos de variações na saúde que ocorrem em estágios mais avançados do ciclo de vida (Elo e Preston, 1996). De fato, educação é uma das variáveis

mais utilizadas nos trabalhos sobre diferenciais socioeconômicos na mortalidade, tanto nos países desenvolvidos (Preston e Taubman, 1994; Kunst e Mackenbach, 1994; Huisman *et al.*, 2004), como naqueles em desenvolvimento (Liang *et al.*, 2000; Zhu e Xie, 2007; Zimmer *et al.*, 2002; Son, 2001; Son, 2004; Hurt *et al.*, 2004; Rosero-Bixby e Dow, 2009). Ainda assim, vários trabalhos mostram que problemas de saúde na infância afetam negativamente o desempenho escolar, e determinam os anos de estudo que uma pessoa terá ao entrar na vida adulta (Case *et al.*, 2005; Miguel e Kremer, 2004). Além disso, segundo a teoria econômica, existem fatores que afetam tanto a educação como a saúde, interferindo no sentido da causalidade entre as duas variáveis (Groot e Van den Brick, 2007). Segundo Becker e Mulligan (1994). As pessoas que dedicam mais tempo ao estudo têm menor preferência pelo tempo presente ou, dito de outra forma, maior preocupação com o futuro. Este comportamento explica que as pessoas que estudam mais, investem mais na própria saúde e não consomem bens que oferecem prazeres imediatos, mas são negativos para a saúde (Fuchs, 1982; Groot e Van den Brick, 2007; Grossman, 2008). Assim, é importante levar em conta a existência dessa dupla relação, com o objetivo de controlar os seus efeitos e ter cuidado ao se falar em causalidade.

Apesar da existência de causalidade reversa, vários trabalhos têm demonstrado que a associação mais forte se dá no sentido status socioeconômico → saúde, e não ao contrário (Adler *et al.*, 1994; Mulatu e Schooler 2002; Smith e Kington, 1997; Goldman, 2001). No caso da escolaridade, Lleras-Muney (2005) mostrou que a melhora no atendimento à educação nos Estados Unidos, no período de 1915 a 1939, causou, de forma direta, um aumento no número de anos de esperança de vida da população. Desta forma, a autora mostra que, independentemente da existência de causalidade reversa, existe um efeito claro e positivo sobre a redução da mortalidade por parte do aumento da escolaridade.

Vale mencionar, por fim, a correlação existente entre as variáveis socioeconômicas, sendo difícil, muitas vezes, diferenciar os mecanismos causais específicos de cada uma delas em relação ao estado de saúde e ao risco de morte. A renda familiar determina os níveis de educação, e a escolaridade, por sua vez, afeta o tipo de ocupação e a renda do trabalho. Além de todos os argumentos já mencionados, é importante acrescentar que uma das maiores vantagens em se utilizar a variável educação para representar o status socioeconômico é o fato dela ser de fácil coleta para todos os adultos (Elo e Preston, 1996). Finalmente, no caso das mulheres, dos desempregados e dos aposentados, muitas vezes é difícil obter informações

sobre renda individual e sobre o tipo de ocupação que sejam representativas do seu status socioeconômico real. A educação, nesses casos, é a variável mais próxima que existe para poder medir sua posição social de forma mais confiável, apesar de não ser a ideal (McDonough *et al.*, 1999).

2.2. EVIDÊNCIAS EMPÍRICAS E QUESTÕES METODOLÓGICAS

2.2.1. Diferenciais socioeconômicos na mortalidade nos países desenvolvidos: magnitudes e tendências

Os países desenvolvidos têm a vantagem de ter, há décadas, registros de mortalidade de alta qualidade. Além disso, esses países também possuem várias pesquisas, amostrais e longitudinais, que permitem o exame em profundidade da relação entre as condições socioeconômicas e o estado de saúde da população. Isso lhes permite obter resultados muito acurados sobre a relação entre as condições socioeconômicas e a mortalidade da população. O estudo dos diferenciais de saúde e mortalidade nesses países serve, portanto, como referência para os estudos nos países em desenvolvimento, onde a qualidade de dados é inferior. Diante deste fato, é apresentada, a seguir, uma revisão dos resultados e tendências observados durante as décadas mais recentes nos países desenvolvidos.

Os estudos sobre diferenciais socioeconômicos em mortalidade nos Estados Unidos, Canadá ou Europa abrangem uma grande diversidade de variáveis, passando pelas tradicionais, como sexo, raça, estado conjugal, educação, renda ou ocupação, mas incluindo muitas outras como: lugar de nascimento, riqueza, religião ou capital social. Como esta tese foca nos diferenciais em mortalidade adulta por nível de educação, a revisão, a seguir, também irá se restringir a esta área, incluindo alguns exemplos de diferenciais por renda e tipo de ocupação, por serem variáveis intimamente correlacionadas com a educação.

A variável educação, normalmente definida pelos anos de estudo, é uma das medidas mais utilizadas nos Estados Unidos para o estudo de diferenciais socioeconômicos em mortalidade (Preston e Taubman, 1994; Hummer *et al.*, 1998; Wilmoth e Dennis, 2006). Kitagawa e Hauser (1973) foram os primeiros a estimar diferenciais de mortalidade por escolaridade, utilizando dados pareados de registros de óbitos e de uma pesquisa amostral em 1960. O trabalho dos autores indicou a existência de um grande diferencial de mortalidade por nível

de escolaridade, especialmente entre os homens brancos. Por exemplo, os autores mostraram que os homens brancos, com menos de oito anos de estudo e idade entre 25 e 64 anos, apresentavam taxas de mortalidade 60% maiores que os indivíduos de mesma idade e raça que freqüentavam a universidade. A partir desse estudo, outros trabalhos foram realizados, como os de Feldman *et al.* (1989), Duleep (1989); Pappas *et al.* (1993), Elo e Preston (1996), ou o de Rogot *et al.* (1992), entre outros. Todos os trabalhos encontraram diferenças importantes na mortalidade por nível educação. Além disso, todos observaram um aumento destes diferenciais ao longo do tempo, em comparação aos resultados de Kitagawa e Hauser (1973), pertencentes ao ano de 1960; pelo menos no caso dos homens. Segundo o trabalho de Elo e Preston (1996), considerado o mais conservador dos estudos nesta área (Wilmoth e Dennis, 2006), os homens de 25 a 65 anos com menos de 8 anos de estudo, apresentavam taxas de mortalidade duas vezes maiores do que os que tinham chegado à universidade em 1971; portanto, uma variação maior do que a apresentada por Kitagawa e Hauser (1973). No que tange às mulheres, os diferenciais de mortalidade por educação são sempre inferiores aos dos homens e parecem não aumentar significativamente com o tempo. As estimativas de Kitagawa e Hauser (1973) para o ano de 1960 indicam que as mulheres brancas com escolaridade menos de 8 anos de estudo tinham taxas de mortalidade 50% maiores que as mulheres com escolaridade universitária (Kitagawa e Hauser, 1973). Elo e Preston (1996) encontraram um diferencial de 52% no período de 1971-1984 (Elo e Preston, 1996).

Cutler *et al.* (2010) também encontraram grandes diferenças de mortalidade nos Estados Unidos entre homens e mulheres universitários e aqueles que não chegaram à universidade, entre 1971 e 1996. No estudo da evolução dos níveis de mortalidade por escolaridade durante esse período, constataram um aumento substancial dos diferenciais relativos durante o período e que as diferenças foram maiores entre os homens do que entre as mulheres, durante todos os anos.

Outros trabalhos referentes à população dos Estados Unidos estimaram diferenciais de mortalidade por escolaridade, através da esperança de vida, ao invés de utilizar taxas de mortalidade ou razões de sobrevivência (Rogot *et al.*, 1992; Crimmins e Saito, 2001; Molla *et al.*, 2004). Rogot *et al.* (1992), por exemplo, estimaram que, entre 1979 e 1985, a diferença na esperança de vida aos 25 anos, entre os mais escolarizados e os menos escolarizados, era de 6 anos para os homens brancos e de 5 anos para as mulheres brancas. Molla *et al.* (2004) apresentaram estimativas para 1998, onde a diferença encontrada na esperança de vida aos 30

anos, entre universitários e aqueles com menos de 8 anos de estudo, foi de sete anos para os homens e de cinco anos para as mulheres. Esta diferença entre as esperanças de vida dos mais escolarizados e os menos se reduziu a quatro anos quando se referia à esperança de vida aos 50 anos dos homens. Para as mulheres, as diferenças foram de cinco e quatro anos, respectivamente. Estes dois trabalhos (Rogot *et al.*, 1992; Molla *et al.*, 2004) confirmam o constatado nos trabalhos que estimaram diferenciais nas taxas de mortalidade, mostrando que os diferenciais na mortalidade por nível de escolaridade parecem ter aumentado nas décadas recentes, e são maiores entre a população mais jovem.

Apesar da escolaridade não ser uma das variáveis mais utilizadas para explorar os diferenciais socioeconômicos na mortalidade adulta na Europa, a produção de estudos em diversos países e de trabalhos comparativos têm proliferado (Doombos e Kromhout, 1990; Valkonen, 1993; Fox e Goldblatt, 1982; Desplanques, 1984; Pagnanelli, 1991; Vagero e Lundberg, 1995; Kunst e Mackenbach, 1994; Kunst *et al.*, 2004; Huisman *et al.*, 2004). Os resultados confirmaram uma evolução parecida à encontrada nos Estados Unidos, apesar de existirem diferenças. Um dos estudos comparativos sobre diferenciais de mortalidade adulta por nível de escolaridade é o de Kunst e Mackenbach (1994). Os autores utilizam estimativas de mortalidade masculina geradas em trabalhos referentes à década de 70, relativos a oito países europeus e Estados Unidos. Os maiores diferenciais de mortalidade por educação nas idades de 35 a 44 anos, foram observados nos Estados Unidos, França e Itália, enquanto a Holanda, Noruega e a Inglaterra com o País de Gales, apresentavam os menores diferenciais para esse mesmo grupo de idade. As maiores variações entre os países foram observadas nas idades avançadas (entre 55 e 64 anos). O menor diferencial de mortalidade por educação (36%) foi registrado na Suécia. Como os autores indicam, é importante destacar que esta medida reflete, em parte, a desigualdade que existe entre categorias educacionais nos diferentes países, sendo que a França e os Estados Unidos são os países com maiores desigualdades educacionais (Kunst e Mackenbach, 1994).

Em outro estudo comparativo para 11 países europeus, Huisman *et al.* (2004) também constataram a existência de grandes diferenciais de mortalidade por nível de escolaridade para as pessoas de 50 a 90 anos, no início da década de 1990. Ao contrário do estudo anterior de Kunst e Mackenbach (1994), nesse trabalho, a variável de educação é dividida em três categorias, para cada um dos países, para maior comparabilidade. O país com maior diferencial de mortalidade masculina por educação foi a Áustria, apresentando uma razão

entre as taxas de mortalidade do grupo de menor escolaridade, em relação aos outros grupos de educação, de 1,86; seguido pela Dinamarca e a Inglaterra, onde os quocientes foram respectivamente de 1,37 e 1,36. Como nos Estados Unidos, os diferenciais são menores entre as mulheres, à exceção da Noruega. No caso da França, por exemplo, a desigualdade é mínima: a razão entre o grupo de menor escolaridade e o grupo de maior escolaridade é de apenas 1,08 para as mulheres de 50 a 59 anos. Ao igual que o constatado em vários trabalhos para os Estados Unidos descritos acima, nos países europeus também se observa um aumento geral dos diferenciais relativos na mortalidade por nível educacional, nas décadas recentes, especialmente entre os homens (Kunst *et al.*, 2004). Apesar do aumento do diferencial relativo, a diferença absoluta de mortalidade entre grupos socioeconômicos diminuiu, devido, em grande parte, à diminuição geral dos níveis de mortalidade. De qualquer forma, não é possível descrever um padrão único para a evolução dos diferenciais de mortalidade ao longo do tempo, já que há processos únicos dentro de cada país.

Os resultados encontrados por escolaridade são consistentes com os resultados estimados por nível de ocupação e renda, tanto nos Estados Unidos (Duleep, 1989; Elo e Preston, 1996; Johnson *et al.*, 1999; Kitagawa e Hauser, 1973; McDonough *et al.*, 1997; Pappas *et al.*, 1993; Rogot *et al.*, 1992), como na Europa (Townsend *et al.*, 1982; Kunst *et al.*, 1998; Marmot e Shipley, 1996; Wilkinson, 1990; Huisman *et al.*, 2004; Smith *et al.*, 1990), onde os estudos dos diferenciais por tipo de ocupação se destacam. Os estudos clássicos, na literatura recente, se concentram na Grã-Bretanha, onde, começando pelo chamado “*Black Report*” (Townsend *et al.*, 1982) e passando pelos chamados “*Whitehall Studies*” (Ferrie (ed.), 2004), a variável ocupação tem sido crucial para mostrar a existência de desigualdades na mortalidade, apesar do acesso universal a serviços de saúde naquele país. Além disso, existem outros trabalhos realizados no Canadá (Etches, 2009), no Japão (Liang *et al.*, 2002, Kagamimori *et al.*, 1983), na Austrália (Najman e Congalton, 2008) ou na Nova Zelândia (Pearce *et al.*, 1983), que confirmam as mesmas tendências observadas nos diferenciais em mortalidade por educação. Os diferenciais por renda e por tipo de ocupação são expressivos em todos os países, especialmente entre os homens, mostrando-se inferiores entre as mulheres, nos casos em que foi possível gerar estimativas para elas.

Um ponto em comum em todos os resultados apresentados acima é que os diferenciais relativos de mortalidade adulta, entre os grupos socioeconômicos, estão aumentando nas últimas décadas, apesar da diminuição no nível geral de mortalidade; o que contraria as

expectativas iniciais. Esse aumento da desigualdade é mais marcante na população masculina. Cabe destacar que o aumento do diferencial é sempre relativo, já que com a diminuição dos níveis de mortalidade, muitas vezes acontece uma redução do diferencial absoluto (Hoffman, 2005b). No caso das mulheres, além dos diferenciais serem inferiores aos dos homens na maioria dos estudos e países, a tendência temporal de aumento da desigualdade não foi observada em todos os casos. Nos Estados Unidos, por exemplo, Preston e Elo (1995) estimam que, entre 1979 e 1985, a desigualdade relativa por nível educacional diminuiu levemente para as mulheres de 25 a 64 anos e manteve-se estável para as mulheres entre 65 e 74 anos. O mesmo padrão é observado no caso de outros estudos realizados com outras variáveis socioeconômicas e em outros países, como na Europa (Kunst *et al.*, 2004; Mackenbach *et al.*, 2003; Borrell *et al.*, 1997). Vale ressaltar que as mulheres são, muitas vezes, excluídas das análises, pela dificuldade em se medir seu status socioeconômico através das variáveis de renda e ocupação. O aumento dos diferenciais é uma questão que preocupa os especialistas, especialmente porque essa tendência está acontecendo apesar do acesso universal aos serviços de saúde (Smith *et al.*, 1990; Marmot e Shipley, 1996). Um outro fator que pode explicar em parte essa tendência, é a mudança do conceito de status socioeconômico medido por cada uma das variáveis. No caso de educação, por exemplo, sabe-se que no passado, indivíduos que tinham baixa escolaridade podiam apresentar status socioeconômico elevado, com maior frequência do que na atualidade (Warren e Hernandez, 2007).

Outro padrão observado na grande maioria dos trabalhos é a constatação de que os diferenciais socioeconômicos em mortalidade são menores quanto maior é a idade. Mais uma vez, esse padrão é universal no caso dos homens, mas não no das mulheres, para as quais os diferenciais se mantêm estáveis com a idade, em alguns casos (Huisman *et al.*, 2004). O padrão dos diferenciais de mortalidade masculina por idade tem sido observado desde os primeiros estudos sobre mortalidade nos Estados Unidos (Kitagawa e Hauser, 1973), e reiterado em estudos posteriores, tanto nos Estados Unidos (Elo e Preston, 1996; Lauderdale, 2001), como em outros países (Marmot e Shipley, 1996; Kunst *et al.*, 2004; Hoffman, 2005b). Existem várias explicações para a diminuição dos diferenciais na mortalidade com a idade. Em primeiro lugar, o envelhecimento pode ser um agente nivelador das desigualdades sociais, se for considerado que os determinantes biológicos se sobrepõem aos sociais nas idades mais velhas (Liang *et al.*, 2002). Neste sentido, Hoffman (2005b) observa que os diferenciais convergem, não com a idade, mas à medida que a saúde piora. Por outro lado, à

medida que as pessoas envelhecem, os principais fatores que determinam os diferenciais socioeconômicos, como, por exemplo, as condições no trabalho, deixam de existir, o que leva a uma redução da heterogeneidade das condições socioeconômicas, diminuindo, também, os diferenciais na mortalidade (Lauderdale, 2001). Além disso, podem existir argumentos específicos para cada país. No caso dos Estados Unidos, por exemplo, onde existe o Medicare, um seguro público de saúde para os indivíduos maiores de 60 anos, a redução dos diferenciais de mortalidade em idades superiores a 60 anos pode ser causada pelo aumento no acesso a serviços de saúde para os subgrupos da população que, durante a fase adulta, estavam desatendidos pelo sistema privado (Preston e Taubman, 1994).

Outra explicação para a convergência das taxas de mortalidade de diferentes grupos socioeconômicos nas idades mais avançadas seria a presença de efeitos de seleção. Os grupos de menor status socioeconômico experimentam maiores níveis de mortalidade durante todo o ciclo de vida, o que poderia favorecer a seleção de indivíduos menos suscetíveis à morte nas idades mais avançadas, em comparação aos indivíduos dos grupos de status socioeconômico mais elevado, reduzindo os diferenciais observados em idades mais jovens (Hoffman, 2005b). A seleção da mortalidade pode implicar, inclusive, em uma inversão dos diferenciais de mortalidade nas idades mais avançadas, como é o caso dos diferenciais entre brancos e negros nos Estados Unidos (Manton e Stallard, 1981; Kestenbaum, 1992) ou como é o caso dos diferenciais por educação do Japão (Liang *et al.*, 2002). Vale ressaltar, no entanto, que a seleção de mortalidade não seria o único componente que explica a diminuição dos diferenciais com a idade. Estas mudanças podem ser fruto de vários fatores combinados (Beckett 2000). Alguns autores defendem a hipótese contrária, de que o diferencial de mortalidade entre os grupos socioeconômicos aumenta com a idade, em função do acúmulo de adversidades sociais e econômicas ao longo do ciclo de vida (Ross e Wu 1996), e de que a vulnerabilidade frente à morte aumenta com a idade, provocando um maior impacto das condições sociais (House *et al.*, 1994). Apesar destes argumentos, tal tendência apenas pode-se observar nos estudos sobre morbidade.

Talvez a melhor forma de entender se a desigualdade relativa na mortalidade aumenta ou diminui com a idade, seja utilizar um enfoque de coorte (Preston e Elo, 1995; Lauderdale, 2001; Hoffman, 2005b). A desigualdade da mortalidade pode ser menor entre os indivíduos mais velhos, em relação aos mais jovens, por pertencerem a coortes mais antigas, que experimentaram, ao longo da vida, padrões distintos de mortalidade. O trabalho de

Lauderdale (2001), por exemplo, confirma que, dentro de uma mesma coorte, os diferenciais na mortalidade por educação aumentam com a idade, apesar de que, entre os indivíduos de um mesmo período, os diferenciais sejam inferiores para aquelas idades mais avançadas (pertencentes a coortes mais antigas).

2.2.2. Diferenciais socioeconômicos na mortalidade em países em desenvolvimento

Os trabalhos sobre diferenciais socioeconômicos na mortalidade adulta, em países em desenvolvimento, são mais escassos do que nos países desenvolvidos. A falta de estudos se deve à carência de dados confiáveis para correlacionar características socioeconômicas, no nível individual, com os registros de óbitos. Além disso, como em muitos desses países a mortalidade infantil ainda se mantém em níveis relativamente elevados, há um maior enfoque nos diferenciais de mortalidade nesta faixa etária da população, deixando de lado comparações nas idades adultas (Cleland, 1992; Caldwell, 1989). Segundo esta literatura, o nível socioeconômico da família e, em especial, das mães apresenta uma correlação positiva com o risco de sobreviver dos recém-nascidos e das crianças abaixo de 5 anos. Sabe-se, também, que as condições socioeconômicas na infância têm um efeito sobre o estado de saúde na idade adulta (Elo e Preston, 1992; Dolbhammer, 2003). Por isso, pode-se esperar que a presença de desigualdades de mortalidade entre grupos socioeconômicos, durante os primeiros anos de vida sejam um indicador da existência de gradientes parecidos entre os adultos (Case *et al.*, 2002).

Apesar da falta de dados, alguns estudos sobre esses diferenciais em países em desenvolvimento merecem destaque. Os resultados desses trabalhos são diversos e nem sempre são consistentes com os resultados apresentados para os países desenvolvidos. Entre os trabalhos publicados, são notáveis os estudos sobre países asiáticos, como China (Liang *et al.*, 2000; Zhu e Xie, 2007), Taiwan (Zimmer *et al.*, 2002), Índia (Subramaninan *et al.*, 2006), Coréia do Sul (Son, 2001; Son, 2004) ou Bangladesh (Hurt *et al.*, 2004). Na América Latina, destacam-se os trabalhos realizados com dados da Argentina (Rofman, 1994), Costa Rica (Rosero-Bixby e Dow, 2009) e Chile (Koch *et al.*, 2007). Por último, um único trabalho foi encontrado para a África, com dados da Etiópia (Berhane *et al.*, 2002). Na maioria desses estudos, enfatiza-se o uso da variável de educação, embora, em muitos deles, sejam utilizadas

medidas complementares, tais como riqueza domiciliar, casta - no caso da Índia- ou região de nascimento (rural/urbana).

Em geral, aqueles indivíduos com maior número de anos de estudo, em posições socioeconômicas mais altas e com melhor padrão de vida apresentaram menores riscos de mortalidade. Em alguns casos, os diferenciais de mortalidade encontrados foram maiores do que os observados para os Estados Unidos ou a Europa. Por exemplo, no Chile (Koch *et al.*, 2007), indivíduos (homens e mulheres) com menos de 8 anos de estudo apresentaram um nível de mortalidade 1,5 vezes superior daquele dos indivíduos com mais de 8 anos de estudo. No caso da Coreia do Sul (Son, 2001; Son, 2004), os homens com menor escolaridade (menos de 6 anos de estudo) apresentaram taxas de mortalidade 5 vezes superiores às dos universitários, enquanto o diferencial para as mulheres foi de 3,42 vezes. Já na Índia (Subramanian *et al.*, 2006), os adultos jovens (19-44 anos) do quintil de renda mais baixo apresentaram 3 vezes mais chances de morrer do que os do quintil mais alto. Além disso, nos estudos que apresentam estimativas por grupos de idade, como na Índia (Subramanian *et al.*, 2006) e na China (Liang *et al.*, 2000, Zhu e Xie, 2007), também foi observado um diferencial socioeconômico em mortalidade inferior entre as idades mais velhas.

No entanto, nem todos os resultados para os países em desenvolvimento concordam com os resultados encontrados para os países desenvolvidos. No que se refere às diferenças entre sexos, por exemplo, foram encontrados grandes diferenciais de mortalidade, por grupos socioeconômicos, entre as mulheres (Hurt *et al.*, 2004; Liang *et al.*, 2000), embora o diferencial masculino continue sendo maior em alguns países. Em especial, vale a pena destacar os resultados encontrados para a Costa Rica. Trata-se de um país que se caracteriza por ter uma esperança de vida comparável a de países como os Estados Unidos, apesar de apresentar uma renda per capita ainda baixa (Rosero-Bixby e Dow, 2009). Entre as particularidades da Costa Rica, não foi possível observar diferenciais de mortalidade, por grupos socioeconômicos, para os idosos. Entre as idades mais avançadas, o diferencial encontrado é contrário ao esperado, onde os indivíduos de menor status socioeconômico apresentam menores taxas de mortalidade Rosero-Bixby e Dow (2009). Um estudo anterior (Rosero-Bixby, 1996) constatava que as áreas mais desenvolvidas do país eram as que apresentavam maiores taxas de mortalidade por doenças cardiovasculares e diabetes. Este resultado pode ser devido a que a população em estudo em Rosero-Bixby e Dow (2009) era

unicamente constituída de indivíduos de 60 anos e mais, onde o efeito de seleção da mortalidade é mais intenso. Os autores argumentam que o estilo de vida (em relação à dieta e ao consumo de tabaco e álcool) dos indivíduos de Costa Rica que se encontram em posições socioeconômicas mais elevadas, é mais prejudicial para a saúde do que nos países desenvolvidos. O argumento é consistente com o observado por Smith e Goldman (2007) para o México, onde a renda dos homens está associada positivamente com a obesidade, o consumo habitual de cigarros e o consumo excessivo de álcool.

2.2.3. Diferenciais socioeconômicos na mortalidade no Brasil: estratégias metodológicas e evidências empíricas

Apesar do Brasil ser um dos países com maiores níveis de desigualdade social, e ter uma longa tradição de pesquisa nesta área, principalmente na avaliação da desigualdade de renda (Barros *et al.*, 2007), ainda sabe-se muito pouco sobre as desigualdades de mortalidade adulta, entre os diversos grupos socioeconômicos. A grande dificuldade reside na obtenção de dados confiáveis para a geração de estimativas no nível individual. O Brasil não possui estudos longitudinais pareados com o registro civil, e os dados convencionais (registro civil e censos demográficos) padecem de problemas como a ausência sistemática de informações socioeconômicas ou inconsistências na declaração dessas variáveis. Por exemplo, a escolaridade materna apresentou entre 30 e 60 % de incompletude entre os registros de óbitos de menores de um ano nas grandes regiões do Brasil em 1996/2000 (Romero e Cunha, 2006). Além disso, espera-se que o preenchimento das variáveis socioeconômicas no registro de óbito de um indivíduo adulto seja ainda de menor qualidade, já que, em muitos casos, a pessoa que faz o registro de óbito não conhece bem o indivíduo morto.

Não obstante, a análise da desigualdade da mortalidade adulta, por grupos sociais, já foi abordada a partir de alternativas metodológicas, com base em métodos indiretos ou se valendo de informações agregadas. Os métodos indiretos têm sido especialmente utilizados para estimar diferenciais socioeconômicos na mortalidade infantil (Sastry, 1997 e 2004; Fernandes, 1984; Merrick, 1985), mas, também, para gerar estimativas de esperança de vida através de níveis de mortalidade infantil e tabelas de vida padrão (Carvalho, 1974; Carvalho e Wood, 1977; Wood e Carvalho, 1988). Contudo, existem alguns escassos estudos que analisam os diferenciais na mortalidade adulta por níveis socioeconômicos através dos registros de óbitos (Duncan *et al.*, 1995; Cordeiro e Silva, 2001) e um estudo longitudinal

realizado na cidade de Pelotas (Rio Grande do Sul) que acompanha duas coortes de indivíduos nascidas em 1982 e 1993 (Victoria *et al.*, 1989, Barros *et al.*, 1996), analisando os diferenciais na mortalidade infantil e na saúde nos primeiros anos de vida. Além disso, existem várias pesquisas amostrais sobre diversos indicadores de saúde, que permitem analisar a desigualdade socioeconômica no estado de saúde dos adultos e idosos no Brasil (Dachs, 2002; Barros *et al.*, 2006; Noronha e Andrade, 2005; Lima-Costa *et al.*, 2006).

Os estudos sobre mortalidade infantil já demonstraram que há grandes diferenciais tanto por nível de educação, quanto por outras variáveis socioeconômicas das mães. Hernandez e Warren (2008), em um estudo para todo o Brasil no período de 1970 a 2000, mostraram que a mortalidade de crianças menores de 5 anos, filhos de mães posicionadas no menor índice de posição socioeconômica, era o dobro da mortalidade de crianças, filhas de mães em posições mais elevadas. De forma consistente, Sastry (2004) estimou, para o estado de São Paulo, níveis de mortalidade 1,8 vezes maiores para crianças filhas de mães analfabetas, comparativamente a crianças filhas de mães com educação secundária, em 1970 e 1980, e de até 3 vezes, em 1991.

Entre os trabalhos que utilizaram métodos indiretos para estimar diferenciais na expectativa de vida ao nascer no Brasil, por nível socioeconômico, destaca-se o estudo de Wood e Carvalho (1988). Os resultados desse trabalho mostram que os indivíduos das famílias com maior renda tinham, em 1970, até 12 anos a mais de esperança de vida ao nascer do que os das famílias de menor renda.

Uma vez que o Brasil é um país com grandes disparidades regionais, um grande conjunto de trabalhos enfatiza o estudo dos diferenciais de mortalidade adulta entre as diversas regiões do país, com base em medidas socioeconômicas agregadas para essas regiões. Uma parte desses estudos está centrada na análise da mortalidade infantil (Barros e Sawyer, 1993; Castilla, 1996; Sawyer *et al.*, 1987; Szwarcwald e Castilho, 1995; Simões e Monteiro, 1995), outros utilizaram como medida a esperança de vida ao nascer (Carvalho, 1974; Carvalho e Wood, 1977; Wood e Carvalho, 1988; Duarte *et al.*, 2002; Scarpin e Boff, 2008; Messias, 2003). Entre esses trabalhos, destaca-se a contribuição de Wood e Carvalho (1988). Os autores argumentam que grande parte dos diferenciais regionais de mortalidade se deve à desigual distribuição de renda entre as regiões no Brasil, mostrando que, se diminuísse a proporção de indivíduos com menor renda, reduzindo assim a desigualdade na distribuição da renda, a esperança de vida do país, como um todo, aumentaria. No que diz respeito aos diferenciais de

expectativa de vida na área rural e urbana, Wood e Carvalho (1988) demonstraram que a maior esperança de vida nas áreas urbanas beneficiava, unicamente, as famílias com melhores condições econômicas. Entre as famílias de baixa renda, as que moravam na área rural apresentavam maiores esperanças de vida do que aquelas que moram na área urbana. Vale notar que as estimativas dos autores se referem ao ano de 1970.

Mais recentemente, nota-se um crescimento no número de trabalhos sobre a associação de diferenciais socioeconômicos regionais com taxas de mortalidade por idade ou por causas de morte (Abreu e Rodrigues, 2000; Barros e Ramos, 2006; Barros, 1984; Cerqueira e Paes, 1998; Ishitani *et al.*, 2006; Paes, 2001). De forma geral, estes trabalhos mostraram que melhores indicadores socioeconômicos, tais como melhores níveis de alfabetização, maior grau de urbanização ou maior PIB per capita regional estão correlacionados com menores taxas de mortalidade geral e menores taxas por causas externas ou evitáveis. Além disso, as causas de morte mal definidas estão muito mais presentes nas regiões que têm indicadores socioeconômicos piores. Em relação às taxas de mortalidade cardiovasculares, cerebrovasculares e hipertensivas, Ishitani *et al.* (2006) mostraram uma associação positiva com a proporção de pessoas pobres e de baixa escolaridade. Por outro lado, tanto Cerqueira e Paes (1998), como Ishitani *et al.* (2006), concluíram que as causas de morte por doenças isquêmicas do coração tiveram uma maior representação relativa nas regiões com melhores indicadores socioeconômicos, em relação às regiões mais pobres. Os autores atribuem esta diferença ao maior consumo de gordura nas regiões mais ricas .

Com a intenção de utilizar características socioeconômicas mais próximas do nível individual, vários trabalhos utilizaram variáveis socioeconômicas no nível da vizinhança ou do bairro (Abreu *et al.*, 2007; Bassanessi *et al.*, 2008; Drummond e Barros, 1999; Macedo *et al.*, 2001; Paes-Souza, 2002; Silva *et al.*, 1999; Szwarcwald *et al.*, 1999). Entre os indicadores socioeconômicos utilizados, encontram-se o PIB per capita, a taxa de alfabetização ou a qualidade da infra-estrutura de serviços de saúde, utilizados para analisar normalmente sua associação com o nível de mortalidade e a distribuição das causas de morte, ou com algum tipo de causa concreta, como os homicídios. Da mesma forma que nos estudos com variáveis agregadas no nível de região ou estado, esses estudos mostraram que piores condições socioeconômicas estavam associadas com maiores taxas de mortalidade por violência, mas, também, por doenças cardiovasculares ou cerebrovasculares.

Outros trabalhos se valeram da coleta de dados em pequenas populações para examinar os diferenciais socioeconômicos na mortalidade infantil e adulta. Os resultados desses trabalhos corroboram as conclusões sugeridas anteriormente, utilizando metodologias indiretas ou variáveis agregadas. No caso da mortalidade infantil, o trabalho de Victora *et al.* (1989) e de Barros *et al.* (1996), no estudo de coorte realizado na cidade de Pelotas, mostra que a renda familiar apresenta uma forte associação com a mortalidade infantil, pois a mortalidade das crianças de famílias de menor renda (1 salário mínimo) foi sete vezes maior que a das famílias com mais de 10 salários mínimos.

Em relação à mortalidade adulta, somente dois trabalhos foram encontrados, onde se estimaram taxas de mortalidade por tipo de ocupação, utilizando dados dos registros de óbitos. O primeiro foi feito para a cidade de São Paulo (Duncan *et al.*, 1994) e o segundo, baseou-se em uma pesquisa de campo das famílias com óbitos, realizada na cidade de Botucatu, no estado de São Paulo (Cordeiro e Silva, 2001). Nos dois casos, apenas a mortalidade masculina foi examinada. No estudo de Botucatu, o coeficiente de mortalidade padronizado dos homens com ocupações intelectuais foi de 2,38 por 1000 habitantes, enquanto entre os operários, este índice foi de 8,40; quase 4 vezes superior. No caso do trabalho de Duncan *et al.* (1994) para a cidade de São Paulo, os resultados mostraram a presença de grandes diferenciais por tipo de ocupação, onde os intelectuais apresentaram os menores níveis de mortalidade, com uma razão padronizada de taxas de mortalidade de 0,62, em comparação aos 2,3 dos trabalhadores sem qualificação. Vale ressaltar que 34% dos registros não foram classificados pelos autores, por não serem economicamente ativos ou reportar um tipo de ocupação pouco específica. De uma maneira geral, o maior problema desse tipo de estudo é a pequena abrangência geográfica, que não permite chegar a conclusões que possam ser aplicadas ao país como um todo.

Uma outra forma de se aproximar ao estudo dos diferenciais de mortalidade entre grupos socioeconômicos no nível individual e de forma representativa para todo Brasil, é através dos trabalhos que utilizaram pesquisas sobre o estado de saúde, como as PNADs de 1998 e de 2003. Diversos trabalhos mostram que existe uma correlação entre diversas características socioeconômicas individuais e medidas do estado de saúde. Dachs (2002), utilizando dados da PNAD 1998, observa que o incremento de um ano de estudo aumenta a probabilidade de estar em boa saúde em 5,5%, e que cada decil a mais de renda per capita aumenta esta probabilidade em 8,2%. Também com a PNAD 1998, Noronha e Andrade (2005) mostraram

que a renda familiar per capita e a escolaridade estão correlacionadas positivamente com a probabilidade de estar saudável. Um outro estudo de Barros *et al.*, (2006), com dados da PNAD de 2003, constatara que as pessoas com menor escolaridade e imigrantes de outros estados, residentes em áreas urbanas ou moradores da região Sul, apresentavam maiores chances de padecer pelo menos de uma doença crônica. Em relação à evolução dos diferenciais socioeconômicos com a idade, tanto os dados da PNAD 1998 (Lima-Costa *et al.*, 2002) como os da PNAD 2003 (Lima-Costa *et al.*, 2006), mostram que a desigualdade é menor entre os idosos do que entre os adultos jovens.

A modo de conclusão sobre o Brasil, observa-se que o principal problema em relação ao estudo de diferenciais de mortalidade adulta, entre grupos socioeconômicos, é a falta de dados adequados para a sua análise. Os trabalhos sobre diferenciais na saúde mostraram que existem grandes diferenciais individuais em todas as medidas de saúde estudadas. Além disso, os poucos trabalhos que foram capazes de analisar a mortalidade em uma perspectiva individual chamam a atenção para a grande desigualdade existente na mortalidade infantil e na adulta, entre os estratos sociais; diferenciais que parecem ser maiores do que os encontrados nos países desenvolvidos. Considerando que o Brasil é um país marcado por uma profunda desigualdade de renda e de escolaridade e, ao mesmo tempo, como dito acima, um país que ainda tem grande deficiência em seus dados de mortalidade, o objetivo desta tese é contribuir para a ampliação do conhecimento nesta área.

3. MÉTODOS INDIRETOS PARA ESTIMAR A MORTALIDADE ADULTA A PARTIR DE INFORMAÇÕES RETROSPECTIVAS DE SOBREVIVÊNCIA

3.1. A ORIGEM: LOTKA, HENRY E O MÉTODO PARA ESTIMAR O NÚMERO DE PARENTES SOBREVIVENTES

O primeiro autor a desenvolver a relação entre a sobrevivência de parentes com a mortalidade por idade foi Lotka, em 1939. O autor utilizou essa relação para estimar a proporção de órfãos através de dados observados de mortalidade. A partir das relações descritas por Lotka (1939), Goodman *et al.* (1975) formalizaram a frequência das relações de parentesco, utilizando funções de fecundidade e mortalidade. Essas técnicas mostram que, conhecendo e fixando no tempo os regimes de mortalidade e de fecundidade de uma população, é possível estimar o número esperado de parentes vivos que uma pessoa de uma determinada idade teria. Por exemplo, pode-se descrever o número esperado de mães vivas com filhos de idade a de uma determinada coorte a partir de:

$$M(a) = \int_{\alpha}^{\beta} \frac{l_{a+x}}{l_x} A(x|t-a) dx \quad (1)$$

onde l_{a+x}/l_x é a probabilidade das mães que tiveram um filho no momento $t-a$ estarem vivas em t (calculada a partir de funções de mortalidade), $A(x|t-a)$ é a distribuição por idade das mulheres que tiveram filhos no momento $t-a$ (calculada a partir de funções de fecundidade e mortalidade) e α e β são os limites etários do período reprodutivo das mulheres, isto é, 15 (α) e 49 (β) anos.

Se a estrutura por idades das mães, $A(x|t-a)$, não é conhecida, e a população não é estável, é possível estimar esta estrutura a partir de outras informações demográficas. Essas informações são o número de nascimentos de mulheres ocorridos $t-a-x$ anos antes, $B(t-a-x)$, a probabilidade de sobrevivência das mulheres desde o nascimento até as idades do período reprodutivo, l_x , e as taxas de fecundidade que experimentaram essas mulheres quando foram mães (com x anos de idade), $f(x)$ (Goodman *et al.*, 1975):

$$A(x|t-a) = \frac{B(t-a-x) \cdot l_x \cdot f(x)}{\int_a^{\hat{x}} B(t-a-x) \cdot l_x \cdot f(x) \cdot dx} \quad (2)$$

Henry (1960) foi o primeiro autor a explorar a relação entre orfandade e mortalidade materna no sentido contrário, isto é, estimando a mortalidade em uma população que não apresentava registros de óbitos, mas onde havia informação sobre a sobrevivência das mães. Seguindo o mesmo raciocínio das relações de parentesco, ao conhecer a fecundidade da população e o número de parentes vivos dos indivíduos sobreviventes, é possível, também, estimar a função de mortalidade por idade destes parentes. A idéia básica por trás desses métodos é que a idade dos indivíduos de referência serve como estimativa do tempo de exposição à morte dos parentes. A partir dessa idéia, Brass (1971) e Brass e Hill (1973) formalizaram um método indireto para estimar a mortalidade adulta, também chamado de método de orfandade ou método para estimar mortalidade adulta de Brass. Como será discutido na próxima seção, o objetivo do método é estimar funções de sobrevivência a partir de informações sobre orfandade, coletadas em censos demográficos e pesquisas domiciliares.

3.2. O MÉTODO DE ORFANDADE

O método de orfandade baseia-se na informação de sobrevivência das mães dos entrevistados de uma pesquisa amostral ou um censo demográfico para estimar probabilidades de morrer na idade adulta (United Nations, 1983). O mesmo procedimento pode ser aplicado para estimar a mortalidade adulta feminina e a masculina. Não obstante, o uso do método para se estimar a mortalidade masculina é, em geral, evitado, dada a maior dificuldade em se definir a paternidade e a maior imprecisão sobre os limites (início e fim) do período reprodutivo masculino. Portanto, para uma maior conveniência, todas as referências e aplicações empíricas do método, nesta tese, não estão centradas na mortalidade feminina.

A primeira formalização desse método indireto aparece em Brass e Hill (1973) e desde então recebe o nome de “método de orfandade” ou “método para estimar mortalidade adulta de Brass”. A partir dessa publicação, outros autores continuaram desenvolvendo essa metodologia, introduzindo melhoras e adaptações (Hill e Trussell, 1977; Hill *et al.*, 1983; Chackiel e Orellana, 1985, Palloni e Heligman, 1985; Timaeus, 1991a; Timaeus, 1991b). Outros autores exploraram outras alternativas metodológicas, trabalhando, por exemplo, com

duas pesquisas amostrais no tempo, como em Zlotnik e Hill (1981) ou Timaeus (1986). A possibilidade de trabalhar com duas pesquisas amostrais separadas por poucos anos (por exemplo, cinco anos), permite melhorar o método de orfandade clássico através da criação de uma coorte sintética que é acompanhada em dois momentos do tempo. Este tipo de aplicação é especialmente adequado quando a fecundidade ou a mortalidade não são constantes, ou quando existe subregistro.

O método de orfandade clássico parte da constatação que as mães dos entrevistados estavam vivas no nascimento dos filhos. Assim, quando os filhos são entrevistados e perguntados sobre o status de sobrevivência das mães, a idade dos filhos à data da entrevista representa o período de exposição ao risco de morrer das mães. A proporção de mães sobreviventes é, então, uma aproximação da probabilidade condicional de sobrevivência da coorte de mães, da idade em que tiveram os filhos até esta idade mais a idade do próprio filho no momento da entrevista. A idade de início do período de risco das mães dependerá da distribuição de fecundidade da época. O método de orfandade formaliza, então, a relação entre a proporção de mães sobreviventes com o tempo de exposição ao risco definido pela idade do filho, estimando a probabilidade de sobrevivência através da proporção de órfãos e de uma série de pesos estimados por meio de distribuições modelos de fecundidade e mortalidade (Brass e Hill, 1973).

Seguindo o Manual X das Nações Unidas (United Nations, 1983), observa-se que o número de mães que tiveram filhos c anos atrás (idade do filho) pode ser aproximado pelo número de nascidos vivos c anos atrás, e estimado por uma combinação do número de mães em idade fértil daquele ano, $A_c(x)$, e das taxas de fecundidade, também daquele ano, $f_c(x)$, expressada por:

$$b(c) = \int_{\alpha}^{\beta} A_c(x) \cdot f_c(x) \cdot dx. \quad (3)$$

Sendo assim, a proporção de mães sobreviventes, no momento t da entrevista, de uma coorte inicial de nascidos vivos é igual ao número de mulheres que foram mães c anos antes, representadas por $b(c)$ depois de experimentarem a probabilidade de morrer até o momento da entrevista, divididas pelo total de mães que tiveram filhos c anos atrás:

$$S(x) = \frac{b(c) \cdot \frac{l_{x+c}}{l_x}}{b(c)} \quad (4)$$

onde $\frac{l_{x+c}}{l_x}$ refere-se à probabilidade de sobrevivência condicional das mães dos entrevistados entre a idade que as mães tiveram o filho (x) e esta idade mais a idade do filho (c) no momento da entrevista. Como discutido anteriormente, a idade do filho (c) representa o período de exposição das mães ao risco de morrer, que está sendo estudado. A distribuição de mulheres está compreendida entre as idades α e β , limites do período reprodutivo (15 e 49 anos). Combinando as fórmulas (3) e (4), a proporção de mães sobreviventes dos indivíduos entrevistados, com idade c no momento da realização da pesquisa, pode ser reescrita como

$$S(x) = \frac{\int_{\alpha}^{\beta} A(x) \cdot f(x) \cdot \frac{l_{x+c}}{l_x} \cdot dx}{\int_{\alpha}^{\beta} A(x) \cdot f(x) \cdot dx} \quad (5)$$

A partir da equação (5), e conhecendo a proporção de mães sobreviventes e a estrutura de fecundidade, poderia ser estimada a probabilidade de sobreviver desde a idade que a mãe teve o filho (x) até a idade $x + n$ (n sendo a idade do filho). Infelizmente, a equação (5) expressa a idade de forma contínua, impedindo o uso de dados provenientes de pesquisas domiciliares ou censos demográficos, onde, na maioria dos casos, os dados disponíveis são discretos e normalmente agrupados em intervalos de idade de cinco anos. Com o objetivo de resolver esse inconveniente e facilitar a aplicação do método, Brass e Hill (1973) propõem uma adaptação da expressão (5) fixando a idade x à idade de 25 anos, muito próxima de uma da idade à maternidade de cada população. Como as mães de uma mesma coorte de filhos os têm em idades diversas, distribuídas dentro do período reprodutivo, a idade média à maternidade, x_b , pode ser usada como uma média da idade das mulheres que tiveram seus filhos c anos atrás. Porém, a idade de 25 anos escolhida por Brass e Hill (1973), dificilmente coincide de forma exata com a idade à maternidade de cada população em estudo. Os autores estimam, então, uma série de fatores de ponderação ou pesos para ajustar a fórmula à estrutura de fecundidade de cada população. Estes fatores foram originalmente calculados a partir de simulações feitas com base em diferentes funções de fecundidade, pressupondo uma população estável com taxa de crescimento anual de 2% e funções de mortalidade da tabela de vida Logito Standard de Brass (Brass, 1971).

Portanto, a probabilidade condicional de sobrevivência das mães, desde a idade de 25 anos até a idade 25 + n anos (n=idade do filho), pode ser estimada através da proporção ponderada de mães sobreviventes pelos pesos de Brass da seguinte forma:

$$\frac{l_{25+n}}{l_{25}} = W(n) \cdot S(n-5) + (1-W(n)) \cdot S(n), \quad (6)$$

onde $S(n)$ é a proporção de respondentes entre a idade n e $n+4$ com a mãe viva na data da entrevista, e $W(n)$ ¹ é o conjunto de pesos específicos para cada população.

Hill e Trussell (1977) elaboraram uma estratégia alternativa para ajustar as proporções de órfãos à probabilidade condicional de sobrevivência das mães desde os 25 anos até os 25 + n anos. Em vez de usar fatores de ponderação, realizaram regressões a partir de múltiplos padrões de mortalidade e fecundidade, estimando até 900 proporções de mães sobreviventes diferentes. A partir das regressões estimaram-se três parâmetros (a , b e c), permitindo utilizar a proporção de mães sobreviventes de um grupo de 5 anos de idade e a idade média à maternidade (M) para ajustar a probabilidade condicional de sobrevivência das mães desde os 25 anos até os 25 mais a idade dos filhos:

$$\frac{l_{25+n}}{l_{25}} = a(n) + b(n) \cdot M + c(n) \cdot S_{n-5}. \quad (7)$$

As duas técnicas apresentadas aqui sobre o método de orfandade precisam do cálculo da idade média à maternidade para poder estimar as probabilidades condicionais de sobrevivência a partir das proporções de órfãos. Essa idade é diferente da idade média à fecundidade, já que está ponderada pela estrutura por idade das mulheres, e pode ser calculada através da seguinte equação:

$$M = \frac{\sum_{i=1}^7 a(i) \cdot B(i)}{\sum_{i=1}^7 B(i)} \quad (8)$$

¹ Os pesos podem ser encontrados na Tabela A 1 do Anexo.

onde o índice i representa o grupo etária quinquenal das mães dentro do período reprodutivo (de 15 a 49 anos); $B(i)$ é o número de filhos nascidos vivos, em cada grupo de idade i , que ocorreram nos 12 meses anteriores à entrevista; e $a(i)$ é o ponto médio no intervalo etário i . Dado que os nascimentos ocorrem durante o ano prévio à entrevista e a idade das mães é medida no momento da entrevista, a idade da mãe no momento do nascimento do filho deve ser, em média, 6 meses inferior à idade reportada pela mãe. Portanto, $a(i)$, a idade de referência do grupo de idade, deve ser meio ano inferior ao ponto médio do grupo de idade da mãe.

3.2.1. Avaliação da confiabilidade do método de orfandade

O método de orfandade, seja utilizando a fórmula apresentada em Brass e Hill (1973), seja a de Hill e Trussell (1977), permite estimar de forma simples as probabilidades de morte na idade adulta. A principal vantagem reside na facilidade de obter a informação necessária para aplicar o método, já que basta conhecermos o status de sobrevivência dos pais, a idade do entrevistado (filhos) e a idade média à maternidade.

Não obstante, o método se baseia em uma série de fortes pressupostos que, caso não se confirmem, podem adicionar viés aos resultados, devendo ser analisados com detalhe. Considerando unicamente as estimativas de mortalidade das mães, existem, principalmente, três pressupostos do método que, ao serem violados, gerariam uma diferença entre a mortalidade estimada e a mortalidade real. O primeiro pressuposto é que não existe seletividade como consequência da possível relação entre a mortalidade das mães e a dos filhos, e entre a fecundidade e a mortalidade (tanto das mães como a dos filhos). O segundo pressuposto é que os filhos adotivos respondam sobre a sobrevivência das mães naturais e não das mães adotivas. Por último, assume-se que a mortalidade é constante no tempo. Se a mortalidade estivesse caindo, por exemplo, as estimativas de mortalidade não corresponderiam ao nível de mortalidade no momento da entrevista, já que as mães dos indivíduos mais velhos teriam morrido em períodos mais distantes do passado, onde a mortalidade era mais elevada.

Passamos, então, a analisar os possíveis vieses derivados da quebra de cada um desses pressupostos. Em relação ao pressuposto de inexistência de seletividade, há duas relações a serem consideradas. Em primeiro lugar, sabe-se que tanto a mortalidade infantil quanto a

mortalidade adulta estão fortemente correlacionadas com as condições de vida. Sendo que mães e filhos expostos a piores condições de vida, experimentam também maiores taxas de mortalidade. Além disso, uma das causas importantes de mortalidade das mulheres está relacionada com complicações no parto, que também aumentam a probabilidade de morte do recém-nascido. Portanto, há uma maior chance de filhos de mães mortas estarem sub-representados na amostra de entrevistados, o que poderia produzir um erro por falta nas estimativas de mortalidade. Em segundo lugar, deve-se considerar que existe uma relação entre fecundidade e mortalidade das mães que pode provocar vieses nas estimativas de mortalidade em dois sentidos (Palloni *et al.*, 1984). Por um lado, há uma correlação positiva entre fecundidade e mortalidade das mães, o que geraria uma sobre-representação de entrevistados cujas mães estariam mortas na data da pesquisa. Por outro lado, as mulheres que vivem mais estão mais expostas aos riscos de terem filhos e esta correlação negativa entre fecundidade e mortalidade levaria a um viés no sentido contrário. Segundo Palloni *et al.* (1984), o primeiro efeito de seletividade, isto é entre a mortalidade infantil e a adulta, costuma produzir erros mais significativos. Contudo, a existência desse viés é o que permite que os outros dois vieses derivados da relação entre fecundidade e mortalidade, se mantenham pequenos.

Outro pressuposto do método é a inexistência de adoção. Se essa prática fosse usual no Brasil, a quebra do pressuposto deve afetar principalmente à resposta sobre a sobrevivência das mães dos indivíduos mais jovens, já que seguramente sejam as mães adotivas as que respondem diretamente a pergunta do censo ou pesquisas domiciliares, em vez dos filhos. Entre os filhos adultos, se espera que estes, mesmo e ter sido adotados, respondam a informação de sobrevivência da mãe natural. A substituição de mães adotivas sobreviventes pelas mães naturais que seguramente estão mortas provoca uma subestimação do nível de mortalidade adulta. Porém, esse efeito depende da cultura prevalente em cada população (Palloni *et al.*, 1984). No caso do Brasil, a adoção ainda não é uma prática muito comum e, além disso, é possível reduzir o viés gerado pela quebra deste pressuposto retirando da amostra aqueles indivíduos mais jovens, como por exemplo, os menores de 15 anos de idade (United Nations, 1983; Preston *et al.*, 2001).

O último pressuposto a ser analisado é o de mortalidade constante. Espera-se que este seja o que traga maiores problemas, já que a maioria de populações está experimentando uma persistente diminuição da mortalidade. A queda na mortalidade impede que as probabilidades

de sobrevivência, estimadas a partir do método, possam ser usadas como estimativas da mortalidade corrente. No entanto, já houve avanços no método proposto inicialmente que permitem referenciar no tempo as estimativas de mortalidade a partir da informação sobre a idade do filho. Em outro ponto deste capítulo, se discutirá como essa alocação pode ser realizada.

Por último, há de ser considerado ainda o erro de declaração na idade do entrevistado na pesquisa. Este é um erro que não decorre do método e é incorporado em todas as metodologias que estimam a mortalidade, mesmo que diretamente. Normalmente, no caso de populações da América Latina ou da Ásia, as mulheres tendem a reportar uma idade inferior à real, superestimando a mortalidade (Preston *et al.*, 2001).

O principal problema dos vieses mencionados acima é que dificilmente podem ser estimados. Palloni *et al.* (1984) fizeram uma estimativa das implicações da quebra dos pressupostos e mostraram que a maioria dos vieses acaba se anulando, à exceção do pressuposto de mortalidade constante, que pode ser solucionado alocando-se as estimativas no tempo. Apesar dos inconvenientes citados, essa técnica continua sendo amplamente utilizada para estimar a mortalidade adulta em países sem disponibilidade de registros civis de boa qualidade. Diversos autores continuam melhorando sua aplicação; e seu uso aparece em novas aplicações, como o cálculo da mortalidade adulta em países com alta proporção de mortes por HIV/AIDS (Timaeus e Nunn, 1997) ou para estimar diferenciais de mortalidade por nível de escolaridade ou categoria de ocupação (Luy, 2009).

3.2.2. Aplicação do método de orfandade para estimar mortalidade adulta para subgrupos de população

O método de orfandade tradicional produz estimativas de probabilidade de morte por grupo de idade. O mesmo método pode ser aplicado para subgrupos populacionais, desde que exista informação da proporção de órfãos por idade, referente a cada subgrupo. O exemplo mais destacado desse tipo de aplicação, realizado recentemente, é o de Luy (2009) que, com dados da Itália, aplicou o método de orfandade para estimar diferenciais de mortalidade adulta por nível de escolaridade e de ocupação. Para tal, Luy (2009) utilizou uma pesquisa amostral italiana em que os entrevistados responderam questões sobre a sobrevivência das mães e dos pais e, também, sobre a escolaridade e a ocupação dos mesmos. A existência de perguntas

sobre o status socioeconômico do pai permitiu que Luy (2009) aplicasse uma extensão do método de orfandade de Brass, por subgrupo populacional. Além disso, Luy (2009) também tentou reduzir a rigidez do método clássico de orfandade. O autor estimou novos fatores de ponderação para estimar probabilidades condicionais de sobrevivência a partir de proporções de órfãos, baseados em padrões de fecundidade e mortalidade da Itália, mais adequados que o padrão standard de Brass utilizado no método de orfandade (Brass e Hill, 1973). Para aplicar esse método, então, é preciso ter dados históricos de boa qualidade sobre padrões de fecundidade e de mortalidade por idade.

Em relação à aplicação do método para subgrupos populacionais, Luy (2009) não utiliza funções de fecundidade diferenciadas segundo esses subgrupos populacionais (por escolaridade e ocupação dos pais). O que significa que, na verdade, a inovação de Luy (2009) se reduz a estimação de novos pesos, baseados na população em estudo, para ajustar as probabilidades de sobrevivência por idade e ao uso de informações de subgrupos populacionais. Ele não introduz, porém, nenhum ajuste para tratar estes grupos de forma diferenciada.

Neste trabalho, optou-se por aplicar o método tradicional de orfandade de Brass para a geração de estimativas de mortalidade adulta por nível de escolaridade. Como já foi explicado, o método precisa, para ser aplicado, unicamente da informação de órfãos e da informação de fecundidade das mães. Graças à pergunta aos entrevistados sobre escolaridade das mães, é possível obter a proporção de órfãos segundo esta característica. No caso da fecundidade, é preciso encontrar distribuições de fecundidade das mães, diferenciadas por grupo de escolaridade. Estas distribuições podem ser obtidas no momento em que o filho é entrevistado, utilizando a estrutura de fecundidade das entrevistadas (filhas), porém, o ideal seria obter esta informação para períodos mais próximos do momento do nascimento dos entrevistados (filhos). Uma vez obtidas estas duas informações, a aplicação do método segue os mesmos passos expostos acima para obter estimativas por idade.

A aplicação do método de orfandade para subgrupos de população, diferenciados neste caso pelo seu nível de escolaridade, implica novas quebras de pressupostos. O pressuposto de inexistência de seletividade deve prevalecer dentro de cada grupo de escolaridade. Porém, está claro que existe uma relação negativa entre a mortalidade da mãe e do filho com o status socioeconômico da família. Assim, ter uma menor escolaridade, implica uma maior mortalidade, tanto da mãe, como do filho, provocando uma sub-representação na amostra e

diminuindo artificialmente o diferencial de mortalidade existente entre os grupos de escolaridade. Igualmente, existe a associação entre fecundidade e mortalidade, diferente segundo o nível de escolaridade. A relação negativa entre escolaridade e fecundidade tem como consequência que aquelas mães mais escolarizadas apresentem menor número de filhos, ou até nenhum filho. Se esta relação existe, pode levar a uma subestimação do nível de mortalidade das mães menos escolarizadas em relação às mães mais escolarizadas. Esses dois vieses tomam direções opostas que, espera-se, acabem se anulando na estimativa final.

3.2.3. Alocação no tempo das estimativas de mortalidade

É imprescindível alocar as estimativas de mortalidade no tempo, tendo em vista as constantes reduções nos níveis de mortalidade, registradas na maioria das populações (Palloni *et al.*, 1984). A forma mais simples de se estimar o ano de referência da estimativa de mortalidade seria através do ponto médio do período de exposição ao risco. Esta estimativa é relativamente eficiente quando a idade dos entrevistados é pequena. Ao ir aumentando a idade, o período de exposição ao risco de morrer das mães aumenta, e com ele o viés que implica utilizar o ponto médio deste período de risco como referência temporal da estimativa de mortalidade (Palloni *et al.*, 1984).

A alocação no tempo das estimativas de mortalidade, geradas através do método de orfandade, depende de três parâmetros: do ritmo de declínio da mortalidade, da estrutura por idade da fecundidade e da interação entre o padrão de fecundidade e o tipo de declínio da mortalidade (Palloni *et al.*, 1984). Brass e Bamgboye (1981) foram os primeiros autores a desenvolver um método para referenciar as estimativas de mortalidade, levando em conta o seu declínio na mortalidade. Segundo argumentam estes autores:

“Baseando-se em Feeney, a idéia básica é determinar, para cada medida retrospectiva de sobrevivência de coorte, um ponto equivalente no tempo, tal que a sobrevivência de período na tabela de vida desse momento ou a correspondente exposição ao risco de morrer seja a mesma” (Brass e Bamgboye, 1981, p. 2).

Para poder realizar a associação entre coorte e período, os autores utilizaram uma série de pressupostos. Assume-se que a mortalidade declina, ao longo do tempo, de forma linear na escala logística. Esse pressuposto é necessário para que a alocação temporal das probabilidades de sobrevivência seja independente da taxa de mudança da mortalidade (Brass

e Bamgboye, 1981). Vale notar que essa independência só pode ser pressuposta para o nível geral de mortalidade e não para a distribuição de mortalidade por idade. Portanto, é necessário adotar uma estrutura de mortalidade para poder alocar as probabilidades estimadas no tempo, sendo escolhida a tabela de vida do Standard Logito de Brass (Brass, 1971) como padrão.

Utilizando esses componentes, a fórmula para referenciar no tempo as probabilidades de morrer estimadas pelo método de orfandade é:

$$t(n) = \frac{1}{2} n(1 - u(n)). \quad (9)$$

A data da entrevista menos o valor de $t(n)$ indica o ano de referência para cada estimativa de mortalidade. Nesta fórmula, o $u(n)$ é estimado por:

$$u(n) = \frac{1}{3} \cdot \ln({}_{10}S_{n-5}) + Z(M + n) + 0,0037 \cdot (27 - M), \quad (10)$$

onde $Z(M+n)$ é uma função² que depende de M (idade à maternidade) e de n (idade de início do intervalo de idade) e pode ser obtida por interpolação.

Palloni *et al.* (1984) também propõem um modelo de alocação no tempo mais complexo, baseado na estimativa de uma regressão linear. Nessa regressão, o tempo de referência é modelado em função de várias idades média à maternidade, de várias razões de parturição de fecundidade, e em função de várias proporções de órfãos por idade, resultando na estimativa de seis coeficientes. Da mesma forma que em Brass e Bamgboye (1981), os autores pressupõem um declínio linear da mortalidade na escala logística. Um problema desta aplicação é que acrescenta mais complexidade ao método anterior, sem adicionar muito mais eficiência. Por essa mesma razão, no caso desta Tese, o método escolhido para alocar no tempo as probabilidades condicionais de sobrevivência estimadas pelo método de orfandade foi o proposto por Brass e Bamgboye (1981).

Uma vez alocadas no tempo as probabilidades de sobrevivência estimadas é possível gerar tabelas de vida completas para cada uma das estimativas. A técnica mais simples consiste em

² A função de $Z(M+n)$ pode ser encontrada na TAB.A 2 do Anexo.

comparar a probabilidade condicional de sobrevivência estimada com as probabilidades condicionais de diversas tabelas-modelo de sobrevivência. As tabelas-modelo estão classificadas segundo padrões (Norte, Sul, Este, Oeste, ou outros), que se diferenciam pela estrutura de mortalidade em função da idade (Coale e Demeny, 1966; Nações Unidas, 1983b). Cada padrão apresenta várias tabelas-modelo de sobrevivência caracterizadas por níveis de mortalidade diferentes, classificados pela esperança de vida.

Antes de determinar a tabela de sobrevivência que vai ser utilizada para representar o nível de mortalidade da probabilidade condicional estimada, é preciso escolher o padrão de tabelas-modelo. Uma vez escolhido, é interpolada uma tabela de sobrevivência através de duas tabelas-modelo adjacentes com uma probabilidade condicional próxima à estimada.

4. MÉTODO ESTOCÁSTICO PARA ESTIMAR MORTALIDADE ADULTA

4.1. UMA VISÃO GERAL DO MÉTODO

O desenvolvimento de um método indireto estocástico que permita estimar diferenciais de mortalidade adulta entre grupos socioeconômicos é o foco principal desta Tese. Ele surge com o objetivo de lidar com a falta de dados de boa qualidade em um país como o Brasil. O método aqui proposto se inspira nas relações de parentesco apresentadas por Goodman *et al.* (1975), já que está baseado em informações retrospectivas sobre a sobrevivência das mães. Portanto, a lógica por trás do seu desenvolvimento é similar à do método de orfandade formalizado por Brass e Hill (1973), e ao método apresentado por Luy (2009), mais recente, e onde apresenta uma extensão do método de Brass e Hill (1973) para obter estimativas de mortalidade por nível socioeconômico. Não obstante, o método aqui desenvolvido apresenta duas diferenças importantes em relação ao método clássico de orfandade.

Em primeiro lugar, o método desenvolvido aqui, que passa a ser chamado de método estocástico (para diferenciá-lo do método de orfandade), ao invés de utilizar informações agregadas por grupo de idade, como no método clássico de orfandade, trabalha com microdados de pesquisas amostrais. Assim, para cada indivíduo entrevistado, é possível acompanhar longitudinalmente a experiência de mortalidade da mãe. Para poder realizar esse acompanhamento longitudinal, o método estocástico estima o período de exposição ao risco de morrer da mãe do entrevistado, através de três parâmetros: 1) a idade com que a mãe do entrevistado teve seu filho, que equivale à idade de início do período de observação ao risco de morte; 2) a idade de morte da mãe ou a sua idade na data de entrevista do filho (se ela for sobrevivente), que equivale ao fim do período de observação do risco de morte; 3) o período de exposição ao risco de morrer, obtido a partir dos dois parâmetros anteriores (idade de início e fim).

Em segundo lugar, o método estocástico está baseado no uso de diversas técnicas estatísticas. O uso destas técnicas permite gerar aleatoriamente as idades à maternidade e à morte para poder obter os períodos de exposição ao risco de morrer das mães dos entrevistados de forma individual. É por esse motivo que o método é estocástico. Sorteio de idades que é realizado

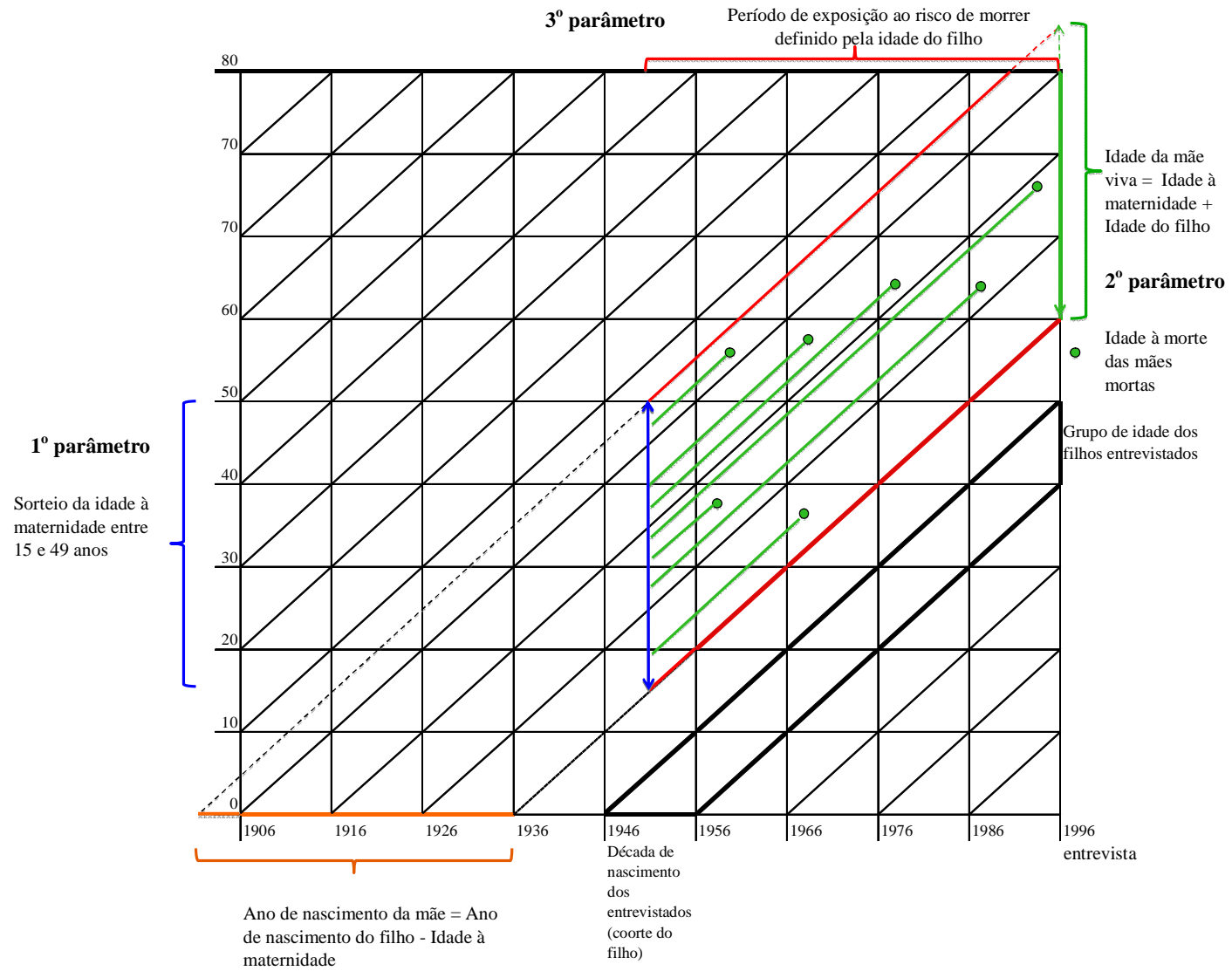
condicionado a funções de probabilidades, criadas a partir de funções reais de fecundidade e mortalidade no Brasil.

Uma vez estimado o período de exposição ao risco de morrer de cada mãe (morta e viva), a partir do método estocástico, podem ser aplicados modelos de regressão para estimar as medidas de mortalidade. A possibilidade de poder aplicar modelos de regressão aos dados obtidos dá uma maior eficiência aos resultados, além de permitir o uso de testes que medem a sua significância.

O diagrama de Lexis da FIG. 2 ilustra a concepção do método estocástico. O eixo horizontal do diagrama representa o tempo de calendário ou os períodos, enquanto as idades se posicionam no eixo vertical. As linhas diagonais representam a evolução das coortes de nascimento. O exemplo hipotético representado no diagrama da FIG. 2 refere-se a um grupo de entrevistados com idade entre 30 e 40 anos, nascidos entre 1946 e 1956 (assumindo que a data de sua entrevista foi em 1996).

Como mencionado acima, o método estocástico parte da estimação de três parâmetros. O primeiro deles consiste em estimar o ponto inicial do período de exposição ao risco de morrer por parte das mães dos entrevistados. Este ponto inicial refere-se à idade em que as mães tiveram seu filho (entrevistado). Para obter essa estimativa parte-se do momento em que o entrevistado nasceu, que, no exemplo apresentado, é a década de 1946-1956. A partir da distribuição de fecundidade dessa década (1946-1956) sorteia-se, então, uma idade à maternidade entre 15 e 49 anos, representada pelo intervalo em azul no diagrama. O seguinte passo consiste em estimar o segundo parâmetro, momento final do período de risco de morrer para cada mãe. Se a mãe está viva no momento da entrevista, a idade dela é obtida somando a idade do filho e a idade estimada à maternidade, como mostrado no intervalo verde. Caso esteja morta, é realizado um sorteio da idade à morte, que pode cair em algum ponto entre o momento do nascimento do entrevistado e o momento da entrevista, representada pelos diversos pontos verdes. O último parâmetro é o período de exposição ao risco de morrer, delimitado na FIG.2 pelas linhas vermelhas. Caso a mãe tenha falecido, este período é delimitado pela idade à maternidade e a idade à morte. Se a mãe está viva no momento da entrevista, esse parâmetro é definido pela idade do filho entrevistado.

FIGURA 2. Diagrama de Lexis representando a estimação dos parâmetros do método estocástico.



4.2. ESTIMAR OS PARÂMETROS DO MÉTODO ESTOCÁSTICO

4.2.1. Idade de início do período de exposição

O ponto de início do período de exposição para todas as mães é o momento de nascimento do filho entrevistado, ou seja, a idade à maternidade. Para as mães que continuam vivas e residem no mesmo domicílio dos entrevistados, é possível conhecer essa idade diretamente a partir da pesquisa. Unicamente é preciso subtrair da idade da mãe (igualmente entrevistada), a idade do filho³. Para as mães que não moram com o filho entrevistado ou que estão mortas na data da entrevista dos filhos, é preciso estimar sua idade à maternidade. O procedimento que se propõe, através do método estocástico, é obter essa informação sorteando aleatoriamente uma idade à maternidade, condicionada a uma distribuição de probabilidade, originada de uma função de taxas específicas de fecundidade por idade, vigentes na década de nascimento do filho. Sortear de forma condicionada a idade à maternidade responde ao fato que as mulheres não têm a mesma probabilidade de ser mães em todas as idades do período reprodutivo (entre 15 e 49 anos). A diferença de probabilidade entre as idades deve depender da estrutura de fecundidade que prevalecia durante o momento do nascimento do filho. Como discutido por Preston (2009)⁴, espera-se que a idade à maternidade varie, também, em função do status de sobrevivência das mães. Ou seja, tendo dois indivíduos com a mesma idade, um cuja mãe está morta e outro cuja mãe está viva, é de se esperar que a mãe morta tenha tido seu filho em uma idade superior à da mãe sobrevivente. Com o objetivo de ajustar a idade à maternidade segundo esta condição, as distribuições de fecundidade das mães mortas foram envelhecidas. Ou seja, para cada função de fecundidade, manteve-se constante o nível da função e alterou-se sua estrutura, através da ponderação das taxas específicas de fecundidade por idade, pelo número de óbitos nos

³ Se a mãe reside com mais de um filho, cada filho vai reportar uma idade à maternidade diferente para a mesma mãe. A sobre-representação de mães com maior número de filhos é um problema que já é considerado no método e uma tentativa de correção é explicada na sua aplicação.

⁴ PRESTON, S. H. Mortality in Brazil. Mensagem recebida por <cmturra@gmail.com> em 23 de Outubro de 2008.

respectivos grupos de idade. Desta forma, o peso das taxas de fecundidade em idades maiores é aumentado em comparação à estrutura original.

Para converter uma distribuição de fecundidade em uma distribuição de probabilidades é preciso realizar uma série de procedimentos. O primeiro passo consiste em converter as taxas específicas de fecundidade em taxas de fecundidade acumuladas. Desta forma, cada taxa de fecundidade acumulada de um grupo de idade vai representar a soma de todas as taxas específicas de fecundidade dos grupos de idade anteriores. Os valores das taxas acumuladas colocados de forma ordenada e crescente são utilizados para delimitar o início e o fim dos intervalos, cujos tamanhos equivalem aos valores das taxas específicas de fecundidade. Em segundo lugar, a distribuição de intervalos pode ser transformada em uma distribuição de valores entre 0 e 1, de forma que o peso relativo de cada intervalo continue idêntico. Assim, é possível sortear um número entre 0 e 1 que se posicionará dentro de algum dos intervalos representando uma das taxas de fecundidade associada a um grupo de idade. Quanto maior seja a taxa de fecundidade, maior será o intervalo e maior a probabilidade de que o número sorteado esteja dentro do intervalo. O grupo de idade associado ao intervalo (taxa) sorteado será a idade à maternidade sorteada. Como as funções de fecundidade são estimadas normalmente em grupos de idade quinquenais, utiliza-se como idade representativa de cada grupo quinquenal, seu ponto médio. Assim, a menor idade possível a ser sorteada é 17,5 anos, e a maior idade é igual a 47,5 anos.

4.2.2. Idade final ao período de exposição

A idade final no período de observação do risco de morte é determinada de diferentes formas, dependendo do status de sobrevivência das mães. No caso das mães que sobreviveram até a data da entrevista dos filhos, esta idade corresponde à sua idade na data da entrevista. Essa idade é fácil de ser determinada para as mães sobreviventes que residem no mesmo domicílio que os filhos, já que elas também foram entrevistadas. Para as mães sobreviventes que não residem no mesmo domicílio que os filhos, sua idade é o resultado da soma da idade do filho com a idade à maternidade estimada no primeiro passo do método.

Caso a mãe não tenha sobrevivido até a data da entrevista, é preciso estimar sua idade na data da morte. Esta estimativa é gerada a partir de um sorteio aleatório similar ao da idade à

maternidade. O sorteio é condicionado a uma distribuição de probabilidades relacionada a uma função de mortalidade vigente na população durante o período de observação. Para converter uma função de mortalidade em uma distribuição de probabilidades foi utilizada a distribuição acumulada de óbitos da tabela de vida (d_x). Da mesma forma que realizado com o sorteio da idade à maternidade, a distribuição acumulada de óbitos por idade permite utilizar estes valores como limites de intervalos, onde aqueles grupos de idade com maior mortalidade serão os que apresentem um intervalo maior, por ter um número de óbitos maior. A distribuição relativa dos intervalos é ajustada a uma distribuição de valores entre 0 e 1. Desta forma, sorteia-se um valor entre 0 e 1 que terá maior probabilidade de cair em aqueles intervalos onde o número de óbitos é superior.

A função de mortalidade utilizada para cada indivíduo varia segundo o período onde se situa o seu tempo de exposição ao risco de morte. Esta função deve ser próxima ao ponto médio desse período de exposição ao risco de morte. O período de risco, antes de conhecer a idade de morte, começa na data de nascimento do filho e termina na data da entrevista, ou no ano em que a mãe teria 100 anos, caso ela tivesse mais de 100 anos até o momento da entrevista.

4.2.3. Período de exposição ao risco de morte

O período de exposição ao risco de morte refere-se ao tempo observado entre a idade à maternidade e a idade da mãe à data de morte ou à data da entrevista (para as mães sobreviventes). Portanto, no caso das mães sobreviventes, o período de observação, é igual à idade dos filhos entrevistados. Como já foi observado na introdução do capítulo, um dos pontos importantes do método estocástico é que o tempo de exposição ao risco de morte estimado pode ser referenciado tanto em função da idade, quanto da coorte de nascimento das mães, e do período em que acontecem os óbitos. Ademais, o tempo de exposição ao risco de morte é estimado para cada um dos indivíduos da população, permitindo obter, posteriormente, estimativas de mortalidade por qualquer característica individual que esteja presente na pesquisa associada de forma única ao indivíduo.

4.3. ANÁLISE DE SOBREVIVÊNCIA COM MODELOS DE REGRESSÃO

A aplicação do método estocástico finaliza com a obtenção dos períodos de exposição ao risco de morte para cada uma das mães na amostra, resultando na criação de banco de dados longitudinal. É possível aplicar a este banco de dados, técnicas de análise de sobrevivência que modelem o número de óbitos em função do tempo de exposição e de quaisquer outras características que forem de interesse e estiverem disponíveis na base de dados, podendo estimar taxas específicas de mortalidade por idade em função dessas características.

A taxa de mortalidade entre a idade x e a idade $x+n$, de uma determinada coorte, denominada aqui por ${}_nM_x$, é calculada pela divisão do número de mortes na idade x (${}_nd_x$) e o tempo de exposição ao risco de morrer, definido pelo número de pessoas-ano (${}_nL_x$). Isto é:

$${}_nM_x = \frac{{}_nd_x}{{}_nL_x}. \quad (11)$$

O número de óbitos é obtido a partir da distribuição do número de mães mortas por idade da mãe e demais características de interesse. Já o número de pessoas-ano representa o número de anos vividos por todas as mães desde o momento que tiveram o filho até o momento da sua morte ou até a data da entrevista, distribuídos entre os grupos de idade e as demais variáveis a serem consideradas.

Para o cálculo do número de pessoas-ano, utiliza-se a função *split* do programa STATA 10.0. Essa função transforma cada indivíduo em um número de pseudo-indivíduos igual ao número de anos que estiveram expostos ao risco de morrer. Por exemplo, se uma mãe esteve exposta ao risco de morte desde os 20 até os 33 anos de idade (idade em que morreu), a função *split* cria 14 pseudo-indivíduos para representar a trajetória desta mãe entre estas idades, de tal forma que é possível conhecer a contribuição exata de cada indivíduo para o tempo total de exposição.

Uma vez calculado o número de pessoas-ano é possível modelar o número de mortes das mães em função do tempo de exposição ao risco de morte. Entre os modelos que existem para trabalhar com dados contáveis, encontram-se a regressão Poisson e a regressão binomial negativa. A diferença entre os dois modelos é que no caso da regressão Poisson, assume-se que a

variância da variável contável é igual a sua média. Já na binomial negativa, é pressuposto que a variável dependente pode ter sobredispersão, e que, portanto, a variância não se equipararia à média (Rodriguez, 2001; Long e Freese, 2003)⁵.

A fórmula da regressão Poisson é definida por uma regressão log-linear que pode ser descrita como:

$$\log(\mu_i) = \beta_i \cdot X_i, \quad (12)$$

onde X representa um vetor de variáveis e os coeficientes β são estimados por máxima verossimilhança.

A regressão binomial negativa consiste em adicionar ao modelo um efeito aleatório que represente a heterogeneidade que não está sendo controlada na regressão Poisson, segundo a seguinte forma:

$$\log(\mu_i) = \beta_i \cdot X_i + \varepsilon_i \quad (13)$$

Para poder decidir qual dos dois modelos é superior, é preciso observar se existe sobredispersão na variável resposta. Dentro da distribuição da regressão binomial negativa existe um parâmetro (chamado de parâmetro alfa em Long e Freese (2003)) que mede o nível de sobredispersão. O programa STATA 10.0 realiza um teste de máxima verossimilhança sobre a hipótese de que o parâmetro alfa seja igual a zero, que implicaria a não existência de sobredispersão. Se o teste não é estatisticamente significativo, significa que a regressão Poisson é superior à regressão binomial negativa⁶.

⁵ Além da regressão binomial, há um outro modelo que controla a sobredispersão da variável contável chamado “*zero inflated poisson*” (zip).

⁶ Uma das causas mais comuns que provoca sobredispersão na variável que parte dos zeros da amostra na variável resposta não sejam reais, ou seja, que existam indivíduos que respondem zero porque não estão expostos ao risco ou porque existe uma variável latente não controlada no modelo. Sendo a variável dependente neste estudo, o número de mortes, assume-se que todos os membros da amostra estejam expostos ao risco de morte. Por isso, espera-se que não exista nenhuma variável latente, e que o teste estatístico não demonstre a existência de sobredispersão (Böhning, 1997).

No caso das duas regressões (Poisson e binomial negativa), para que seja possível modelar o número de óbitos em função da exposição ao risco, permitindo estimar taxas de mortalidade, o tempo de exposição deve ser incorporado à equação (13) como um termo independente do vetor de variáveis,

$$\log E(mortes) = \beta_1 \cdot X_i + \log(exposição) \quad (14)$$

Para obter as taxas de mortalidade em função das características das mães presentes na base de dados é preciso incluir estas variáveis no modelo de regressão, no lugar do vetor X_i da equação (14).

As duas variáveis básicas que vão ser analisadas nesta tese são a idade e a escolaridade. Porém, as variáveis incluídas no modelo de regressão poderiam ser quaisquer que estejam presentes na base de dados relacionadas com a mãe do entrevistado. O primeiro modelo a ser estudado incorpora unicamente a variável de idade, permitindo estimar taxas específicas de mortalidade por idade. Este modelo seria:

$$\log E(mortes) = \beta_1 + \sum_{i=1}^{j-1} \beta_{2i} \cdot idade_i + \log(exposição), \quad (15)$$

onde i representa um dos j grupos de idade em que se divide esta variável, sendo que o grupo de referência será omitido da regressão.

O outro modelo a ser testado incorpora a idade e a escolaridade. Este modelo permite analisar como a mortalidade muda em função da escolaridade, uma vez controlando pelo efeito de idade, sempre presente. O modelo de regressão ficaria igual a:

$$\log E(mortes) = \beta_1 + \sum_{b=1}^{B-1} \beta_{2b} \cdot idade_b + \sum_{c=1}^{C-1} \beta_{3c} \cdot escolaridade_c + \log(exposição). \quad (16)$$

Por último, o modelo de regressão permite introduzir a interação entre as variáveis idade e escolaridade. Este tipo de análise é utilizado para separar o efeito conjunto da escolaridade e da idade, do efeito puro destas variáveis sobre a mortalidade das mães. Além disso, a interação entre variáveis dentro de um modelo de regressão permite observar como o comportamento de uma

variável muda em função da outra. Por exemplo, neste caso, podem ser analisadas taxas de mortalidade em função da escolaridade dentro de cada grupo de idade, permitindo comparar se este comportamento varia de um grupo de idade a outro. O modelo de regressão Poisson com a interação seria:

$$\log E(\text{mortes}) = \beta_1 + \sum_{b=1}^{B-1} \beta_{2b} \cdot \text{idade}_b + \sum_{c=1}^{C-1} \beta_{3c} \cdot \text{escolaridade}_c + \sum_{d=1}^{D-(B+C-1)} \beta_{3d} \cdot (\text{idade}_b \times \text{escolaridade}_c) + \log(\text{exp})$$

(17)

onde B é o número de categorias da variável idade, C é o número de categorias da variável escolaridade, e D é o número de interações possíveis entre as categorias das duas variáveis.

Em todos os modelos de regressão Poisson, os coeficientes unicamente explicam o valor diferencial em relação à categoria de referência. Porém, estes coeficientes de regressão estão controlados pelo tempo de exposição ao risco de morte, e por tanto, se referem também a taxas de mortalidade. A transformação é efetuada através do inverso do logaritmo dos coeficientes, ou seja, através do seu exponencial. Tomando como exemplo o modelo (15), onde a única variável analisada são os grupos de idade, suponha-se que esteja dividida em $j=4$ categorias, onde o grupo 1 é o de referência (omitido no modelo). Para calcular a taxa de mortalidade do grupo de idade 2, por exemplo, é necessário calcular o exponencial da soma dos coeficientes da constante (β_1) e do coeficiente do grupo de idade número 2 (β_{22}).

4.3.1. Coeficientes *bootstrap* das regressões Poisson

Um dos problemas que derivam do uso de números aleatórios é a falta de confiabilidade nos resultados. Para incrementar esta confiabilidade, uma das estratégias é calcular estimadores *bootstrap* dos coeficientes das regressões Poisson.

Para explicar o que é um estimador *bootstrap*, consideremos uma amostra S de observações independentes, onde $\hat{\theta}$ é um estimador consistente do parâmetro θ . A partir da amostra original S, é possível criar uma subamostra S_b obtida através de amostragem aleatória com substituição, onde b refere-se à ordem da amostra aleatória. Nessa sub-amostra S(b), pode-se obter igualmente

o estimador $\hat{\theta}^b$ do parâmetro θ^b , aplicando o mesmo método de estimação aplicado na amostra original S. Assim, $(\hat{\theta}^1, \dots, \hat{\theta}^B)$ seria o conjunto de estimadores obtidos por repetição do mesmo método B vezes, em B subamostras. Uma vez tendo todos os B estimadores, o estimador *bootstrap* de θ seria a média de todos os estimadores (Little e Rubin, 2002):

$$\hat{\theta}_{boot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^b \quad (18)$$

e a variância estimada *bootstrap* de $\hat{\theta}_{boot}$ é:

$$\hat{V}_{boot} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^b - \hat{\theta}_{boot})^2 \quad (19)$$

Vários estudos mostram que sob certas condições o estimador *bootstrap* $\hat{\theta}_{boot}$ é menos enviesado que o estimador original e que em condições gerais \hat{V}_{boot} é um estimador consistente da variância de $\hat{\theta}$, quando B tende ao infinito (Little e Rubin, 2002; Davison e Hinkley, 1997).

Diante desses fatos, é preciso realizar múltiplas amostras dos períodos de exposição ao risco de morrer das mães, podendo gerar múltiplas regressões Poisson. Dessa forma, em vez de apresentar os estimadores dos parâmetros de uma única regressão, serão mostrados como resultados finais, estimadores *bootstrap* resultantes de 500 regressões Poisson.

4.3.2. Alocação no tempo das estimativas de mortalidade

Para referenciar no tempo as estimativas de mortalidade, é preciso adicionar uma outra variável temporal ao modelo de regressão, seja a coorte de nascimento das mães ou o período em que elas morreram, além da idade. Sabendo a idade do entrevistado na data da entrevista, a idade à maternidade e a idade de morte, conhecer o ano em que a mãe do entrevistado nasceu (coorte de nascimento) e o ano em que ela morreu (período de morte) é uma fácil operação.

Dado que a idade é definida pelo período e a coorte como:

$$\text{Idade} = \text{Período} + \text{Coorte}, \quad (20)$$

incluir as três variáveis temporais no modelo provocaria uma colinearidade não permitida estatisticamente (Caselli et al., 1989). Por esse motivo, unicamente duas dessas variáveis temporais podem ser incluídas no modelo. Nesta Tese, as variáveis temporais utilizadas serão a idade e a coorte de nascimento. Assim, além dos modelos de regressão por idade e por escolaridade, serão apresentados modelos controlando por idade e por coorte, permitindo referenciar no tempo as estimativas.

4.4. VANTAGENS E LIMITAÇÕES DO MÉTODO ESTOCÁSTICO

Apesar do método estocástico estar baseado em pressupostos similares ao do método de orfandade de Brass, ele apresenta várias vantagens. A primeira vantagem é a maior flexibilidade ao estimar o período de risco de morte das mães. Para estimar a idade inicial desse período, o método clássico de orfandade utiliza a mesma idade média à maternidade para todas as mães. No método estocástico, esta idade muda em função do momento de nascimento do filho, já que o sorteio é realizado em função do padrão de fecundidade da década de nascimento do filho. Este fato permite que as estimativas sejam mais confiáveis e próximas da realidade da população em estudo. Além disso, o sorteio da idade permite dar maior variabilidade aos dados. No método estocástico também são ajustados os padrões de fecundidade para aquelas mães que já estavam mortas, tentando controlar pelo viés que implica pressupor a mesma fecundidade para todas as mães, independentemente do seu status de sobrevivência (Preston, 2008). Este ajuste não é realizado no método de orfandade e nem é mencionado nos pressupostos.

Em relação à estimativa de idade à morte, os dois métodos utilizam diversos padrões de mortalidade. Porém, no método de orfandade estes padrões de mortalidade e fecundidade provêm de tabelas-modelo, enquanto no método estocástico, a distribuição de mortalidade utilizada para sortear a idade de morte é representativa do padrão da população em estudo.

A segunda vantagem do método estocástico é a possibilidade de aplicar técnicas estatísticas para a sua análise. O uso de modelos de análise de sobrevivência como a regressão Poisson, por exemplo, permite obter estimativas que podem ser testadas estatisticamente para conhecer sua significância e seus intervalos de confiança. Esta flexibilidade permite, por exemplo, explorar

quais são as categorias que dão mais significância ao modelo, ou seja, que têm maior poder de explicação.

Além disso, o fato de que a base de dados seja longitudinal é essencial para poder controlar as estimativas de mortalidade em função da idade e de outra variável temporal, seja a coorte de nascimento ou o período. Desta forma, as estimativas podem ser alocadas no tempo com mais exatidão, fugindo do pressuposto de mortalidade constante do método clássico de orfandade.

Apesar das vantagens apresentadas até aqui, o método estocástico continua baseado em muitos dos pressupostos do método de orfandade, já que utiliza a mesma pergunta sobre sobrevivência das mães, respondida pelos filhos entrevistados. Os pressupostos de inexistência de seletividade e de adoção continuam prevalecendo, implicando nos mesmos problemas derivados da quebra destes pressupostos (ponto 3.2.1. desta Tese). Assim, como explicado no método de orfandade, os mesmos vieses podem se revelar com o método estocástico, mas, da mesma forma, espera-se que todos os vieses acabem se anulando. Por outro lado, o pressuposto de adoção pode ser parcialmente corrigido com o método estocástico de forma mais eficiente do que com o método de orfandade. Por estar trabalhando com microdados, é possível conhecer a idade das mães sobreviventes, e que diferença de idade elas têm com os seus filhos. Caso esta diferença seja estranhamente grande, esses casos podem ser eliminados da amostra. Além disso, como no método de orfandade, a amostra pode ser reduzida a aqueles indivíduos com mais de 15 anos de idade. Este tipo de controle também permite reduzir os possíveis vieses derivados da sobrerepresentação de mães com filhos que residem no mesmo domicílio e que multiplicam as informações sobre a mesma mãe.

5. BASES DE DADOS

Para a aplicação do método de orfandade de Brass e do método estocástico para estimar a mortalidade adulta feminina por nível de escolaridade, é preciso conhecer o status de sobrevivência e o status socioeconômico das mães. Embora a informação sobre mãe viva seja comumente coletada no Brasil, tanto em censos demográficos quanto em pesquisas domiciliares, a informação sobre o status socioeconômico das mães é rara. Recentemente, o IBGE incluiu esta informação em duas de suas pesquisas domiciliares: a Pesquisa sobre Padrões de Vida e a Pesquisa Nacional por Amostra de Domicílios, ambas de 1996.

5.1. PESQUISA SOBRE PADRÕES DE VIDA

A Pesquisa sobre Padrões de Vida (PPV) foi originalmente concebida dentro do projeto conhecido como *Living Standards Measurement Study (LSMS)*, realizado pelo Banco Mundial desde 1980. O projeto LSMS tem como objetivo principal melhorar a qualidade e a quantidade de informações disponíveis sobre o bem-estar dos membros de domicílios situados nos países em desenvolvimento. No caso do Brasil, a PPV foi realizada no ano de 1996, nas regiões Nordeste e Sudeste, que respondem por 70% da população do Brasil (Turra, 2000). A pesquisa utilizou um questionário extenso e coletou diversas informações sobre os domicílios, incluindo dados sobre as condições de moradia, informações demográficas (migração, fecundidade etc.) e econômicas (renda, ocupação, transferências interdomiliares, consumo etc.), além de pesquisar a história de educação e dados sobre a saúde de cada um dos membros dos domicílios.

Foram entrevistados inicialmente 4.940 domicílios, o que corresponde a 19.409 indivíduos. Neste trabalho, foram excluídos 91 entrevistados que nasceram fora do Brasil. Esta exclusão se justifica pelo fato de que as mães provavelmente se educaram em outros países e, portanto, poderiam apresentar uma relação entre status socioeconômico e mortalidade distinta do examinado para as brasileiras. Além disso, com o objetivo de evitar, ao menos em parte, os vieses que são típicos do efeito de adoção, foram excluídos 24 indivíduos que apresentavam uma

diferença de idade com suas mães, residindo no mesmo domicílio, superior a 55 anos. Finalmente, a informação sobre sobrevivência das mães não foi respondida por 84 indivíduos ou 0,44% da amostra. Como o número de casos é muito pequeno, optou-se por excluí-los. Portanto, dos 19.409 indivíduos entrevistados, foram considerados 19.209 (98,97%) na análise. Destes, 22,26% responderam que a mãe estava morta na data da entrevista.

Dentro dos 19.209 indivíduos considerados, a variável de escolaridade das mães, obtida através da resposta do filho entrevistado, apresenta uma proporção pequena de não resposta, igual a 3,83%. Uma vez que parte destas mães residia no mesmo domicílio dos filhos entrevistados, foi possível completar a informação com bases na própria pesquisa, para uma pequena parcela (0,41%). Assim, ao final, restaram 3,42% de casos sem informação sobre a escolaridade da mãe. Embora esta proporção seja pequena, ela é suficientemente alta para provocar viés nas estimativas e por isso, como será visto adiante, serão utilizadas técnicas de tratamento de dados faltantes para obter informações completas para a escolaridade das mães.

5.2. PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS

A Pesquisa Nacional por Amostra de Domicílios (PNAD) é uma pesquisa amostral realizada todos os anos, a não ser em anos censitários, desde 1971. A PNAD examina tanto as características gerais da população brasileira, como suas diferentes dimensões socioeconômicas e demográficas (ex: educação, trabalho, rendimento e habitação, migração, fecundidade, entre outros). A disponibilidade de informações varia a cada ano, segundo as necessidades do País.

Neste trabalho, foi utilizada a PNAD de 1996. Nesse ano, a pesquisa é representativa da população não institucionalizada residente em todo o país, com exceção da área rural da região Norte. A escolha da PNAD de 1996 se deve a existência de um suplemento sobre mobilidade social, que permite determinar o status socioeconômico do entrevistado e de seus pais. Além disso, a partir da década de 1990, as PNADs começaram a coletar informações sobre a sobrevivência das mães em seus questionários.

De um total de 331.219 pessoas entrevistadas na PNAD de 1996, foram excluídos 1.880 casos cujos entrevistados moravam nos domicílios como empregados, familiares de empregados e

pensionistas, para os quais não existiam informações sobre educação das mães. Como realizado na base de dados da PPV, não foram considerados 1.289 casos que se referem a entrevistados que nasceram fora do Brasil. Além disso, a fim de minimizar possíveis vieses do efeito de adoção, foram excluídos outros 613 entrevistados cuja diferença de idade com a mãe residente no mesmo domicílio fosse superior a 55 anos. A pergunta sobre sobrevivência das mães apresenta 0,39% de dados faltantes (1.271 casos), que também foram eliminados da amostra por ser uma proporção muito pequena. Portanto, ao final, das 331.219 pessoas entrevistadas na PNAD de 1996, foram considerados para a análise 326.226 casos ou 98,5% da amostra original. Destas, 21,19% responderam que a mãe estava morta na data da entrevista.

Em relação à variável sobre a escolaridade das mães, a proporção de dados faltantes é significativa, uma vez que esta informação não foi coletada de todos os indivíduos da PNAD. A pergunta foi feita, exclusivamente, aos entrevistados com 15 anos ou mais, que se declararam “chefes de domicílio” ou “cônjuges” dos chefes de domicílio. A partir desta pergunta pode-se obter a resposta de apenas 37,75% da amostra, restando 62,25% sem resposta. Contudo, uma vez que uma proporção das mães residia com os filhos entrevistados, foi possível conhecer sua escolaridade diretamente pela pesquisa e reduzir a quantidade de dados faltantes para 16,55% da amostra total. Essa proporção de dados faltantes é ainda superior à existente na amostra da PPV. Diante desse fato, é imprescindível entender quais tipos de riscos e vieses resultam da análise de uma variável com uma proporção grande de dados faltantes. Por isso, torna-se necessário aplicar alguma técnica estatística que permita o tratamento dos dados faltantes na variável de escolaridade das mães. A seguir, é apresentado um estudo em profundidade das implicações derivadas da existência de dados faltantes e dos métodos disponíveis que podem ser utilizados para os eu tratamento.

5.3. DADOS FALTANTES E ESTRATÉGIAS PARA O SEU TRATAMENTO

Como foi apresentado na seção anterior, a informação sobre a escolaridade das mães não foi fornecida por uma proporção elevada de indivíduos, principalmente no caso da PNAD. Este é um fenômeno bastante comum nas bases de dados de pesquisas socioeconômicas. O problema da existência de dados faltantes, chamados *missing values* na literatura anglo-saxã, é que a maioria

dos métodos estatísticos de análise supõe que as bases de dados são completas. Para lidar com esta questão, foram desenvolvidos diferentes métodos (Allison, 2001), alguns dos quais serão discutidos a seguir.

5.3.1. Tipos de valores faltantes

Em primeiro lugar, é necessário conhecer o tipo de mecanismo que define a distribuição da não resposta na base de dados, já que este mecanismo é que determina o método estatístico mais adequado para seu tratamento. Seguindo Little e Rubin (2002), uma base completa como Y e a matriz de unidades não respondidas como M . A composição dos valores faltantes na amostra é definida pela distribuição de M condicionada à base Y , ou seja, por $f(M|Y, \phi)$, onde ϕ faz referência ao conjunto de parâmetros desconhecidos.

O primeiro mecanismo de valores faltantes seria aquele em que a distribuição dos valores não respondidos depende da base Y (que inclui os valores observados e os não respondidos), de tal forma que:

$$f(M|Y, \phi) = f(M|\phi) \text{ para toda } Y, \phi. \quad (x)$$

Esse tipo de mecanismo é chamado de “completamente aleatório”. Para identificá-lo, usa-se a sigla em inglês MCAR (*Missing Completely At Random*). Para mostrar como funciona esse mecanismo, apresenta-se o seguinte exemplo: suponha-se uma base de dados Y , onde há uma única variável Y_K com dados incompletos, e as demais variáveis, $Y_1 \dots Y_{K-1}$, são completamente observadas. M é um indicador com valor 1 quando não há informação na variável Y_K e 0 quando o valor é observado. Supondo que Y e M são independentes, os valores faltantes seriam MCAR se:

$$\Pr(M_i = 1 | y_{i1}, \dots, y_{iK}; \phi) = \phi \quad (21)$$

Nesse caso, os dados completos são equivalentes a uma amostra aleatória da base total, que inclui os valores faltantes. A condição MCAR é muito forte e pouco habitual, já que a não resposta está, quase sempre, ligada às outras variáveis em estudo.

Definamos agora os dados observados da base Y como Y_{obs} , e os dados faltantes como Y_{miss} . Quando os dados não respondidos estão condicionados unicamente aos Y_{obs} , o mecanismo é chamado de aleatório e denominado pela sigla MAR (*Missing At Random*). Esta é uma condição mais flexível que a anterior e consiste em:

$$f(M|Y, \phi) = f(M|Y_{obs}, \phi) \text{ para todo } Y_{miss}, \phi \quad (22)$$

Seguindo o exemplo anterior (21) os dados seriam MAR no caso em que:

$$\Pr(M_i = 1 | y_{i1}, \dots, y_{iK}; \phi) = \Pr(M_i = 1 | y_{i1}, \dots, y_{iK-1}; \phi) \quad (23)$$

Desta forma, a distribuição dos valores faltantes é aleatória unicamente ao condicioná-la aos valores das variáveis completas, sem incluir a variável com dados faltantes, Y_K .

Por último, existe a possibilidade de que a distribuição de valores faltantes esteja condicionada a esses mesmos valores não respondidos da base de dados Y . Neste caso, o mecanismo que gerou a não resposta seria não aleatório e denominado pela sigla MNAR (*Missing Not At Random*). Seria o caso em que a distribuição de valores não respondidos da variável Y_K dependesse também da variável Y_K . Este último caso é o mais difícil de tratar analiticamente.

5.3.2. Métodos para tratar os valores faltantes

Existem inúmeros métodos para analisar bases de dados com valores faltantes. Eles têm proliferado, especialmente nas últimas décadas (Allison, 2001), graças à existência de pacotes estatísticos que permitem utilizar métodos de alta intensidade computacional. São vários os livros e artigos que descrevem exaustivamente cada um desses métodos e suas variantes, como, por exemplo: Little e Rubin (2002), Allison (2001) ou Schafer (1997), entre outros. Neste capítulo, são discutidos os métodos que são considerados mais relevantes e aqueles que serão empregados nesta tese para o tratamento da variável de escolaridade das mães.

A apresentação de vários métodos de tratamento de dados faltantes é destinada à comparação de várias estratégias que podem ser aplicadas para resolver o problema da não resposta neste trabalho. No final da seção, aqueles métodos que se adaptem melhor ao tratamento da não

resposta na escolaridade das mães serão escolhidos para ser aplicados. A escolha de mais de um método vai permitir comparar os resultados de várias metodologias e obter resultados mais confiáveis e consistentes.

5.3.2.1. Métodos baseados em dados completos

Esses métodos propõem restringir a análise dos dados àquelas unidades respondidas, excluindo os dados faltantes. Estão neste grupo, os métodos que utilizam estratégias como o *listwise* ou o *pairwise* (Allison, 2001) para eliminar total ou parcialmente as unidades não respondidas. Podem ser aplicados quando a quantidade de não resposta é pequena, poucas variáveis são afetadas, e assumindo que os valores faltantes são MCAR (Nie *et al.*, 1975). Além disso, a base de dados completa precisa ser uma amostra aleatória da amostra total. Quando estas condições não são observadas, este grupo de métodos pode induzir a grandes vieses e ser muito ineficiente (Little e Rubin, 2002). Além disso eliminar dados da amostra pode implicar uma perda de precisão. O tamanho do viés provocado pela aplicação destas técnicas depende da proporção de unidades completas existentes na amostra, do tipo de distribuição dos dados faltantes, e das diferenças existentes entre os valores faltantes e os valores completos (Little e Rubin, 2002).

5.3.2.2. Variável *dummy* para os dados faltantes

Este método propõe criar uma variável adicional que recolha todos os valores faltantes da variável incompleta. Por exemplo (Allison, 2001): tendo uma variável Y_K com dados incompletos, deve ser criada a variável D (*dummy*), com valor 1 quando Y_K não tem informação e com valor 0 quando o variável de Y_K é respondida. Da mesma forma, pode se recodificar Y_K como Y_K^* de modo que:

$$Y_K^* = \begin{cases} Y_K & \text{quando o valor é completo} \\ c & \text{quando o valor é incompleto} \end{cases} \quad (24)$$

onde c é uma constante. No caso em que a variável é categórica, o tratamento dos valores faltantes é ainda mais simples, podendo ser criada uma categoria adicional que reúna todos os valores não respondidos da variável em questão.

A principal vantagem desse método é que a informação de variáveis completas, de indivíduos sem informação em outras variáveis, não se perde. O problema derivado desse método é que, ao realizar uma regressão com a variável com *dummy*, a estimação dos coeficientes incorpora, inevitavelmente, viés. Segundo Jones (1996), este método sobreestima a variância do resíduo da regressão, diminuindo a significância do modelo.

5.3.2.3. Métodos de Imputação

A imputação consiste em substituir os valores não respondidos por valores gerados através de valores preditos. Esta substituição é realizada com base em pressupostos extraídos dos dados completos (Little e Rubin, 2002). Ou seja, analisando o comportamento das variáveis completas associadas à variável com não resposta é possível prever o valor dos dados faltantes. Entre os muitos métodos que permitem a imputação das cédulas sem informação, apresentam-se aqui o da imputação pela média, da imputação por regressão, da imputação múltipla, assim como os métodos denominados *hot deck*.

A. Imputação pela média

Neste caso, as cédulas sem informação em uma variável Y_K são substituídas pela média das unidades com resposta da mesma variável Y_K . A média pode ser calculada condicionada a um grupo de variáveis completas (Little e Rubin, 2002). Este método provoca viés na estimação dos parâmetros e dos desvios padrão naqueles modelos que incluam a variável com a imputação. Este viés deriva do fato de transformar aqueles valores faltantes que poderiam ser extremos em valores no centro da distribuição, reduzindo artificialmente a variabilidade com a imputação.

B. Imputação por regressão

Neste método, os dados faltantes são estimados através de valores preditos de uma regressão das unidades observadas, em função de diversas variáveis completas (Little e Rubin, 2002). Considere-se o exemplo de uma base composta por um conjunto de variáveis, Y_1, \dots, Y_{K-1} completamente observadas, e uma única variável Y_K com as primeiras r observações completas e as últimas $n-r$ observações sem resposta. A regressão para imputar os valores sem resposta tem como variável dependente Y_K em função de Y_1, \dots, Y_{K-1} , baseada unicamente nos r casos completos. Os valores faltantes são substituídos por predições da regressão. Por exemplo, temos o caso i com y_{iK} sem responder e y_{i1}, \dots, y_{iK-1} observados. A cédula faltante, y_{iK} , é substituída pelo valor estimado \hat{y}_{iK} através dos parâmetros estimados da regressão de Y_K em função de Y_1, \dots, Y_{K-1} , baseada nos r casos completos:

$$\hat{y}_{iK} = \tilde{\beta}_{K0.1.2\dots K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj.1.2\dots K-1} \cdot y_{ij}, \quad (25)$$

onde $\tilde{\beta}_{K0.1.2\dots K-1}$ é o intercepto, e $\tilde{\beta}_{Kj.1.2\dots K-1}$ são os coeficientes de Y_j . A imputação da média pode ser considerada um caso especial da imputação por regressão, onde as variáveis preditivas são indicadores *dummy* dos valores das variáveis completas (Little e Rubin, 2002).

C. Imputação por regressão estocástica

Segue-se neste caso o mesmo princípio proposto na imputação por regressão, só que adicionando à regressão um resíduo gerado aleatoriamente, que agrega incerteza ao valor predito (Little e Rubin, 2002). Seguindo o exemplo anterior (fórmula 25), em vez de imputar um valor condicionado a médias, neste caso, imputa-se um valor condicionado, além das médias, a um componente aleatório:

$$\hat{y}_{iK} = \tilde{\beta}_{K0.1.2\dots K-1} + \sum_{j=1}^{K-1} \tilde{\beta}_{Kj.1.2\dots K-1} \cdot y_{ij} + z_{iK}, \quad (26)$$

onde z_{iK} é um número aleatório de uma distribuição normal com média zero e variância $\tilde{\sigma}_{KK-1,2,\dots,K-1}$, que é a variância do resíduo da regressão de Y_K em Y_1, \dots, Y_{K-1} baseada nos casos completos. Ao adicionar o número aleatório, a imputação se torna um número aleatório da distribuição de predições dos valores faltantes, deixando de ser unicamente a média. O resultado é que as distorções provocadas na imputação são minimizadas.

Este último método é claramente superior aos outros métodos de imputação apresentados até agora (pela média e por regressão simples), como Little e Rubin (2002) demonstram, através de um exemplo com uma base de duas variáveis, onde uma delas apresenta dados faltantes MCAR. Comparado com os outros métodos, é o único que obtém estimadores consistentes, mesmo com dados faltantes MAR. De qualquer modo, apresenta alguns inconvenientes. O número aleatório reduz a eficiência. Além disso, os desvios padrão dos parâmetros estimados com a imputação por regressão estocástica são muito pequenos, já que não incorporam a incerteza da imputação. Já a imputação múltipla, apresentada a seguir, tenta resolver estes dois inconvenientes.

D. Imputação múltipla (IM)

Rubin (1976, 1987) foi o primeiro autor a apresentar a IM. Esse método consiste em gerar um vetor de valores imputados para substituir os valores faltantes, onde são gerados dois ou mais valores ($D \geq 2$) para cada valor não respondido, ao invés de utilizar um único valor imputado como nas imputações anteriores (Little e Rubin, 2002). Cada uma das D imputações é obtida através de um mesmo modelo de não resposta, com um componente aleatório, que reflete a incerteza dos dados. Através destas imputações podem ser geradas D amostras de dados completas. Cada uma destas amostras pode ser analisada, separadamente, através do mesmo método estatístico escolhido, tal como se analisariam os dados, caso estes fossem completos. Adicionalmente, os resultados derivados da análise de cada uma das amostras por separado deve ser posteriormente combinado para ter uma única resposta.

No seguinte exemplo, assumamos que θ é o parâmetro que se deseja estimar, e $\hat{\theta}_d$ e \hat{W}_d ($d=1, \dots, D$) são os D estimadores e suas associadas variâncias desse parâmetro, resultantes da análise com

o mesmo modelo estatístico, de cada uma das D amostras imputadas. A combinação destes parâmetros estimados para obter um único parâmetro final é:

$$\bar{\theta}_D = \frac{1}{D} \sum_{d=1}^D \hat{\theta}_d \quad (27)$$

seguindo os mesmos princípios das estimativas *bootstrap* apresentadas no ponto 4.3.1 desta Tese. O modelo de não resposta utilizado na imputação múltipla depende das características da variável com dados faltantes.

Esse método corrige algumas das desvantagens da imputação simples. Em primeiro lugar, permite utilizar modelos diferentes para cada uma das variáveis faltantes, permitindo aplicar modelos que se adaptem a variáveis com distribuições categóricas. Além disso, esse método reflete melhor a variabilidade amostral e a incerteza do modelo de não resposta, já que as estimativas são resultado de várias imputações, ao invés de uma única imputação, oferecendo estimadores mais consistentes (Little e Rubin, 2002).

E. Imputação por *hot deck*

A imputação por *hot deck* consiste em substituir os valores faltantes daquelas unidades sem informação em uma variável concreta, por valores sorteados aleatoriamente de unidades “semelhantes” nas outras variáveis repostas e que também possuem informação na variável onde as primeiras unidades têm informação omitida (Little e Rubin, 2002). O método varia segundo o tipo de modelo que seja utilizado para encontrar essas unidades “semelhantes. Por exemplo: considere-se que em uma amostra de n unidades selecionados de uma amostra maior N , r das n unidades da variável Y responderam. N , n e r são fixos. Os primeiros $r < n$ unidades de n são os que respondem. A média de Y pode ser estimada como uma combinação da média das unidades respostas (\bar{y}_R) e a média das unidades faltantes (\bar{y}_{NR}) da seguinte forma:

$$\bar{y}_{HD} = \frac{(r\bar{y}_R + (n-r)\bar{y}_{NR}^*)}{n}, \quad (28)$$

considerando que

$$\bar{y}_{NR}^* = \sum_{i=1}^r \frac{H_i y_i}{n-r}, \quad (29)$$

onde H_i é o número de vezes que y_i é usado como um substituto para os valores não respondidos de Y , e $\sum_{i=1}^r H_i = n-r$ é o número de valores faltantes. As propriedades de y_{HD} dependem do procedimento utilizado para gerar os números $[H_1, \dots, H_r]$. Os valores imputados podem ser entendidos como valores selecionados da amostra de valores completos, podendo-se conhecer a distribuição de $[H_1, \dots, H_r]$ depois de aplicar repetidas vezes o método *hotdeck*.

O método de imputação por *hotdeck* gera estimativas com um viés menor do que no caso da imputação pela média ou da imputação simples (Little e Rubin, 2002). O principal problema do método *hotdeck* é que se torna difícil de aplicar quando a amostra de dados a ser analisada contém muitas variáveis com dados faltantes, reduzindo o número de unidades completas que podem substituir as unidades incompletas. Além disso, o método *hotdeck* estima desvios padrões dos parâmetros menores que os dos esperados e não leva em conta a incerteza da imputação (Little e Rubin, 2002). Portanto, o método de imputação múltipla se mostra superior a esse método.

5.3.2.4. O algoritmo de Expectativa e Maximização (EM)

O algoritmo EM é um algoritmo de iteração baseado na estimativa de máxima verossimilhança (MV) no caso de dados incompletos. Foi apresentado pela primeira vez por Dempster *et al.* (1977).

Suponha-se uma base de dados Y , com presença de dados faltantes, e uma função de densidade $f(Y|\theta)$, onde θ é o conjunto de parâmetros desconhecidos. Esta base está composta por valores observados, Y_{obs} , e faltantes, Y_{miss} . O objetivo é maximizar uma função de verossimilhança que, assumindo uma distribuição MAR dos valores faltantes, seria:

$$L(\theta|Y_{obs}) = \int f(Y_{obs}, Y_{miss}|\theta) dY_{miss} \quad (30)$$

Usualmente, adota-se o logaritmo natural da função de verossimilhança para poder maximizá-la, representada por $l(\theta|Y_{obs})$. A maximização desta função, com a base decomposta em valores observados e não observados, pode levar a soluções muito complexas, usando os métodos convencionais (Little e Rubin, 2002). Uma alternativa é utilizar o algoritmo EM, que relaciona a estimação de MV dos parâmetros θ de $l(\theta|Y_{obs})$ com a estimativa da verossimilhança da base de dados inteira, $l(\theta|Y)$, incluindo os valores faltantes.

O nome do algoritmo EM se refere à aplicação dos passos E e M repetidamente. O passo M consiste em realizar a MV do modelo para estimar os parâmetros θ , utilizando a base Y inteira com os dados faltantes imputados. Na maioria das vezes, a primeira imputação é realizada seguindo um método bem simples, como a imputação pela média.

O passo E acha a expectativa condicional dos valores faltantes, utilizando os parâmetros estimados no passo M. Seguidamente, os valores faltantes são substituídos pelas expectativas estimadas.

A repetição dos passos E e M segue esta ordem (Allison, 2001):

1. Substituir os valores faltantes por valores esperados, usando um método de imputação. Este é o primeiro passo E, que é realizado antes da estimação dos parâmetros.
2. Estimar os parâmetros do modelo, utilizando máxima verossimilhança, baseado nos dados completos imputados. Este é o denominado passo M do algoritmo.
3. Re-estimar os valores imputados, tomando os parâmetros estimados no passo M como certos. De novo, o passo E, já que a estimação dos valores imputados é, em realidade, uma expectativa.
4. Re-estimar os parâmetros, utilizando os novos valores faltantes imputados, que consiste em uma repetição do passo 2.
5. Continuar a repetição dos passos 2 e 3 até chegar à convergência dos resultados.

É importante notar que o passo E não necessariamente substitui literalmente os dados sem informação. A ideia do EM é que não trata os valores faltantes como os Y_{miss} , mas como funções de Y_{miss} dentro da base inteira Y no modelo de verossimilhança $l(\theta|Y)$ (Little e Rubin, 2002).

A vantagem principal do algoritmo EM é que podem ser usadas todas as variáveis da base completa, se for preciso, no modelo. Além disso, permite que diferentes variáveis com dados faltantes apresentem padrões diferentes de não resposta, como no caso da imputação múltipla (Allison, 2001).

O principal problema desse método é que ele não permite estimar erros padrão. Além disso, a implementação do algoritmo EM precisa de uma programação complexa, que é difícil de encontrar nos diversos pacotes estatísticos. Os programas existentes nos pacotes de tratamento de bases de dados mais usuais acabam restringindo a aplicação do algoritmo EM a casos específicos, reduzindo a flexibilidade que o método pode oferecer.

5.4. APLICAÇÃO DOS MÉTODOS DE TRATAMENTO DE DADOS FALTANTES DA VARIÁVEL DE ESCOLARIDADE DAS MÃES NA PPV E NA PNAD

A seguir são apresentados os procedimentos usados para tratar os valores faltantes da variável escolaridade nas duas bases de dados. Para se conhecer as técnicas adequadas para o tratamento dos valores faltantes da variável escolaridade, é preciso fazer um estudo em profundidade da não resposta e sua associação com outras variáveis da base de dados. A seguir é realizada a análise da não resposta no caso das duas bases de dados (PPV e PNAD) utilizadas, finalizando-se com a apresentação dos modelos escolhidos para o tratamento dos dados faltantes.

5.4.1. Análise da informação sobre escolaridade das mães na PPV

A variável escolaridade apresenta 3,4% de dados não respondidos. Porém, esses dados não estão distribuídos de forma uniforme. Por exemplo, entre os que não responderam sobre a escolaridade da mãe, há uma maioria de mães mortas: cerca de 52% dos casos. Essa é uma proporção superior a duas vezes àquela encontrada na amostra total, cuja proporção de mães mortas é de 22,3%. Os

dados não são, portanto, MCAR. Porém, ainda se pode considerar que sejam MAR, controlando por outras variáveis da amostra.

Determinar que os dados são MAR é uma tarefa difícil. De qualquer forma, é importante definir as variáveis que vão servir para os modelos de tratamento dos dados faltantes da variável de escolaridade. Essas variáveis podem estar correlacionadas com a escolaridade das mães, mas de forma exógena, ou seja, a escolaridade das mães não deve determinar as variáveis escolhidas para evitar dupla causalidade nos modelos. Além disso, como a variável escolaridade vai ser utilizada em um modelo de regressão do número de óbitos, a sobrevivência das mães não pode ser utilizada para imputar os valores faltantes dessa variável. Apesar disso, a sobrevivência da mãe é um fator muito importante para definir a existência de não resposta na escolaridade da mãe, já que a morte da mãe pode estar influenciando na falta de memória por parte do filho.

Diante dessas restrições, as variáveis escolhidas na PPV, para analisar a distribuição da não resposta, incluem variáveis sobre as características do filho e sobre as características do pai do entrevistado⁷. Apesar de não ser possível saber se o pai e a mãe estavam casados, pode-se esperar que as características socioeconômicas do pai sejam uma aproximação das características da mãe. Entre as variáveis relacionadas ao filho entrevistado, foram incluídas: a idade, a região de nascimento, se morou em área rural ou urbana até os 15 anos, em que área mora atualmente e a cor. Entre as características relacionadas ao pai, foram incluídas: se está vivo no momento da entrevista e seu nível de escolaridade. Várias dessas variáveis apresentam dados faltantes, mas estes foram tratados, em todos os casos, como uma categoria adicional.

A distribuição da não resposta segundo essas variáveis pode ser observada na TAB. 1. Como já mencionado, a proporção de mães mortas entre os filhos que não responderam sobre a escolaridade das mães (52,2%) é muito superior, quando comparada aos que responderam (21%). Isto pode estar correlacionado com o fato de que a idade média dos entrevistados que não responderam é 12 anos maior do que a idade média dos que responderam, sugerindo que, por ser pessoas geralmente mais velhas e que, portanto, podem ter esquecido ou desconheciam a escolaridade das mães. A distribuição por região de nascimento é muito parecida entre as duas

⁷ As perguntas que geraram estas variáveis estão recolhidas no Anexo, na TAB. A 3

categorias, tal como o teste estatístico indica. Apesar de observar que entre os que não responderam, existe uma proporção maior de filhos que moravam na área rural antes dos 15 anos, esta diferença entre as duas categorias não é significativa. No entanto, é significativa a maior proporção de pessoas morando na área urbana na data da entrevista entre os que responderam sobre a escolaridade das mães. Finalmente, a proporção de filhos que se declararam de cor branca parece não ser diferente, segundo a resposta ou não resposta sobre a escolaridade das mães.

Em relação às características dos pais, a proporção de pais sobreviventes entre os que não responderam sobre a escolaridade das mães (32%) é superior aos que responderam (56%), diferença que é altamente significativa. Essas diferenças também são expressivas e estatisticamente significativas por grupo de escolaridade dos pais, concentrando-se uma maior proporção de indivíduos que não conheciam a escolaridade dos pais entre os que também não responderam sobre a escolaridade das mães.

A distribuição da não resposta segundo as variáveis apresentadas foi significativamente diferente à da distribuição dos dados completos da escolaridade das mães unicamente em algumas das variáveis analisadas. Estas variáveis são, então, adequadas para serem utilizadas no modelo de não resposta. Talvez o único problema pode decorrer de se estar trabalhando com variáveis que também apresentam não resposta, como as características dos pais dos entrevistados. No entanto, como estas variáveis são especialmente importantes para determinar o status socioeconômico das mães, espera-se que os benefícios de se utilizar estas variáveis sejam superiores ao viés que possam introduzir. Além disso, o fato de que algumas variáveis não foram significativas indica que os dados faltantes são aleatórios em função de algumas variáveis, podendo assumir com mais certeza que são MAR (Little e Rubin, 2002).

TABELA 1. Distribuição da resposta e não resposta da variável escolaridade das mães, segundo diversas características dos entrevistados na PPV 1996

	Escolaridade da mãe		Valor P
	Resposta	Não resposta	
Mãe morta (<i>proporção</i>)	21,06	52,17	<0.001
Idade (<i>média</i>)	27,17	39,78	<0.001
Região de nascimento (<i>proporção</i>)			
Nordeste	45,50	46,76	0,144
Sudeste	52,78	51,51	0,093
Morou na área rural até os 15 anos (<i>proporção</i>)	29,91	34,60	0,337
Região urbana (<i>proporção</i>)	78,19	81,44	0,003
Cor branca (<i>proporção</i>)	54,07	47,03	0,110
Pai morto (<i>proporção</i>)	30,94	51,36	<0.001
não resposta	1,05	3,53	<0.001
Educação do pai (<i>proporção</i>)			
Sem escolaridade	34,25	28,76	0,017
1 a 8 anos	47,34	24,77	<0.001
9 e mais	12,24	2,54	<0.001
não resposta	6,17	43,93	<0.001

Fonte: IBGE, PPV 1996

5.4.2. Análise da informação sobre a escolaridade das mães na PNAD

Como mencionado anteriormente, no caso da PNAD, a informação sobre escolaridade da mãe não foi coletada para todos os membros do domicílio, resultando em uma proporção de dados faltantes superior ao encontrado na PPV; cerca de 16,6%. Entre os indivíduos que não responderam sobre a escolaridade das mães, 41,4% são pessoas cuja mãe estava morta na data da

entrevista; uma prevalência duas vezes maior do que a encontrada para o total da amostra (21,2%). Ao constatar essa forte associação entre o status de sobrevivência das mães e a presença de dados faltantes na escolaridade, a possibilidade de que a não resposta apresente uma distribuição MCAR desaparece. Para saber se os valores são MAR é necessário realizar uma análise da distribuição da não resposta segundo várias características que podem ser utilizadas no posterior modelo de tratamento dos dados faltantes.

A disponibilidade de variáveis aceitáveis sobre informação dos pais na PNAD é menor do que na PPV, já que as características socioeconômicas dos pais dos entrevistados apresentam uma proporção de dados faltantes muito elevada, para serem utilizadas na aplicação de um modelo de não resposta. Por isso, no caso da PNAD, as variáveis tiveram que ser restringidas unicamente às características dos filhos. Foram incluídas: a idade, a região de nascimento, a região de residência, a situação de residência (urbana ou rural), e a cor⁸. Além disso, incluiu-se a renda domiciliar per capita referente ao filho, já que esta era a única variável existente que fosse de fato socioeconômica. Note-se que a renda per capita dos filhos pode não ser exógena na hora de imputar os valores faltantes da variável de escolaridade das mães. Porém, considerou-se a melhor opção que a PNAD oferece para medir o status socioeconômico das mães dos entrevistados. Optar por outra variável, como o nível de escolaridade dos filhos, por exemplo, seguramente implica uma correlação mais forte com a escolaridade das mães. Espera-se que os benefícios resultantes de utilizar esta variável sejam superiores ao fato de não incluí-la no modelo de não resposta.

A distribuição de cada uma dessas variáveis em função da existência de resposta ou não para a escolaridade das mães é apresentada na TAB. 2. Como já mencionado, a proporção de mães mortas é significativamente inferior entre aqueles que responderam (17%) quando comparada com aqueles que não responderam (41%) sobre a escolaridade das mães. Da mesma forma que com os dados da PPV, a idade média entre os entrevistados que não responderam é de 34 anos, superior à idade média dos que responderam, de 26 anos de idade. Também existem diferenças significativas segundo a região de nascimento e de residência. Entre os que não responderam,

⁸ As perguntas que geraram estas variáveis estão recolhidas no Anexo, na TAB. A 4

existe uma maior proporção de pessoas nascidas e residindo no Nordeste do que entre os entrevistados que conheciam a escolaridade das mães. Finalmente, as diferenças existentes entre a distribuição da resposta e a não resposta, segundo as variáveis área de residência, cor e renda domiciliar per capita, não são muito expressivas, porém, o teste estatístico foi significativo. Entre os filhos que reportaram a escolaridade das mães, a proporção de pessoas morando em áreas urbanas é ligeiramente inferior, a proporção de pessoas brancas é superior e a renda domiciliar per capita é inferior, em comparação com aqueles entrevistados que não conheciam a escolaridade das mães.

A distribuição da não resposta da escolaridade das mães foi significativamente diferente da distribuição da informação completa em todas as variáveis analisadas. Este fato indica que existe uma concentração da não resposta na base de dados, sugerindo que existe algum mecanismo de seleção. Esta conclusão concorda com o fato de que somente os chefes de domicílio e seus cônjuges foram perguntados sobre a escolaridade das mães. Apesar dessa concentração, não pode ser afirmado que os dados não sejam MAR, já que, controlando por essas variáveis a distribuição de dados faltantes pode passar a ser aleatória. Porém, é preciso considerar que existe alguma possibilidade de que os modelos de não resposta escolhidos não sejam tão ajustados como no caso da PPV. Outra estratégia que pode ser aplicada para melhorar o ajuste dos modelos com dados faltantes é restringir a amostra de entrevistados aos adultos, aumentando o número de chefes e cônjuges na amostra.

TABELA 2. Distribuição da resposta e não resposta da variável de escolaridade das mães, segundo diversas características dos entrevistados na PNAD 1996

	Escolaridade da mãe		Valor P
	Resposta	Não resposta	
Mãe não sobrevivente (proporção)	17,34	40,58	<0,001
Idade (média)	26,49	34,57	<0,001
Região de nascimento (proporção)			
Nordeste	33,15	39,80	<0,001
Sudeste	41,45	35,59	<0,001
Região de residência (proporção)			
Nordeste	28,92	33,47	<0,001
Sudeste	44,18	39,82	<0,001
Região urbana (proporção)	78,93	80,84	<0,001
Cor branca (proporção)	56,16	50,12	<0,001
Renda per capita (média)	242,48	250,68	<0,001

Fonte: IBGE, PNAD 1996

5.4.3. Modelos de tratamento dos dados faltantes da variável escolaridade das mães

Uma vez analisada a distribuição dos valores da variável escolaridade e encontradas as possíveis variáveis associadas à sua distribuição, é preciso determinar qual é o melhor método para o tratamento dos dados faltantes. A literatura mostra que algumas técnicas são superiores a outras, porém, parece importante testar várias delas para achar a mais adequada ao objeto de estudo. Com o objetivo de obter estimativas mais consistentes com a variável escolaridade das mães,

neste trabalho serão testados diversos modelos de tratamento dos valores faltantes dessa variável, apresentados à seguir.

Em primeiro lugar, e como forma de controle dos posteriores tratamentos, o primeiro método consiste em não imputar nenhum valor, agrupando os dados faltantes em uma categoria a parte (*dummy*) nos modelos de regressão a serem aplicados. Desta forma, os dados da variável de escolaridade são categorizados em três grupos, mais a categoria de dados faltantes: 1) sem escolaridade; 2) com 1 a 8 anos de estudo; 3) 9 anos de estudo ou mais; 4) dados faltantes. Apesar de que esse tipo de tratamento dos dados faltantes provoca viés nos coeficientes estimados, este modelo pode servir como referência dos outros, além de mostrar como os dados faltantes se comportam em relação à variável dependente.

O segundo método consiste em realizar uma imputação por regressão estocástica. É necessário escolher as variáveis que vão ser utilizadas para a imputação. No caso da PPV, essas foram: a idade do filho, a idade do filho ao quadrado, a região de nascimento, se o filho morou em área rural ou urbana antes dos 15 anos e em que área mora atualmente, a cor, se o pai está vivo ou morto no momento da entrevista e o nível de escolaridade do pai. No caso da PNAD, as variáveis foram: a idade do filho, a idade do filho ao quadrado, a região de nascimento e de residência do filho, se o filho mora em área urbana ou rural, a cor, e a renda domiciliar per capita do filho.

A regressão para imputar os valores faltantes assume uma distribuição normal da variável resposta, o que significa que a variável escolaridade deve ser tratada como contínua. Porém, no questionário, esta variável é codificada em 10 intervalos de anos de estudo. Apesar disso, assume-se que estes intervalos, ordenados de menor a maior escolaridade podem ser uma aproximação de uma distribuição contínua.

Em terceiro lugar, aplicou-se o método de imputação múltipla. A técnica escolhida foi a que oferece o programa Stata 10.0, explicada em detalhe em Royston (2004) e baseada no método descrito por Van Buuren *et al.* (1999). É um método de imputação múltipla de equações em cadeia que, usando a sigla em inglês, é chamado de *MICE* (*Multiple Imputation by Chained Equations*). Esse método consiste em gerar várias amostras aleatórias a partir de regressões univariadas e iterativas de cada uma das variáveis incompletas condicionadas à distribuição das outras variáveis com dados faltantes (Van Buuren *et al.*, 1999). Por exemplo, assumindo-se uma

base de dados com três variáveis com valores incompletos, Y_1 , Y_2 e Y_3 , o método pode ser descrito da seguinte forma: em primeiro lugar, os valores faltantes de Y_1 são imputados de forma aleatória, utilizando todas as variáveis completas e as variáveis com valores faltantes. Em seguida, a variável Y_2 é imputada com as variáveis completas, incluindo a variável Y_1 já imputada. O mesmo passo é realizado com Y_3 . O processo se repete, desta vez usando as variáveis Y_2 e Y_3 imputadas para estimar os valores faltantes de Y_1 novamente, e assim por diante, até a imputação de número vinte que, segundo Van Buuren *et al.* (1999), é um número suficientemente grande para estabilizar a imputação.

O modelo utilizado para imputar a variável com valores faltantes, permite tratar esta como categórica. Assim, a variável escolaridade das mães pode ser dividida em quatro categorias: sem escolaridade, de 1 a 4 anos de estudo, de 5 a 8 anos de estudo, e de 9 e mais. O modelo de regressão escolhido para sua imputação foi uma regressão logística ordenada. As variáveis independentes do modelo são as mesmas que no caso da imputação com a regressão estocástica. Foram realizadas 20 imputações diferentes, onde a média dos valores imputados foi utilizada como resultado final⁹.

Outro modelo de imputação utilizado foi o da imputação *hotdeck*. O método *hotdeck* escolhido é o que oferece o programa estatístico Stata 10.0, baseado no modelo de imputação *bootstrap* de aproximação bayesiana de Rubin e Schenker (1986). A substituição das unidades com valores incompletos pela informação de unidades completas é realizada gerando, inicialmente, uma amostra *bootstrap* de unidades completas. Posteriormente, é gerada uma amostra aleatória da amostra *bootstrap* de unidades completas, que é usada para preencher as unidades com valores sem resposta. Por ser um modelo baseado em uma imputação estocástica, é importante gerar mais de uma imputação para poder incorporar a variabilidade do modelo.

⁹ O ideal seria que cada uma das imputações da variável de escolaridade das mães fosse utilizada para gerar um modelo de número de mortes. Desta forma, os coeficientes finais do modelo de sobrevivência analisados seriam a média de todos os modelos com cada uma das variáveis imputadas. Infelizmente, como o modelo de regressão do número de mortes já é resultado de coeficientes *bootstrap*, optou-se por essa simplificação na imputação. Caso contrário, a distribuição dos coeficientes *bootstrap* teria que ser multiplicada por 20 vezes aumentando o número de regressões até 10.000.

Nesta imputação não foi possível incluir as mesmas variáveis que nos modelos anteriores. O uso de muitas variáveis no método de imputação *hotdeck* diminui a probabilidade de encontrar unidades semelhantes às unidades com dados faltantes. No caso dos dados da PPV, o modelo foi restringido às seguintes variáveis: a idade, a região de nascimento, área de residência (rural ou urbana) até os 15 anos, área de residência (rural/urbana) no momento da entrevista, e o estado de sobrevivência do pai. Com a PNAD, as variáveis escolhidas foram: a idade, a região de nascimento e a região de residência, área de residência (urbana/rural) e cor. A variável de escolaridade foi tratada como categórica, utilizando a mesma classificação adotada no caso da imputação múltipla. Como na imputação múltipla, a imputação final é resultado da média de 20 imputações *hotdeck*.

O último modelo de imputação aplicado à variável escolaridade é o algoritmo de Expectativa e Maximização (EM). Foi utilizado o modelo apresentado pelo programa Amelia II do pacote estatístico R 2.9.0. Neste programa, aplica-se uma versão particular do algoritmo EM combinado com um procedimento *bootstrap*, que permite imputar valores faltantes. O modelo assume que a base de dados é multivariada e com distribuição normal. Apesar desse pressuposto ser uma aproximação bastante limitada da realidade, a literatura mostra que o modelo funciona de forma aceitável para dados discretos ou categóricos (King *et al.*, 2001).

A particularidade deste método é que, ao invés de estimar os parâmetros através de uma simples máxima verossimilhança, o modelo incorpora um componente aleatório na estimativa e cria uma distribuição *bootstrap* dos parâmetros. A partir desses parâmetros, estimam-se as expectativas dos valores faltantes. O procedimento se repete iterativamente, até chegar à convergência dos resultados. Uma vez estimados os parâmetros convergentes, estes são usados para estimar aleatoriamente a imputação dos valores faltantes. Como o procedimento final de imputação é aleatório, é recomendado que se realizem várias imputações, aproximadamente vinte (Honaker *et al.*, 2009). A imputação final é o resultado da média das vinte imputações.

As variáveis utilizadas neste modelo são as mesmas que foram usadas na imputação simples e múltipla, porém, sem incluir a idade ao quadrado. O modelo assume que todas as variáveis são contínuas e com distribuição normal, o que não é o caso das variáveis escolhidas, já que muitas são categóricas ou dicotômicas. Da mesma forma que foi realizado com a imputação estocástica,

assumiu-se que as 10 categorias iniciais da variável de escolaridade aproximavam-se de uma distribuição contínua.

5.4.4. Resultados do tratamento dos dados faltantes da variável escolaridade das mães

Na TAB. 3 podem ser observadas as distribuições dos valores imputados, tanto para a PPV como para a PNAD. As imputações estão separadas em dois grupos, aquelas que trataram a variável escolaridade como contínua (imputação estocástica simples e algoritmo EM) e as que a trataram como categórica (imputação múltipla e imputação *hotdeck*). Quando a variável escolaridade é contínua, assume-se uma distribuição normal, o que faz com que a imputação apresente valores negativos e uma dispersão maior. Esta dispersão é maior no caso da PNAD do que no da PPV, devido ao uso da variável de renda domiciliar per capita nos modelos de não resposta, que apresenta uma dispersão de valores muito ampla. Nas imputações com a variável escolaridade categórica, os valores máximos e mínimos são restringidos pelas categorias existentes (de 0 a 3). É importante observar que a distribuição da imputação foi muito parecida entre as imputações que utilizam a mesma variável de escolaridade, tanto com os dados da PPV como da PNAD. Este resultado dá consistência às técnicas de imputação utilizadas.

Em todos os casos, seja porque a variável foi tratada como contínua no modelo, ou porque a imputação final é o resultado de uma média de imputações, os valores não são exatos. Por isso, é preciso arredondar os dados para proceder à sua categorização e aplicar posteriormente os modelos de mortalidade.

TABELA 3. Distribuição dos valores imputados da variável escolaridade das mães, segundo o tipo de imputação, PPV e PNAD, 1996

Base	Tipo de imputação	Média	Desvio padrão	Min.	Max.
PPV	Imputação por regressão estocástica simples	2,787	2,074	0,6075	10
	Algoritmo EM	2,817	2,074	-0,506	10
	Imputação por regressão múltipla	0,989	0,992	0	3
	Imputação por hotdeck	0,994	0,994	0	3
PNAD	Imputação por regressão estocástica simples	2,854	1,974	-0,444	28,893
	Algoritmo EM	2,866	1,999	-2,513	28,057
	Imputação por regressão múltipla	1,070	0,945	0	3
	Imputação por hotdeck	1,069	0,940	0	3

Fonte: IBGE, PPV e PNAD 1996

O arredondamento pode ser realizado de forma simples, por associação com a categoria mais próxima ao valor imputado. No entanto, vários autores destacam que, se os dados imputados não são monotônicos, esse tipo de técnica pode provocar grandes vieses (Ake, 2005; Yucel e Zaslavsky, 2004). Uma das alternativas é tratar os dados sem arredondar (Ake, 2008), porém, neste caso, essa opção deve ser descartada porque a estrutura da variável imputada é de natureza categórica. Outra estratégia consiste em arredondar os dados segundo uma distribuição aleatória, gerada através dos dados respondidos (Yucel e Zaslavsky, 2004). Estes autores propõem utilizar o conjunto de indivíduos com informação sobre a escolaridade das mães para gerar uma

distribuição aleatória de indivíduos igual ao número de pessoas que não responderam essa variável. Essa distribuição deve estar condicionada às variáveis utilizadas para imputar os valores não respondidos. Dessa forma, gera-se um grupo de dados faltantes “fictícios” e com distribuição MAR, para os quais podem ser observados os valores da variável que vai ser imputada, neste caso, a escolaridade das mães. A partir desse grupo de valores faltantes é possível conhecer a proporção de valores imputados que pertencem a cada categoria da variável escolaridade. Os valores faltantes imputados devem seguir a mesma distribuição nas diversas categorias que os valores faltantes “fictícios”. A distribuição dos valores faltantes gerados aleatoriamente através dos dados completos é apresentada no Anexo na TAB. A 5.

TABELA 4. Proporção de mães mortas segundo as categorias da variável escolaridade das mães, por tipo de tratamento dos valores faltantes, PPV e PNAD, 1996

Base	Tipo de tratamento dos valores faltantes	Sem escolaridade	1 a 4 anos de estudo	1 a 8 anos de estudo	9 e + anos de estudo	Não resposta
PPV	Variável dummy	34,93	17,44	6,48	5,62	52,17
	Imputação por regressão estocástica simples	36,57	19,03	7,86	6,24	-
	Algoritmo EM	36,71	18,94	7,38	6,57	-
	Imputação por regressão múltipla	36,60	18,89	7,97	6,49	-
	Imputação por hotdeck	36,52	18,91	8,31	6,45	-
PNAD	Variável dummy	32,19	15,24	3,45	3,83	41,70
	Imputação por regressão estocástica simples	38,96	18,55	6,83	6,13	-
	Algoritmo EM	37,03	19,59	7,84	6,51	-
	Imputação por regressão múltipla	38,02	19,10	7,40	6,54	-
	Imputação por hotdeck	37,51	19,71	7,19	6,03	-

Fonte: IBGE, PPV e PNAD 1996

A TAB. 4 apresenta as estimativas de proporção de mães mortas, categoria de escolaridade segundo o tipo de imputação da variável escolaridade. Pode-se observar como as diversas distribuições são muito similares, independentemente do tipo de imputação, dando uma maior confiabilidade e consistência aos resultados da imputação.

Dentre todas as imputações realizadas, a que parece mais confiável é a do algoritmo EM. Esta imputação é escolhida por muitos estatísticos porque o uso da máxima verossimilhança para estimar dados imputados permite não impor um comportamento predeterminado às variáveis. Além disso, o método é iterativo e os valores imputados acabam sendo resultado de uma convergência. Porém, neste trabalho, não foi possível adaptar este modelo a uma distribuição categórica, como a da variável em estudo, a escolaridade das mães. Portanto, outro modelo de imputação torna-se mais adequado para este caso, o de imputação múltipla. Este método respeita a categorização da variável a ser imputada e considera a incerteza dos valores faltantes, além de se comportar, também, de forma iterativa, sendo que os dados imputados também são fruto da convergência do modelo.

6. ESTIMATIVAS DE MORTALIDADE ADULTA COM INFORMAÇÕES RETROSPECTIVAS DE SOBREVIVÊNCIA

6.1. ESTIMATIVAS DE MORTALIDADE ADULTA COM BASE NO MÉTODO DE ORFANDE

Antes de discutir os resultados da aplicação do método estocástico, nesta primeira seção, serão apresentadas as estimativas de mortalidade geradas através do método de orfandade de Brass, utilizando dados da PPV e da PNAD. O uso deste método clássico nesta Tese tem como objetivo gerar estimativas de mortalidade que sejam comparáveis às estimativas geradas através do método estocástico, a partir dos mesmos dados. Além de probabilidades de sobrevivência por idade, também serão estimadas probabilidades de sobrevivência segundo a escolaridade das mães. A versão do método escolhida foi a apresentada por Brass e Hill, 1973, seguindo as indicações do Manual X das Nações Unidas (United Nations, 1983).

6.1.1. Estimativas de mortalidade adulta por idade

O primeiro passo a seguir para aplicar o método de orfandade é o cálculo das idades à maternidade. Nesta tese foi utilizada a idade à maternidade derivada dos dados da PPV e da PNAD, diretamente. Assim, estas idades se referem à fecundidade das entrevistadas durante o ano anterior à entrevista. As estimativas são de 25,35 anos para a PPV e de 26,27 anos para a PNAD.

Na TAB. 5 são observados os dados referentes à proporção de órfãos, tanto da PPV, como da PNAD. As estimativas das probabilidades de sobrevivência são apresentadas apenas a partir dos 20 anos de idade do entrevistado, com o objetivo de reduzir o viés provocado pela prática de adoção entre os entrevistados mais novos (Preston *et al.*, 2001).

TABELA 5. Brasil - Estimativas de probabilidades condicionadas de sobreviver de mulheres, através do método de orfandade, PPV e PNAD, 1996

Idade	n	PPV					PNAD				
		Não orfãos (mãe viva)	Orfãos (mãe morta)	S(n)	W(n)	l_{25+n}/l_{25}	Não orfãos (mãe viva)	Orfãos (mãe morta)	S(n)	W(n)	l_{25+n}/l_{25}
0-4	0	10.473.939	47.037				14.741.280	70.277			
5-9	5	10.088.517	185.059				15.612.192	207.314			
10-14	10	11.458.483	289.878				16.992.856	371.936			
15-19	15	10.710.158	473.689	0,95765	0,63880		15.662.700	622.765	0,96176	0,69441	
20-24	20	8.459.405	780.997	0,91548	0,70178	0,94507	12.288.483	829.655	0,93676	0,77843	0,95622
25-29	25	6.802.488	1.033.780	0,86808	0,74041	0,90317	10.449.985	1.269.320	0,89169	0,83745	0,92943
30-34	30	5.924.166	1.552.591	0,79234	0,75169	0,84927	10.098.782	1.907.947	0,84109	0,86765	0,88499
35-39	35	5.093.880	1.724.871	0,74704	0,75058	0,78104	8.212.040	2.493.574	0,76708	0,88284	0,83242
40-44	40	4.233.361	2.257.259	0,65223	0,68583	0,71725	6.216.987	3.161.296	0,66291	0,83449	0,74984
45-49	45	2.932.255	2.261.984	0,56452	0,59499	0,61671	4.180.043	3.524.686	0,54253	0,75614	0,63356
50-54	50	1.478.672	2.403.156	0,38092	0,39641	0,45370	2.446.964	3.515.970	0,41036	0,56461	0,48498
55-59	55	805.812	2.684.419	0,23088	0,15960	0,25482	1.358.210	3.545.570	0,27697	0,32088	0,31977
60-64	60	440.806	2.213.011		-0,03650		629.815	3.570.665		0,09868	
65-69	65	147.927	2.158.937				242.064	3.076.571			
70-74	70	35.970	1.592.759				71.747	2.275.991			
75-79	75	48.300	1.001.818				21.231	1.502.169			
80 e mais	80	0	1.045.005				15.124	1.412.419			

Fonte: IBGE, PPV 1996, PNAD 1996

Em seguida, é aplicada a fórmula proposta por Brass e Bamgboye (1981) para alocar, no tempo, cada uma das probabilidades de sobrevivência. Desta forma, o nível de mortalidade de cada probabilidade condicional vai representar o nível de mortalidade de um ano específico. O nível de mortalidade que representa a probabilidade condicional pode ser determinado interpolando duas tabelas de vida modelo que apresentem uma probabilidade condicional similar à da probabilidade estimada. Neste caso, a família de tabelas modelo escolhida é o padrão América Latina das Nações Unidas feminino (Nações Unidas, 1983). Na TAB. 6, são apresentadas as probabilidades condicionais estimadas com a PPV e PNAD, além dos respectivos anos de referência das estimativas, e as esperanças de vida ao nascer interpoladas a partir do padrão de mortalidade¹⁰ escolhido.

Como era de se esperar, tanto na PPV, como na PNAD, as probabilidades condicionais referentes a períodos mais recentes representam níveis de mortalidade mais baixos. Contudo, no caso da PPV, a tendência de diminuição da esperança de vida ao nascer correspondente às probabilidades de sobrevivência de anos mais distantes no tempo é interrompida a partir da probabilidade condicional l_{65}/l_{25} , a partir da qual, as probabilidades condicionais passam a corresponder a uma maior e_0 do que referentes às idades mais jovens. Provavelmente, este resultado é consequência do tamanho reduzido da amostra da PPV, que se acentua nas idades mais velhas, o que pode levar a vieses nas estimativas. Ao comparar os resultados da PPV e da PNAD, nota-se que as esperanças de vida ao nascer da PNAD são maiores que as da PPV em todos os grupos de idade. Na PNAD, a esperança de vida ao nascer estimada varia entre 63 e 69 anos, enquanto na PPV, ela varia de 60 a 66 anos. Embora seja uma diferença importante, é preciso notar que a amostra da PNAD inclui todo o Brasil, e a da PPV apenas as regiões Nordeste e Sudeste. As regiões faltantes na PPV, Sul, Centro-Oeste e Norte, apresentam, historicamente, uma mortalidade inferior à das

¹⁰ O padrão América Latina foi escolhido por se esperar que tenha uma estrutura de mortalidade por idade mais próxima daquela do Brasil. Contudo, o mesmo exercício foi realizado com outros padrões de tabelas modelo e os níveis de mortalidade estimados para cada probabilidade condicional foram muito semelhantes aos encontrados com o padrão América Latina. Seguramente esta similaridade se deve a que ao trabalhar com mortalidade adulta, o padrão não varia excessivamente de um modelo a outro.

regiões Sudeste e principalmente Nordeste, o que pode ser a causa desta diferença de nível entre as duas amostras (Carvalho, 1974; Carvalho e Pinheiro, 1986).

Estas estimativas podem ser comparadas com outras realizadas anteriormente para períodos próximos aos anos de referência estimados. Carvalho e Pinheiro (1986), por exemplo, estimaram uma esperança de vida ao nascer feminina, para o período de 1970 a 1980, de 63,29 anos, e em IBGE (1981) a estimativa, para o período de 1975-80, foi de 65,5 anos. Essas estimativas podem ser consideradas próximas às das estimativas encontradas tanto para a PPV, como para a PNAD, dando maior confiabilidade aos resultados encontrados.

TABELA 6. Brasil - Estimativas das probabilidades condicionais de sobreviver das mulheres pelo método de orfandade, ano de referência e esperança de via ao nascer

Base	Idade (n)	l_{25+n}/l_{25}	Ano de referência	e_0 interpolada de tabelas modelo
PPV				
	20	0,94507	1986,9	66,40
	25	0,90317	1984,4	63,49
	30	0,84927	1982,0	61,35
	35	0,78104	1979,5	60,07
	40	0,71725	1977,1	61,52
	45	0,61671	1974,6	62,12
PNAD				
	20	0,95622	1987,0	69,11
	25	0,92943	1984,6	68,05
	30	0,88499	1982,2	66,09
	35	0,83242	1979,7	65,58
	40	0,74984	1977,2	64,45
	45	0,63356	1974,7	63,45

Fonte: IBGE, PPV 1996, PNAD 1996

6.1.2. Estimativas de mortalidade por nível de escolaridade

No caso da aplicação do método de orfandade, segundo subgrupos educacionais, foi preciso calcular uma idade média à maternidade para cada grupo de escolaridade das

mães. Para isso, foram utilizados os Censos de 1970, 1980 e 1991 do Brasil, em vez das pesquisas amostrais da PPV e da PNAD.

Através da informação dos censos, é possível calcular as idades médias à maternidade em função do nível de escolaridade das mães e para períodos mais próximos ao momento em que as mães dos entrevistados na PPV e na PNAD tiveram seus filhos (TAB 7). Para aquelas mães de filhos nascidos antes de 1975, foi aplicada a idade à maternidade do Censo de 1970; para aquelas mães de filhos nascidos entre 1975 e 1985, foi aplicada a estrutura de fecundidade do Censo de 1980 e; para todos os outros entrevistados nascidos posteriormente a 1985, foi estimada a idade à maternidade correspondente ao Censo de 1991.

As idades à maternidade das mulheres sem escolaridade são sempre maiores que as das mulheres com alguma escolaridade, independente do ano considerado. Por outro lado, nota-se um rejuvenescimento da idade média à maternidade das mães sem escolaridade e das mães com 1 a 8 anos de estudo, e um envelhecimento da idade média à maternidade das mães com 9 anos e mais de estudo.

TABELA 7. Brasil - Idades à maternidade das mulheres por nível de escolaridade das mães, 1970, 1980, 1991

Ano do Censo	Escolaridade da mãe (anos de estudo)		
	sem escolaridade	1 a 8 anos	9 e mais
<i>1970</i>	29,13	27,36	26,86
<i>1980</i>	28,85	26,17	26,90
<i>1991</i>	28,31	25,31	27,09

Fonte: IBGE, Censos demográficos Brasil, 1970, 1980, 1991

O segundo passo para obter as estimativas por escolaridade consistiu em calcular a proporção de órfãos segundo a escolaridade das mães. As proporções foram calculadas utilizando-se a variável educação, antes da imputação e depois de cada uma das quatro imputações dos dados faltantes. A variável de escolaridade imputada escolhida para mostrar os resultados foi a resultante do método de imputação múltipla, considerada a que está melhor ajustada no Capítulo 5. Para evitar o problema de seleção da distribuição de dados faltantes na amostra da PNAD, apenas foram considerados os entrevistados com 40 anos e mais para estimar as

probabilidades de sobrevivência. A amostra da PPV continuou com os entrevistados de 20 anos e mais, como foi realizado anteriormente quando se estimou a mortalidade, sem separar por nível de escolaridade. As proporções e estimativas das probabilidades condicionadas podem ser observadas no Anexo (TABs. A 6 a A 13).

A alocação no tempo de cada uma das probabilidades foi realizada, utilizando-se o método proposto por Brass e Bamgboye (1981). Como no caso das estimativas de mortalidade por idade, determinou-se um nível de mortalidade para cada uma das probabilidades condicionais de sobrevivência com base no padrão América Latina das Nações Unidas (Nações Unidas, 1983). Na TAB. 8 podem ser observados as probabilidades condicionais, assim como os seus correspondentes anos de referência de cada uma das probabilidades e as esperanças de vida representativas do nível de mortalidade. No caso da PPV, talvez pela amostra ser pequena, as estimativas das probabilidades condicionais de sobrevivência, e conseqüentemente, das esperanças de vida ao nascer, não seguem um padrão claro por idade, dentro de cada grupo de escolaridade. De qualquer forma, as estimativas de esperança de vida das mulheres sem escolaridade variam entre 58 e 63 anos, as das mulheres com 1 a 8 anos de estudo de 63 a 71 anos, e as das mulheres com 9 anos de estudo ou mais, de 69 a 76 anos, demonstrando que há importantes diferenças na mortalidade dos grupos por escolaridade. Além disso, ao comparar as estimativas de esperança de vida ao nascer correspondente à probabilidade condicional de sobrevivência entre as mesmas idades, observa-se que as mulheres com maior escolaridade apresentam, sempre, uma esperança de vida maior do que a das mulheres com menor escolaridade.

No caso da PNAD, ao restringir a amostra aos entrevistados com 40 anos e mais, as probabilidades condicionais apenas foram estimadas para os parâmetros l_{65}/l_{25} e l_{70}/l_{25} . Como no caso das estimativas de mortalidade adulta por idade, os níveis de mortalidade da PNAD são sempre inferiores aos da PPV. A relação positiva entre esperança de vida e escolaridade também é encontrada na PNAD, onde as mulheres sem escolaridade teriam uma esperança de vida ao nascer entre 65 e 66 anos, as mulheres com 1 a 8 anos de estudo teriam uma esperança de vida levemente superior a 67 anos e as mulheres com 9 anos de estudo e mais, entre 73 e 74 anos.

TABELA 8. Brasil - Probabilidades condicionais de sobreviver das mulheres estimadas pelo método de orfandade por nível de escolaridade, ano de referência e esperança de vida ao nascer, PNAD e PPV, 1996

Escolaridade das mães	Idade (n)	PPV			PNAD		
		l_{25+n} / l_{25}	Ano de referência	e_0 da tabela modelo	l_{25+n} / l_{25}	Ano de referência	e_0 da tabela modelo
Sem escolaridade	20	0,91737	1986,9	60,57			
	25	0,90164	1984,4	63,24			
	30	0,86414	1981,9	63,28			
	35	0,76622	1979,4	58,53			
	40	0,72835	1977,1	62,52	0,76726	1977,3	66,02
	45	0,57419	1974,9	58,67	0,65836	1974,9	65,38
1 a 8 anos de estudo	20	0,95902	1987,0	69,84			
	25	0,91824	1984,5	66,03			
	30	0,87916	1982,2	65,29			
	35	0,81256	1980,1	63,42			
	40	0,79600	1977,8	68,63	0,78518	1977,5	67,64
	45	0,74086	1975,3	71,57	0,68307	1975,2	67,27
9 anos e mais de estudo	20	0,97943	1987,1	76,06			
	25	0,95596	1984,8	73,44			
	30	0,93281	1982,5	73,29			
	35	0,90740	1980,1	73,63			
	40	0,80008	1978,2	69,00	0,84841	1978,1	73,46
	45	0,76483	1975,3	72,67	0,78133	1975,8	74,53

Fonte: IBGE, PPV 1996, PNAD 1996, Nações Unidas (1983)

6.2. ESTIMATIVAS DE MORTALIDADE ADULTA COM BASE NO MÉTODO ESTOCÁSTICO

Nesta seção, será aplicada a metodologia estocástica desenvolvida nesta Tese, o método estocástico, para estimar medidas de mortalidade adulta das mulheres no Brasil, por idade e por nível de escolaridade. Como já foi explicado anteriormente e para obter resultados de fontes diferentes que possam dar maior consistência à metodologia, além de pesquisar como que o método lida com cada uma das fontes, serão usadas duas bases de dados, a PPV e a PNAD 1996. A aplicação principal da metodologia será realizada com a PPV, já que a proporção de dados faltantes na variável escolaridade é muito inferior (3,4%) àquela da PNAD (16,55%). Com o objetivo de reduzir os problemas derivados da adoção de crianças e da sobre-representação de mães com maior fecundidade, a amostra foi restringida àqueles

indivíduos com 15 anos e mais, no caso da PPV¹¹. No caso da PNAD, o problema principal é derivado de uma coleta desigual da informação de escolaridade das mães, que leva a que todos aqueles indivíduos que não são chefes de domicílio e têm a mãe morta não respondam sobre a escolaridade dela. Para minimizar os possíveis vieses resultantes desta distribuição de dados faltantes, a amostra da PNAD foi restringida àqueles indivíduos com 40 anos e mais de idade, onde a proporção de chefes de domicílio é bem maior. Diante dessa restrição de idade, os mesmos modelos foram estimados para uma amostra da PPV de indivíduos de 40 anos e mais, podendo-se comparar, assim, os resultados das duas bases de dados. Como, a partir de idades muito avançadas, podem existir mais problemas de seletividade (Hoffman, 2005b), aqueles indivíduos com 80 anos e mais também foram retirados de todas as amostras.

Seguindo o exposto no capítulo metodológico, o primeiro passo é estimar o período de exposição à morte de cada mãe, através da estimativa das idades de início e final desse período. No caso da idade inicial, é preciso sortear a idade à maternidade das progenitoras de todos aqueles indivíduos com mãe morta, e daqueles com idade da mãe não conhecida, por morar fora do domicílio do filho). A realização do sorteio da idade à maternidade é condicionada à estrutura de fecundidade da década de nascimento do entrevistado. As funções de fecundidade utilizadas foram as estimadas por Horta, Carvalho & Frias (2000) para o período de 1930 até 1990¹². Para todos os indivíduos nascidos antes de 1930, utilizaram-se as taxas de fecundidade de 1930-35, assumindo que a estrutura da função manteve-se constante antes desse período.

Uma vez estimada a idade inicial do período de exposição ao risco de morrer, o passo seguinte é estimar a idade final. Esta é resultado de processos diferentes, dependendo se a mãe está morta ou viva. Caso esteja viva, a idade final é calculada pela soma da idade inicial (à maternidade) mais a idade do filho. Se ela está morta, deve ser realizado um outro sorteio, condicionado a uma distribuição de mortalidade. Neste

¹¹ Restringir a idade dos entrevistados provoca seleção na amostra de mães analisadas, o que produz viés nas estimativas, especialmente entre as mães mais jovens. Contudo, espera-se que o viés produzido por esta redução da amostra compense o viés que pode provocar a sobre-representação de mães com mais de um filho entrevistado. Os resultados para a amostra completa da PPV também podem ser contemplados nos Anexos, na TAB. A 26.

¹² Estas funções de fecundidade e aquelas ponderadas pela mortalidade podem ser encontradas nos Anexos (TAB. A 16).

trabalho, foram utilizadas três distribuições de mortalidade: a de 1965 (Carvalho, 1974), a de 1975 (Carvalho e Pinheiro, 1986), e a de 1985 (IBGE, 1991), para o sorteio da idade à morte. Utilizou-se a distribuição de mortalidade de 1965 para os casos em que o ponto médio do período de risco é anterior a 1970; a distribuição de 1975, quando o período de referência estava compreendido entre 1970 e 1980; e, finalmente, a estrutura de mortalidade de 1978, para os casos em que o ponto médio correspondeu a qualquer ano posterior a 1980. Graças a diversas técnicas existentes¹³, os dados de mortalidade, correspondentes a grupo etários quinquenais, podem ser desagregados por idade simples, dando maior flexibilidade ao sorteio.

Com as idades inicial e final estimadas, o período de exposição ao risco de morrer de cada mãe é calculado, automaticamente, como a diferença entre as duas idades. Para estimar taxas de mortalidade a partir desses dados, o período de exposição deve ser transformado em pessoas-ano, utilizando a função SPLIT do Stata 10.0. A partir daqui, o número de mortes das mães pode ser modelado em função do tempo de exposição, através de uma regressão Poisson ou uma binomial negativa, estimando taxas de mortalidade segundo a idade, a escolaridade das mães e, até, segundo a coorte de nascimento. A escolha entre a regressão Poisson e a regressão binomial negativa foi realizada através de um teste de sobredispersão dos valores preditos, disponível no programa Stata 10.0. Esse teste foi realizado para todos os modelos e não foi significativo, indicando que, neste caso, a regressão Poisson é mais adequada para modelar os dados de mortalidade do que a regressão binomial negativa.

Os modelos de regressão Poisson, estimados a seguir, já foram apresentados, de forma geral, no capítulo da metodologia estocástica. No entanto, as especificações das variáveis utilizadas são explicadas aqui com maior detalhe.

O primeiro modelo inclui unicamente a variável de idade da mãe, categorizada em grupos decenais, desde os 30 até os 69 anos, totalizando quatro grupos. Esse modelo é

¹³ Foi empregada a técnica de Karup-King (Shryock e Siegel, 1973) apresentada na TAB. A 16 do Anexo, para distribuir por idades simples as taxas de mortalidade quinquenais das funções de 1965, 1975 e 1985. Os pesos desta técnica se encontram na TAB. 32 do Anexo. Além disso, através de um modelo logito e do padrão de Himes *et al.*(1994) (TAB. 15 dos Anexos), a taxa do grupo de idade final, tipicamente de 80 anos e mais, foram estendidas até a idade de 100 anos e mais. As tabelas de sobrevivência escolhidas para a imputação da idade de morte e as tabelas de óbitos acumulados derivadas delas estão apresentadas nas TAB. A 17 a A 20 do Anexo.

essencial para comprovar se as estimativas de mortalidade encontradas são consistentes com as estimativas precedentes de outros autores, para o mesmo período.

O segundo modelo consiste na estimativa do número de mortes em função da idade e da escolaridade. A variável idade foi codificada, como no modelo anterior. A escolaridade das mães foi categorizada em três grupos, aquelas sem escolaridade, as que tinham de 1 a 8 anos de estudo e as mães com 9 anos de estudo ou mais. Quando a variável escolaridade incluiu uma categoria *dummy* para os valores faltantes, o número de categorias aumentou para quatro.

O terceiro modelo adiciona ao segundo a interação entre a idade e a escolaridade. Esse modelo vai permitir estimar taxas de mortalidade por nível de escolaridade, dentro de cada grupo de idade.

Por último, e unicamente com os dados da PPV, foram realizadas algumas regressões Poisson, adicionando um controle de coorte de nascimento aos modelos de idade e escolaridade. Por ser essa base de dados muito pequena, o número de categorias da variável idade teve que ser reduzido a duas, com amplitude de 20 anos cada uma (30 a 49, 50 a 69), procurando ampliar a significância dos coeficientes e do modelo. A variável coorte foi categorizada em três grupos: as mulheres nascidas em 1926 ou anteriormente, as nascidas entre 1927 e 1946 e as nascidas entre 1947 e 1966 (ano limite de nascimento, já que a idade mais nova incluída no modelo é de 30 anos e o ano de entrevista foi 1996).

Todos os coeficientes apresentados são derivados de um procedimento *bootstrap*, resultado de múltiplas gerações aleatórias¹⁴ de períodos de exposição ao risco de morrer, para cada mãe. Além disso, os modelos de regressão Poisson, que incluem a variável escolaridade, foram realizados para cada uma das quatro imputações de dados faltantes. No entanto, os resultados apresentados aqui estão limitados aos modelos com a variável escolaridade imputada através da imputação múltipla. Esse tipo de imputação foi escolhido por ser o que respeita a distribuição categórica da variável escolaridade e por levar em conta, em maior medida, a incerteza que um

¹⁴ Achou-se que 500 vezes era um número suficiente e adequado para dar consistência às estimativas.

modelo de não resposta representa. Os outros resultados com as outras variáveis de escolaridade imputadas e sem imputar estão disponíveis no Anexo (TABs. A 19 a 23).

6.2.1. Estimativas de mortalidade adulta por idade

São analisados aqui todos os resultados referentes às regressões Poisson por idade, utilizando cada uma das amostras. Os resultados das regressões Poisson referem-se a coeficientes *bootstrap*. Porém, antes, apresenta-se uma tabela descritiva (TAB. 9) com o número de mortes, o número de pessoas-ano, e as taxas de mortalidade por grupos de idade para três amostras de dados, duas a partir da PPV (entrevistados de 15 anos e mais e de 40 anos e mais), e uma a partir da PNAD (entrevistados de 40 anos e mais), todas elas vindo de uma única geração aleatória de idades de início e fim do período de exposição ao risco de morrer.

TABELA 9. Brasil - Número de mortes, pessoas-ano e taxas de mortalidade por grupos de idade, para mulheres de diversas amostras, PPV e PNAD, 1996

Variável	PPV (15 anos e mais)			PNAD (40 anos e mais)			PPV (40 anos e mais)		
	Mortes	Pessoas - Ano	Taxa (x1000)	Mortes	Pessoas - Ano	Taxa (x1000)	Mortes	Pessoas - Ano	Taxa (x1000)
Idade									
30 a 39	207	90.864	2,28	1.853	584.960	3,17	115	35.137	3,27
40 a 49	428	100.065	4,28	4.048	774.455	5,23	247	46.335	5,33
50 a 59	627	75.282	8,33	7.152	753.377	9,49	393	45.037	8,73
60 a 69	871	45.458	19,16	12.278	617.390	19,89	758	37.128	20,42

Fonte: IBGE, PPV 1996, PNAD 1996

Como esperado, as taxas indicam que, quanto maior a idade, maior é a taxa de mortalidade. Além disso, comparando-se as amostras pode-se ver que a mortalidade das mães dos entrevistados com 40 anos e mais é maior do que a das mães dos entrevistados de 15 anos e mais. Essa diferença se deve ao fato de que as mães dos entrevistados com 40 anos e mais morreram, em média, em períodos mais distantes, quando o nível de mortalidade era mais alto.

Uma vez calculado o número de pessoas-ano das múltiplas bases de dados, resultado da geração aleatória de múltiplos períodos de exposição ao risco, é possível estimar os coeficientes *bootstrap* das regressões Poisson, controlando pela idade das mulheres.

Os resultados estão expostos na TAB. 10. Como pode ser observado, todos os coeficientes, independentemente da amostra, foram estatisticamente significativos ($P < 0.001$) e crescentes com o aumento da idade. Isso significa, como esperado, que aquelas mulheres com maior idade apresentavam maiores taxas de mortalidade. É importante chamar a atenção para o tamanho das amostras. Como já discutido, tanto no capítulo de bases de dados, como na seção sobre os resultados do método de orfandade, a amostra da PPV totaliza um número inferior de pessoas-ano do que a PNAD. Esta é uma das explicações para o maior valor dos desvios padrão dos coeficientes estimados pela PPV, comparativamente a PNAD. De qualquer forma, os resultados da PNAD e da PPV com a amostra de 40 anos e mais são bastante semelhantes.

TABELA 10. Brasil - Coeficientes da regressão Poisson do número de mortes de mulheres em função da idade, PPV e PNAD 1996

Variável	PPV (15 anos e mais)		PNAD (40 anos e mais)		PPV (40 anos e mais)	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade						
30 a 39	-		-		-	
40 a 49	0,645	(0,079) ***	0,517	(0,028) ***	0,515	(0,115) ***
50 a 59	1,328	(0,075) ***	1,132	(0,026) ***	1,134	(0,106) ***
60 a 69	2,085	(0,073) ***	1,865	(0,025) ***	1,868	(0,102) ***
constante	-6,075	(0,065) ***	-5,772	(0,023) ***	-5,763	(0,095) ***
N	38209		307037		18364	

*** $P > 0,01$; ** $P > 0,05$; * $P > 0,1$

Fonte: IBGE, PPV 1996, PNAD 1996

Para conhecer com exatidão o nível de mortalidade ao que se referem os coeficientes, é preciso calcular as taxas de mortalidade através do inverso do logaritmo natural dos coeficientes de cada categoria. A partir dessas taxas, utilizou-se o sistema logito para ajustar uma tabela-modelo de sobrevivência¹⁵ e obter uma distribuição de mortalidade

¹⁵ O ajuste foi realizado obtendo o logito de cada uma das taxas de mortalidade por idade estimadas e de uma tabela-modelo do padrão da América Latina que estivesse mais próximo do nível de mortalidade observado. Através dos logitos se estimou uma regressão linear entre as duas estruturas a partir da qual se obteve os parâmetros α e β para ajustar a tabela-modelo.

representativa das estimativas que permitisse calcular uma esperança de vida ao nascer e aos 30 anos. O padrão de mortalidade utilizado foi o das mulheres da América Latina (Nações Unidas, 1983) como no caso do método de orfandade. Os resultados podem ser observados na TAB. 11.

TABELA 11. Brasil - Taxas de mortalidade específicas de mulheres por idade (x 1000) estimadas através do método estocástico e esperanças de vida ao nascer e aos 30 anos, PPV e PNAD, 1996

Base	Idade	Taxas			e_0			e_{30}		
			Intervalo			Intervalo			Intervalo	
PPV (15 anos e mais)	30 a 39 anos	0,00230	0,00201	0,00263	69,16	74,27	61,86	46,12	49,72	41,23
	40 a 49 anos	0,00439	0,00325	0,00593						
	50 a 59 anos	0,00868	0,00648	0,01162						
	60 a 69 anos	0,01850	0,01388	0,02466						
PNAD (40 anos e mais)	30 a 39 anos	0,00311	0,00297	0,00326	67,47	69,15	61,42	45,63	46,44	39,59
	40 a 49 anos	0,00522	0,00472	0,00578						
	50 a 59 anos	0,00965	0,00876	0,01064						
	60 a 69 anos	0,02009	0,01827	0,02209						
PPV (40 anos e mais)	30 a 39 anos	0,00314	0,00261	0,00378	66,57	70,96	60,50	45,09	47,30	42,28
	40 a 49 anos	0,00526	0,00349	0,00793						
	50 a 59 anos	0,00976	0,00658	0,01448						
	60 a 69 anos	0,02034	0,01383	0,02992						

Fonte: IBGE, PPV 1996, PNAD 1996, Nações Unidas (1983)

Como já observado na TAB. 9, a amostra da PPV, que inclui todos os indivíduos de 15 anos ou mais, apresentou um nível de mortalidade inferior ao dos outros casos, onde a amostra era de 40 anos e mais. A esperança de vida aos 30 anos estimada foi de 46,12 anos, para a amostra de 15 anos e mais da PPV, em comparação aos 45,6 e 45,1 anos das amostras da PNAD e PPV com entrevistados de 40 anos e mais. A diferença é explicada pela inclusão de mães mais novas na amostra de 15 anos e mais da PPV, que, além de mais jovens, estiveram expostas ao risco de morte em um período mais recente em média, onde a mortalidade era menor. Para obter uma medida do período de referência para a mortalidade em cada amostra, estimou-se a mediana de todos os pontos médios dos períodos de exposição à morte das mulheres. No caso da PPV com a amostra de entrevistados de 15 anos e mais, este ponto é o ano 1979. Para a PNAD e PPV com entrevistados de 40 anos e mais, o ano de referência é 1965. Os resultados seguem as tendências esperadas e coincidem com estimativas

apresentadas em trabalhos anteriores. A esperança de vida aos 30, com base na amostra da PPV com entrevistados de 15 anos e mais foi de 46,12 anos, com um intervalo de confiança de 41,23 a 49,72 anos. Ao comparar este resultado com a esperança de vida aos 30 anos estimada por Carvalho e Pinheiro (1986) para a década 1970/1980, de 42,2 anos, observa-se que esta última está dentro do intervalo de confiança estimado pelo método estocástico. Da mesma forma, as esperanças de vida estimadas aos 30 anos da PNAD e da PPV com as amostras de 40 anos e mais, foram, respectivamente, de 45,63 (intervalo de confiança: 39,59 a 46,44 anos), e de 45,09 (intervalo de confiança: 42,38 a 45,97 anos), que inclui, no primeiro caso, a esperança de vida de 40,4 anos, estimada por Carvalho (1976) para a década de 1960/1970. Diante disso, pode-se dizer que as estimativas de mortalidade por idade encontradas através do método estocástico são consistentes (pelo menos no caso da amostra da PPV com 15 anos e mais e da PNAD com 40 anos e mais) com outras estimativas realizadas para os mesmos períodos de referência.

6.2.1. Estimativas de mortalidade adulta por nível de escolaridade

Nesta seção, são apresentadas as estimativas de mortalidade adulta, por nível de escolaridade. Serão discutidos apenas os resultados referentes aos modelos que incorporam a variável escolaridade imputada através da imputação múltipla. Os resultados referentes às outras imputações podem ser encontrados nos anexos (TABs. A 19 a A 23). Vale dizer que os coeficientes calculados com base nas outras imputações são consistentes.

Antes de apresentar os resultados das regressões Poisson, a TAB. 12 expõe o número de mortes, pessoas-ano e o cálculo das taxas de mortalidade segundo o nível de escolaridade das mulheres¹⁶. As taxas de mortalidade indicam, como esperado, que mulheres com maior escolaridade tem menor mortalidade, comparativamente às mulheres sem escolaridade.

¹⁶ Os dados apresentados são o resultado de uma única geração aleatória das idades de início e fim do período de exposição ao risco de morrer como no caso da TAB. 9. Não são a média das 500 gerações aleatórias que são utilizadas para estimar os coeficientes *bootstrap*.

TABELA 12. Brasil - Número de mortes, pessoas-ano e taxas de mortalidade brutas, por grupos de idade e de escolaridade para mulheres, PPV, PNAD, 1996

Variável	PPV (15 anos e mais)			PNAD (40 anos e mais)			PPV (40 anos e mais)		
	Mortes	Pessoas - Ano	Taxa (x1000)	Mortes	Pessoas - Ano	Taxa (x1000)	Mortes	Pessoas - Ano	Taxa (x1000)
Escolaridade									
Sem escolaridade	2.159	177994	12,13	34.003	1983823	17,14	2.088	119995	17,40
1 a 8 anos de estudo	1.271	172661	7,36	18.078	1361408	13,28	1.050	78748	13,33
9 anos e mais de estudo	124	29931	4,14	1.333	124740	10,69	96	9205	10,43

Fonte: IBGE, PPV 1996, PNAD 1996

Na TAB. 13, apresentam-se os coeficientes do modelo de regressão Poisson, incorporando a variável escolaridade, além da idade. Mais uma vez, todos os coeficientes (idade e escolaridade) são significativos com um nível de confiança de 1%. Em comparação com o modelo que controla apenas por idade (TAB. 10), há uma redução no tamanho dos coeficientes dos grupos de idade, neste segundo modelo. Esta mudança pode ser explicada pelo fato das mulheres mais jovens terem também maior escolaridade. Em relação aos coeficientes dos grupos de escolaridade, eles mostram a existência de uma clara relação negativa entre os anos de estudo e a mortalidade¹⁷.

¹⁷ Cabe mencionar que na regressão Poisson, onde os dados faltantes da variável escolaridade não foram imputados, mas agrupados em uma categoria *dummy*, esta categoria apresentou uma mortalidade expressivamente superior aos outros grupos, tal como a análise descritiva das bases de dados indicaram. Estes resultados podem ser observados nas TABs. A 21 a A 26 do Anexo.

TABELA 13. Brasil - Coeficientes da regressão Poisson do número de mortes de mulheres em função da idade e da escolaridade, PPV, PNAD 1996

Variável	PPV (15 anos e mais)		PNAD (40 anos e mais)		PPV (40 anos e mais)	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade						
30 a 39 anos	-		-		-	
40 a 49 anos	0,627	0,084 ***	0,516	0,028 ***	0,514	0,115 ***
50 a 59 anos	1,292	0,080 ***	1,132	0,026 ***	1,134	0,106 ***
60 a 69 anos	2,037	0,078 ***	1,865	0,025 ***	1,869	0,102 ***
Escolaridade						
Sem escolaridade	-		-		-	
1 a 8 anos de estudo	-0,340	0,045 ***	-0,196	0,013 ***	-0,249	0,054 ***
9 anos e + de estudo	-0,772	0,115 ***	-0,437	0,040 ***	-0,532	0,152 ***
Constante	-5,863	0,072 ***	-5,685	0,024 ***	-5,655	0,097 ***
N	38209		307037		18364	

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996, PNAD 1996

As taxas de mortalidade estimadas a partir dos coeficientes acima são apresentadas na TAB. 14. Os dados indicam que as taxas aumentam com a idade e diminuem com a escolaridade, dentro de cada um dos grupos de idade. O modelo sem a interação entre as variáveis idade e escolaridade, não permite, no entanto, avaliar como os diferenciais de escolaridade variam dentro de cada grupo de idade. Apesar disso, estas taxas permitem observar qual é o diferencial relativo entre os grupos de escolaridade, em média, para todos os grupos de idade. Na amostra da PPV com os entrevistados de 15 anos e mais a taxa de mortalidade das mulheres sem escolaridade é 2,16 vezes maior da taxa das mulheres com 9 anos de estudo e mais. Este diferencial também é expressivo, apesar de inferior, para as amostras da PNAD e da PPV com entrevistados de 40 anos e mais, onde as taxas de mortalidade das mais escolarizadas são 1,5 e 1,7 vezes superiores à das taxas das mais escolarizadas, respectivamente. O diferencial também existe entre as mulheres do grupo intermédio de escolaridade (com 1 a 8 anos de estudo) e os outros dois grupos (sem escolaridade e com 9 anos e mais de estudo), porém é inferior ao encontrado entre os grupos extremos.

As diferenças entre as amostras indicariam que quando o ano de referência da amostra é mais recente (1979 para a PPV com entrevistados de 15 anos e mais), os diferenciais

de mortalidade por escolaridade são superiores aos encontrados para as amostras com um ano de referência mais distante (1965 para as amostras com entrevistados de 40 anos e mais).

TABELA 14. Brasil - Taxas de mortalidade específicas (x 1000) de mulheres em função da idade e da escolaridade, PPV e PNAD, 1996

Variáveis		PPV 15 anos e mais	PNAD 40 anos e mais	PPV 40 anos e mais
Idade	Escolaridade			
30 a 39 anos	Sem escolaridade	2,843	3,395	3,500
	1 a 8 anos de estudo	2,024	2,790	2,729
	9 anos e mais de estudo	1,313	2,192	2,055
40 a 49 anos	Sem escolaridade	5,320	5,690	5,849
	1 a 8 anos de estudo	3,787	4,677	4,561
	9 anos e mais de estudo	2,457	3,674	3,435
50 a 59 anos	Sem escolaridade	10,352	10,525	10,873
	1 a 8 anos de estudo	7,369	8,651	8,478
	9 anos e mais de estudo	4,782	6,797	6,385
60 a 69 anos	Sem escolaridade	21,803	21,912	22,679
	1 a 8 anos de estudo	15,519	18,011	17,684
	9 anos e mais de estudo	10,072	14,150	13,317
Razão entre "Sem escolaridade" e "9 anos de estudo e mais"		2,165	1,549	1,703

Fonte: IBGE, PPV 1996, PNAD 1996

Diante desses resultados, parece importante introduzir no modelo de regressão a interação entre a idade e a escolaridade. As estimativas desse modelo são apresentadas na TAB. 15. Os coeficientes da idade e da escolaridade continuam significativos, à exceção da amostra da PPV com entrevistados de 40 anos e mais, onde os coeficientes da escolaridade perdem sua significância. No caso dos coeficientes das interações, estes não foram estatisticamente significativos em nenhum dos modelos. Portanto, a análise do valor desses coeficientes não é relevante. Este resultado pode estar indicando, na verdade, que o diferencial entre os grupos de escolaridade não muda em função da idade. Porém, e especialmente no caso da PPV,

a falta de significância pode ser causada pelo tamanho relativamente pequeno da amostra, que provoca uma falta de poder estatístico. Neste caso, a análise dos coeficientes continua sendo importante, porque pode apontar uma tendência não necessariamente errada. No caso da PPV, os coeficientes da interação foram positivos e crescentes com a idade e com o nível de escolaridade. Dado que os coeficientes dos grupos de escolaridade são negativos, esse tipo de resultado sugere que os diferenciais relativos de mortalidade entre grupos educacionais são menores entre os grupos de idades mais elevados do que entre as idades mais jovens. Já no caso da PNAD, o tamanho da amostra é suficientemente grande, porém, os coeficientes continuam não sendo significativos, indicando que o diferencial de mortalidade por nível de escolaridade praticamente não varia com a idade, apesar de existir um diferencial de mortalidade em função da idade e da escolaridade, separadamente. Embora os resultados das três amostras não possam ser diretamente comparados, parecem sugerir que o diferencial de mortalidade por nível de escolaridade é maior no período mais recente e diminui com a idade, enquanto nos períodos mais antigos, é menor e varia menos em função da idade.

TABELA 15. Brasil - Coeficientes da regressão Poisson do número de mortes de mulheres em função da idade, da escolaridade e da interação entre as duas variáveis, PPV e PNAD, 1996

Variável	PPV (15 anos e mais)		PNAD (40 anos e mais)		PPV (40 anos e mais)	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade						
30 a 39 anos						
40 a 49 anos	0,571	0,112 ***	0,504	0,036 ***	0,506	0,143 ***
50 a 59 anos	1,205	0,106 ***	1,115	0,033 ***	1,126	0,133 ***
60 a 69 anos	1,945	0,102 ***	1,849	0,032 ***	1,865	0,127 ***
Escolaridade						
Sem escolaridade						
1 a 8 anos de estudo	-0,482	0,144 ***	-0,235	0,050 ***	-0,274	0,205
9 anos e + de estudo	-1,213	0,370 ***	-0,465	0,151 **	-1,235	27,595
Idade x Escolaridade						
40 a 49 x 1 a 8 a. estudo	0,104	0,175	0,033	0,060	0,030	0,247
40 a 49 x 9 e + a. estudo	0,316	0,447	0,042	0,182	0,601	28,772
50 a 59 x 1 a 8 a. estudo	0,168	0,166	0,045	0,055	0,030	0,229
50 a 59 x 9 e + a. estudo	0,503	0,426	0,031	0,169	0,685	27,661
60 a 69 x 1 a 8 a. estudo	0,178	0,161	0,044	0,053	0,021	0,219
60 a 69 x 9 e + a. estudo	0,567	0,418	0,020	0,162	0,691	27,633
constante	-5,788	0,092 ***	-5,670	0,029 ***	-5,650	0,119 ***
N	38209		307037		18364	

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996, PNAD 1996

6.2.3. Estimativas de mortalidade por coorte de nascimento

Uma das várias vantagens do método estocástico é a construção de uma base de dados de formato longitudinal, que permite a decomposição de efeitos de idade, período e coorte. Nesta tese, em função da falta de poder estatístico das amostras, optou-se por referenciar as estimativas de mortalidade apenas por idade e coorte de nascimento. Para tal, foram estimados dois modelos de regressão Poisson: o primeiro controlando apenas por idade e coorte e o segundo, controlando por idade, coorte e escolaridade,

considerando apenas os efeitos puros de cada variável. A TAB. 16 apresenta os coeficientes *bootstrap* dos dois modelos, utilizando a amostra de entrevistados de 15 anos e mais da PPV.

Todos os coeficientes foram estatisticamente significativos e com o sentido esperado. Ou seja, o nível de mortalidade é maior para as coortes mais velhas e as mulheres mais idosas. Ao adicionar a variável de escolaridade no modelo, os coeficientes das coortes perdem força explicativa e estatística, mas continuam sendo significativos ($P < 0.05$). Isso pode indicar que parte da menor mortalidade das mulheres de coortes mais recentes se deve ao aumento da escolaridade. A escolaridade também apresentou o resultado esperado, sendo que as mulheres com mais escolaridade tinham menor mortalidade do que as mulheres com escolaridade inferior.

TABELA 16. Brasil - Coeficientes da regressão Poisson do número de mortes de mulheres em função da idade, da coorte e da escolaridade, PPV 1996

Variável	Coef.	D. P.	Coef.	D. P.
Idade				
30 a 49 anos	-		-	
50 a 69 anos	1,208	0,051 ***	1,209	0,051 ***
Coorte de nascimento				
1926 ou antes	-		-	
1927 a 1946	-0,164	0,047 ***	-0,106	0,047 *
1947 a 1966	-0,399	0,115 ***	-0,265	0,116 *
Escolaridade				
Sem escolaridade			-	
1 a 8 anos de estudo			-0,341	0,046 ***
9 anos e + de estudo			-0,775	0,116 ***
constante	-5,545	0,049 ***	-5,393	0,051 ***
N	38469		38469	

*** $P > 0,01$; ** $P > 0,05$; * $P > 0,1$

Fonte: IBGE, PPV 1996, PNAD 1996

A partir dos coeficientes desses modelos, é possível estimar as taxas de mortalidade correspondentes a cada grupo de idade, coorte de nascimento e escolaridade. Essas taxas são apresentadas na TAB. 17. O maior nível de mortalidade foi experimentado pelas mulheres nascidas em 1926 ou antes, e com idades entre 50 e 69 anos. A mortalidade diminui entre as coortes mais recentes. Pode ser observado a existência de um expressivo diferencial de mortalidade em função da escolaridade, onde as taxas de mortalidade das mulheres sem escolaridade é 2,2 vezes maiores do que as taxas das mulheres com 9 anos de estudo ou mais. Como o modelo não controla pela interação entre as variáveis, esse diferencial é representativo de todos os grupos de idade e de coorte. Por último, é interessante notar que os níveis de mortalidade das mulheres de 30 a 49 anos sem escolaridade e com 1 a 8 anos de estudo, da coorte mais recente (nascidas entre 1947 e 1966), são maiores do que o das mulheres com a mesma idade, nascidas antes de 1927 e com 9 anos de estudo ou mais. Apesar da mortalidade das coortes mais recentes ter caído substancialmente, as mulheres de 30 a 49 anos sem escolaridade e aquelas com 1 a 8 anos de estudo, nascidas entre 1947 e 1966, ainda morriam mais do que as mulheres com maior escolaridade, nascidas antes de 1946.

TABELA 17. Brasil - Taxas de mortalidade (x 1000) de mulheres segundo a idade, a coorte e a escolaridade, PPV

Coorte	Idade	Total	Por Escolaridade		
			Sem escolaridade	1 a 8 anos de estudo	9 anos de estudo ou mais
antes de 1926	30 a 49 anos	3,91	4,55	3,23	2,09
	50 a 69 anos	13,08	15,24	10,83	7,02
1927 a 1946	30 a 49 anos	3,32	4,09	2,91	1,88
	50 a 69 anos	11,10	13,70	9,74	6,31
1947 a 1966	30 a 49 anos	2,62	3,49	2,48	1,61
	50 a 69 anos	-	-	-	-

Fonte: IBGE, PPV 1996, PNAD 1996

Para examinar como o diferencial de mortalidade por educação varia entre as coortes, seria necessário controlar pela interação entre a idade, a coorte e a escolaridade. Infelizmente, os coeficientes dessas interações não são significativos, impedindo,

mais uma vez, que alguns dos desdobramentos do gradiente educacional em mortalidade fossem discutidos nesta tese.

6.3. Comparação dos resultados do método de orfandade com os do método estocástico

Finalmente, nesta seção, são comparadas as estimativas de mortalidade geradas através do método de orfandade e do método estocástico. Para tal, estimou-se a esperança de vida ao nascer a partir dos resultados de cada uma das metodologias, tendo, como padrão de mortalidade, aquele da América Latina, gerado pelas Nações Unidas (Nações Unidas, 1983)¹⁸. As amostras de dados escolhidas para a comparação são aquelas que mostraram resultados mais consistentes e, por tanto, mais confiáveis, neste caso, a amostra de indivíduos de 15 anos e mais da PPV e a amostra de 40 anos e mais da PNAD.

¹⁸ Os métodos para estimar as esperanças de vida com os resultados de cada um dos métodos de mortalidade por nível de escolaridade, é o mesmo que foi utilizado para estimar as esperanças de vida para todos os grupos de idade, mostrado anteriormente nesta Tese.

TABELA 18. Brasil – Esperanças de vida ao nascer femininas estimadas através do método de orfandade e do método estocástico, para toda a amostra e por nível de escolaridade, PNAD, PPV, 1996

	Escolaridade	Orfandade	Método estocástico		
		e_0	e_0	Intervalo	
PPV (15 anos e mais)	Total	61,52	69,16	74,27	61,86
	Sem escolaridade	62,52	65,45	68,43	62,47
	1 a 8 anos de estudo	68,63	70,41	73,37	67,44
	9 anos de estudo e mais	69,00	74,34	75,25	69,46
PNAD (40 anos e mais)	Total	63,45	67,47	69,15	61,42
	Sem escolaridade	65,38	63,45	65,45	61,48
	1 a 8 anos de estudo	67,27	67,44	68,43	64,45
	9 anos de estudo e mais	74,53	70,41	72,34	67,47

Fonte: IBGE, PPV, PNAD 1996

Na maioria dos grupos de idade e de escolaridade, pode-se dizer que as estimativas dos dois métodos são consistentes, desde que os resultados do método de orfandade estão incluídos dentro dos intervalos de confiança das estimativas do método estocástico (TAB. 18). Os resultados de ambas as metodologias mostram que as mulheres com mais escolaridade são as que apresentam esperanças de vida ao nascer maiores. A diferença, em anos, entre as esperanças de vida das menos escolarizadas e das mais escolarizadas, nas duas metodologias, varia entre 6 (na PPV) e 9 anos (na PNAD); uma diferença, sem dúvida, não trivial. Vale ressaltar que foram escolhidas as esperanças de vida do método de orfandade que tem como período de referência uma data próxima à das estimativas feitas através do método estocástico. Estas estimativas são representativas de períodos diferentes no caso de cada amostra, os dados da PPV se referem ao ano 1979 e os da PNAD ao ano de 1965. Porém, nos dois casos, as diferenças de esperança de vida entre as mulheres de diferentes grupos educacionais são muito expressivas.

Uma diferença importante entre as estimativas de um e outro método é o fato dos níveis de mortalidade estimados pelo método estocástico serem inferiores aos estimados pelo método de orfandade na PPV, acontecendo o contrário quando usados os dados da PNAD. A comparação dos resultados mostra que o método estocástico consegue obter estimativas aparentemente confiáveis, desde que se considerem os métodos baseados em dados retrospectivos de sobrevivência válidos. Além disso, o método estocástico demonstrou apresentar mais de uma vantagem em relação ao método clássico de orfandade. Em primeiro lugar, os resultados do método estocástico foram validados na regressão Poisson como estatisticamente significativos, o que significa que os níveis e diferenças, sempre que se confie nos dados utilizados, são confiáveis. Além disso, os grandes intervalos de confiança, derivados também das regressões Poisson, indicam, neste caso, que as estimativas encontradas através desses dados são frágeis e que seria mais adequado encontrar uma base de dados que oferecesse mais confiança. Finalmente, o método estocástico permite realizar uma decomposição das taxas em idade-período-coorte que pode ser extremamente útil para entender a evolução dos diferenciais de mortalidade no tempo, caso os dados permitam esse tipo de análise.

7. CONCLUSÃO

O principal objetivo deste trabalho foi desenvolver um método, aqui chamado de método estocástico que permitisse estimar taxas de mortalidade adulta feminina em função da idade e do nível de escolaridade. O método estocástico combina dados retrospectivos de sobrevivência das mães dos entrevistados com informações socioeconômicas das mesmas, declaradas em pesquisas domiciliares. Este novo método é inspirado no método clássico de orfandade para estimar mortalidade adulta, porém, aplica técnicas estocásticas para criar um banco de dados longitudinais, a partir do qual podem ser aplicados modelos de regressão para estimar medidas de mortalidade. Da mesma forma que o método de orfandade, o método estocástico utiliza a idade do entrevistado como estimativa do tempo de exposição ao risco de morte das mães. A principal diferença com o método de orfandade de Brass (Brass e Hill, 1973) é o uso de microdados, ao invés de dados agregados, além de estimar o início e o fim do período de exposição ao risco de morte para cada mãe através do sorteio aleatório de idades à maternidade e à morte. Desta forma, é gerada uma base de dados de sobrevivência com estrutura longitudinal à qual podem ser aplicadas técnicas de regressão que modelem o número de mortes. O método estocástico é um método mais complexo de ser utilizado que o método de orfandade clássico, já que requer o uso de padrões históricos de fecundidade e mortalidade da população em estudo. Por outro lado, os seus resultados permitem o uso de técnicas de sobrevivência, que proporcionam estimativas de mortalidade mais confiáveis e que podem ser testados estatisticamente.

Uma vez realizada a apresentação e o desenvolvimento do método estocástico, este método foi aplicado ao caso do Brasil, para estimar taxas de mortalidade adulta feminina por grandes grupos etários e por nível de escolaridade, utilizando-se para isto, duas bases de dados: PPV e PNAD 1996. Além disso, foram estimadas probabilidades de sobrevivência utilizando o método de orfandade, para comparar com os resultados do método estocástico. Os resultados do método mostraram-se robustos em relação às estimativas de mortalidade por idade, independente do nível de educação. Observou-se que os resultados do método estocástico seguem as tendências

esperadas e coincidem com estimativas apresentadas em trabalhos anteriores. Por exemplo, foi calculada uma esperança de vida aos 30 anos de 46,12 anos (intervalo de confiança: 41,23 a 49,72 anos) com base na PPV, tendo 1979 como ano de referência. Ao comparar este resultado com a esperança de vida aos 30 anos estimada por Carvalho e Pinheiro (1986) para a década 1970/1980, de 42,2 anos, concluiu-se que esta última está dentro do intervalo de confiança estimado pelo método estocástico. Da mesma forma, a esperança de vida aos 30 anos de 40,4 anos, estimada por Carvalho (1976) para a década de 1960/1970, também se encontra dentro do intervalo de confiança das estimativas do método estocástico com base na PNAD (intervalo de confiança: 39,59 a 46,44 anos), que tem como ano de referência, 1965.

Os resultados do método estocástico e do método de orfandade coincidiram, em grande medida, em parte graças ao fato de que os intervalos de confiança do modelo de regressão são muito amplos. Esta comparação permite confirmar que o método é, no mínimo, tão válido quanto o método de orfandade para estimar mortalidade adulta. Porém, pode-se dizer que o método estocástico se mostrou superior ao método de orfandade em vários aspectos que é importante destacar. Por um lado, o método estocástico permite gerar estimativas através de modelos de regressão, a partir dos quais podem se obter intervalos de confiança e testes de significância estatística que permitem comprovar a robustez dos resultados. Além disso, a possibilidade de testar estatisticamente os dados, e que o método trabalhe com dados individuais, permite, por exemplo, testar várias divisões de variáveis em categorias, até obter aquela que ofereça resultados mais significativos e confiáveis. Esse argumento demonstra a maior flexibilidade do método estocástico em relação ao método de orfandade. Outra das vantagens relacionada com sua flexibilidade é que permite modelar o número de mortes em função de várias características socioeconômicas, interagindo entre si, característica que se torna mais complexa no caso do método de orfandade. Finalmente, o método estocástico permite gerar estimativas de mortalidade por coorte de nascimento das mães e/ou período. Desta forma, é possível alocar temporalmente as estimativas de mortalidade sem precisar de uma técnica adicional, como no caso do método de orfandade.

Outro dos objetivos principais da Tese foi a aplicação, tanto do método de orfandade como do método estocástico, para estimar diferenciais de mortalidade por nível de

escolaridade. A aplicação do método estocástico permitiu constatar, através de dados individuais, a existência de grandes diferenciais de mortalidade adulta feminina por nível de escolaridade no Brasil. As estimativas variaram principalmente em função da amostra de indivíduos considerada. Porém, as diferenças encontradas entre a base de dados da PPV e da PNAD, apesar de não ser desprezíveis em alguns casos, caminham na mesma direção, dando robustez aos resultados. No caso da amostra da PPV, composta por entrevistados (filhos) de 15 anos e mais, estimou-se uma mortalidade para as mulheres sem nenhuma escolaridade mais de duas vezes maior à mortalidade das mulheres com 9 anos de estudo e mais. O diferencial relativo de mortalidade é igualmente expressivo entre as mulheres sem escolaridade e as que apresentam de 1 a 8 anos de estudo, sendo as taxas de mortalidade do primeiro grupo cerca de 70% maiores do que as do segundo grupo. Em outros países ocidentais, encontram-se diferenças desta magnitude apenas quando se estuda a mortalidade masculina (Elo e Preston, 1996; Kunst e Mackenbach, 1994; Huisman *et al.*, 2007). Por exemplo, segundo o trabalho de Elo e Preston (1996), os homens de 25 a 65 anos, com menos de 8 anos de estudo, apresentavam taxas de mortalidade duas vezes maiores aos que tinham chegado à universidade em 1971. Nestes países, é difícil encontrar grandes diferenciais de mortalidade feminina entre os diversos grupos socioeconômicos (Elo e Preston, 1996; Molla *et al.*, 2004; Kunst e Mackenbach, 1994; Huisman *et al.*, 2007). Por outro lado, as estimativas encontradas para o Brasil coincidem com as de outros países em desenvolvimento, como a Coreia do Sul (Son, 2001; Son, 2004), onde as taxas de mortalidade das mulheres com menos de 8 anos de estudo eram até três vezes maiores que as taxas das mulheres universitárias. Contudo, é difícil fazer uma comparação clara com outros países devido ao uso de diferentes técnicas e de diferentes medidas socioeconômicas. No caso do Brasil, a proporção de mulheres analfabetas é tão elevada que estas constituem uma categoria própria. Já nos países ocidentais, em função da obrigatoriedade da educação básica, o grupo com menor escolaridade se refere normalmente aos 8 anos de estudo. A distribuição dos indivíduos por nível de escolaridade também é diferente segundo o país, o que também pode estar afetando a desigualdade observada na mortalidade (Kunst *et al.*, 2004).

Como foi mencionado acima, o método estocástico permite obter dados através dos quais podem ser estimadas medidas de mortalidade por coorte de nascimento das

mães e/ou período. Os resultados mostraram que as mulheres nascidas em coortes mais antigas apresentaram uma maior mortalidade que as mulheres de coortes mais recentes, nos mesmos grupos de idade. Além disso, o mesmo diferencial de mortalidade por nível de escolaridade encontrado nos modelos anteriores continua existindo dentro de cada coorte. Por outro lado, embora tenha sido tentado medir variações nos níveis de mortalidade por idade e escolaridade, dentro de cada coorte, através de modelos de interação, o tamanho reduzido da amostra da PPV e os problemas de dados faltantes da variável escolaridade da amostra da PNAD, impediram análises com a devida precisão estatística. Sendo assim, os valores estimados e apresentados nesta tese se referem a uma média dos diferenciais nas últimas décadas.

A aplicação do método estocástico para o Brasil apresentou outras limitações a ser destacadas. As pesquisas amostrais utilizadas contêm uma proporção significativa de dados faltantes na variável escolaridade das mães dos entrevistados: 3,4% na PPV e de 16% na PNAD. Em ambas as bases, a distribuição de dados faltantes não é completamente aleatória, já que está correlacionada à sobrevivência das mães, não sendo aconselhável retirá-las da amostra, por se tratar de um estudo de diferenciais de mortalidade. Este problema foi solucionado através do uso de várias técnicas para o tratamento da falta de informação e assumindo que os dados faltantes apresentavam uma distribuição aleatória, ao controlar por diversas variáveis da amostra. Através de cada uma das técnicas, foram imputados novos valores para substituir aqueles faltantes. O resultado da imputação foi considerado satisfatório, no caso da PPV, já que todas as imputações estimaram resultados semelhantes. No entanto, a imputação para a PNAD foi menos eficiente, devido à maior proporção de casos faltantes provocada pelo fato de que a pergunta sobre escolaridade das mães foi respondida unicamente pelos chefes de domicílio e os cônjuges. Conseqüentemente, os dados faltantes se concentraram entre as crianças e jovens com a mãe morta, gerando uma seleção que pode induzir viés nos modelos de imputação.

Como discutido acima, o tamanho reduzido da amostra da PPV, constituiu um problema a mais para o cálculo dos diferenciais de mortalidade por educação. Esta limitação impediu obter significância estatística para observar como as taxas de mortalidade por nível de escolaridade evoluem em função da idade, para

comprovar se os diferenciais de mortalidade encontrados no Brasil reduzem com a idade, como nos EUA e Europa. Não foi possível saber se de fato as interações entre idade e escolaridade não são significativas no Brasil ou se este resultado é apenas uma consequência da falta de poder estatístico da PPV. Como mencionado acima, também não foram significativos aqueles modelos controlando pela coorte, idade e escolaridade. De qualquer forma, é importante destacar que esta é uma limitação dos dados e não da metodologia. Neste sentido, uma solução seria tornar a questão sobre escolaridade das mães na PNAD obrigatória para todos os moradores dos domicílios entrevistados. Assim, haveria uma redução significativa na proporção de dados faltantes nessa pesquisa, provavelmente, para chegar a um nível similar ao encontrado no caso da PPV. Como a PNAD tem uma amostra substancialmente maior que a PPV, seria possível examinar o comportamento dos diferenciais de educação na mortalidade por idade, período e coorte. Esta mudança na pesquisa seria de fácil aplicação e implicaria poder analisar a evolução dos diferenciais, por escolaridade, da mortalidade no Brasil durante as últimas décadas, além de permitir realizar projeções futuras destes diferenciais.

A aplicação do método estocástico poderia ter sido ainda mais refinada, caso as distribuições de fecundidade e mortalidade utilizadas na definição dos períodos de risco à morte tivessem sido diferenciadas segundo a escolaridade das mães. A necessidade de estimar coeficientes bootstrap nos modelos de regressão multiplicou a geração aleatória do início e fim dos períodos de exposição ao risco de morte por 500 vezes, para cada mulher. Introduzir distribuições de fecundidade e mortalidade, diferenciadas segundo o nível de escolaridade das mães teria implicado um aumento importante da complexidade do método. De qualquer forma, é difícil prever se este maior refinamento teria gerado grandes melhoras na precisão do método.

A apresentação de estimativas de mortalidade por nível de escolaridade para mulheres adultas no Brasil representa uma contribuição importante para o estudo das consequências da desigualdade social sobre o bem-estar da população. Este tipo de estudo permite elaborar, também, tabelas de sobrevivência e esperanças de vida por nível de escolaridade, que podem ser de grande utilidade nos estudos sobre previdência social. Igualmente, as projeções populacionais poderiam se beneficiar em grande medida destas estimativas, permitindo observar, por exemplo, como seria a

esperança de vida do Brasil caso mudasse a prevalência dos níveis de educação no país.

Finalmente, espera-se que o desenvolvimento desta proposta metodológica, o método estocástico, contribua para o avanço no conhecimento sobre diferenciais de mortalidade adulta, por nível socioeconômico, em outros países que enfrentam problemas de dados similares aos do Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, D. M. X.; RODRIGUES, N. R. Diferenciais de mortalidade entre as regiões metropolitanas de Belo Horizonte e Salvador, 1985-1995. **Revista de Saúde Pública**, São Paulo, v. 34, n. 5, p. 514-21, out. 2000.

ABREU, G. P.; MEDRONHO, R. A.; ESCOSTEGUY, C. C. Análise dos óbitos por doença isquêmica do coração em idosos no município do Rio de Janeiro. **Revista da Sociedade de Cardiologia**, Rio de Janeiro, v. 20, n. 3, p. 226-232, maio/jun. 2007.

ADLER, N. E., *et al.* Socioeconomic status and health: the challenge of the gradient. **American Psychologist**, Washington, v. 49, n. 1, p. 15-24, Jan. 1994.

AKE, C. F. Rounding after multiple imputation with non-binary categorical covariates. In: ANNUAL MEETING OF THE SAS USERS GROUP INTERNATIONAL-SUGI, 30., 2005, Philadelphia. **Proceedings...** 2005. Focus Session. Paper, 112-30. Disponível em: <<http://www2.sas.com/proceedings/sugi30/toc.html>> . Acesso em: 15 mar. 2010.

ALLISON, P. **Missing data**. Thousand Oaks, Calif.; SAGE University, 2001. (Sage University papers. Quantitative applications in the social sciences, 07-136).

BARROS, F. C. *et al.* Saúde materno-infantil em Pelotas, Rio Grande do Sul, Brasil: Principais conclusões da comparação dos estudos das coortes de 1982 e 1993. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 12, p. 87-92, 1996. Suplemento 1.

BARROS, G. B.; RAMOS, M. Condicionantes da mortalidade na população no extremos sul do Brasil. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 15, 2006, Caxambu. **Anais...** Caxambu: ABEP, 2006.

BARROS, M. B. A. Considerações sobre a mortalidade no Brasil em 1980. **Revista de Saúde Pública**, São Paulo, v. 18, n. 2, p.122-37, abr. 1984.

BARROS, R. P.; FOGUEL, M.; ULYSSEA, G. (Org.). **Desigualdade de renda no Brasil: uma análise da queda recente**. Brasília, DF.: IPEA, 2007. v. 2.

BARROS, R. P.; SAWYER, D. O. Unequal opportunity to survive, education and regional disparities in Brazil. Rio de Janeiro: IPEA, 1993. (Texto para Discussão, 307).

BASSANESSI, S. L.; AZAMBUJA, M. I.; ACHUTTI, A. Mortalidade precoce por doenças cardiovasculares e desigualdades sociais em Porto Alegre: da evidência à ação. **Arquivos da Sociedade Brasileira de Cardiologia**, Porto Alegre, v. 90, n. 6, p. 403-12, jan. 2008.

BECKER, G. S. A theory of the allocation of time. **Economic Journal**, New York, v. 75, n. 299, p. 493-517, Sept. 1965.

BECKER, G. S.; MULLIGAN, C. B. The endogenous determination of time preference. **Quarterly Journal of Economics**, Cambridge, v. 112, n. 3, p. 729–758, Aug. 1997.

BECKETT, M. Converging health inequalities in later-life: an artifact of mortality selection. **Journal of Health and Social Behaviour**, Washington, v. 41, n. 4, p. 106–119, Mar. 2000.

BEHM, H. **Las desigualdades sociales ante la muerte en América Latina**. Santiago de Chile: CELADE, 1992.

BERHANE, Y. *et al.* Gender, literacy, and survival among Ethiopian adults, 1987–96. **Bulletin of the World Health Organization**, New York, v. 80, n. 9, p. 714–720, Sept. 2002.

BLAKELY, T.; HALES, S.; WOODWARD, A. **Poverty**: assessing the distribution of health risks by socioeconomic position at national and local levels. Geneva: World Health Organization, 2004. (Environmental Burden of Disease Series, 10)

BÖHNING, D.; DIETZ, E.; SCHLATTMANN, P. Zero-inflated count models and their applications in public health and social science. In: ROST, J., LANGEHEINE, R. (Ed.) **Applications of latent trait and latent class models in the social sciences**. Münster: Waxmann 1997. p.333–344.

BORRELL, C.; PLASÈNCIA A.; PASARÍN I. Widening social inequalities in mortality: the case of Barcelona, a southern European city. **Journal of Epidemiology and Community Health**, London, v. 51, n. 6, p. 659–67, Dec. 1997.

BRASS, W. On the scale of mortality. In: BRASS, W. (Ed.) **Biological aspects of demography**. London: Taylor & Francis, 1971. (Symposia of the Society for the Study of Human Biology, v. 10).

BRASS, W.; BAMGBOYE, E. A. **The time location of reports of survivorship: estimates for maternal and parental orphanhood and the ever-widowed**. London: Centre for Population Studies, 1981. (Working Paper, 81).

BRASS, W.; HILL, K. H. Estimating adult mortality from orphanhood. In: INTERNATIONAL POPULATION CONFERENCE, 17, 1973, Liege. **Proceedings...** Liege: IUSSP, 1973. v. 3, p. 111–123.

CALDWELL, J. C. Mass education as a determinant of mortality decline. In: CALDWELL, J. C.; SANTOW, G. (Ed.). **Selected readings in the cultural, social and behavioural determinants of health**. Canberra: The Australian National University: Health Transition Centre, 1989. p.101–109.

CARVALHO, J. A. M. **Tendências regionais de fecundidade e mortalidade no Brasil**. Belo Horizonte: CEDEPLAR/UFMG, 1974. (Monografia, 8).

CARVALHO, J. A. M.; PINHEIRO, S. M. G. **Fecundidade e mortalidade no Brasil 1970/1980**. Belo Horizonte, CEDEPLAR/UFMG, 1986. Relatório de Pesquisa.

CARVALHO, J. A. M.; WOOD, C. H. Renda e concentração da mortalidade no Brasil. **Estudos Econômicos**, São Paulo, v.7, n. 1, p. 107-30, jan./abr. 1977.

CASE, A. C.; DEATON, A. Broken down by work and sex: how our health declines. [S.l]: National Bureau of Economic Research, 2003. (NBER Working Papers, 9821). Disponível em: <<http://www.nber.org/papers/w9821>>. Acesso em: 3 abr. 2008.

CASE, A.; FERTIG, A.; PAXSON, C. The lasting impact of childhood health and circumstance. **Journal of Health Economics**, Amsterdam, v. 24, n. 2, p. 365-389, Mar. 2005.

CASELLI, G.; CAPOCACCIA, R. Age, period, cohort and early mortality: an analysis of adult mortality in Italy. **Population Studies**, London, v. 43, n. 1, p. 133-153, Mar. 1989.

CASSEL, J. The contribution of the social environment to host resistance. **American Journal of Epidemiology**, Baltimore, v. 104, n.2, p. 107-123, Aug. 1976.

CASTILLA, F. Uma análise regional dos diferenciais Sócio-Econômicos da mortalidade na infância no Brasil, 1960-1980. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 10, 1996, Caxambu. **Anais...** Caxambu: ABEP, 1996.

CERQUEIRA, C. A.; PAES, N. A. Mortalidade por doenças crônico-degenerativas e relações com indicadores socioeconômicos no Brasil. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 11, 1998, Caxambu. **Anais...** Caxambu: ABEP, 1998.

CHACKIEL, J., ORELLANA, H. Adult female mortality trends from retrospective questions about maternal orphanhood included in censuses and surveys. In: **International Population Conference**, 20., Florence, 1985. Liège, IUSSP, 1985. v. 4, p. 39-51.

CHIRIKOS, T. N.; NESTEL, G. Further evidence on the economic effects of poor health. **The Review of Economics and Statistics**, Cambridge, v. 67, n. 1, p. 61-69, Feb. 1985.

CLELAND, J. G. Maternal education and child survival: further evidence and explanations. In: CALDWELL, J. *et al.* (Ed.). **Health transition: the cultural, social and behavioural determinants of health**. Canberra: The Australian National University: Health Transition Centre, 1990. p. 400-419.

COALE, A. J.; DEMENY, P. **Regional model life tables and stable populations**. Princeton, Princeton University, 1966.

CORDEIRO, R.; SILVA, E. A. Desigualdade da sobrevivência de trabalhadores de Botucatu, São Paulo, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 17, n. 4, p.593-601, jul./ago. 2001.

CRIMMINS E. M.; SAITO Y. Trends in health life expectancy in the United States, 1970-1990: gender, racial, and educational differences. **Social Science and Medicine**, Oxford, v. 52, n. 11, p. 1629-1641, June 2001.

CUTLER, D. M.; DEATON, A. S.; LLERAS-MUNEY, A. **The determinants of mortality**. [S. l.]. National Bureau of Economic Research, 2005. (NBER Working Paper, 11963). Disponível em: <<http://www.nber.org/papers/w11963>> Acesso em: 4 jul. 2008.

CUTLER, D. M. *et al.* **Explaining the rise in educational gradients in mortality**. [S. l.]: National Bureau of Economic Research, 2010. (NBER Working Paper, 15678, 2010). Disponível em: <<http://www.nber.org/papers/w15678>>. Acesso em: 20 fev. 2009.

CUTLER, D. M.; LLERAS-MUNEY, A. **Education and health: evaluating theories and evidence**. [S. l.]: National Bureau of Economic Research, 2006. (NBER Working Paper, 12352). Disponível em: <<http://www.nber.org/papers/w12352>>. Acesso em: 10 out. 2008.

DACHS, J. N. W. Determinantes das desigualdades na auto-avaliação do estado de saúde no Brasil: análise dos dados da PNAD/1998. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 7, n. 4, p.641-657, set. 2002.

DAVISON, A. C.; HINKLEY, D. V. **Bootstrap methods and their application**. Cambridge: Cambridge University Press, 1997.

DESPLANQUES, G. **La mortalite des adults**: resultats de 2 etudes longitudinales (periode 1955-1980). Paris: Institut National d'Etudes Démographiques, 1984.

DOBLHAMMER, G. **The late life legacy of very early life**. Rostock: Max Planck Institute for Demographic Research, 2003. (MPIDR Working Paper, WP-2003-030).

DOORNBOS, G.; KROMHOUT, D. Educational level and mortality in a 32-year follow-up study of 18-year-old men in the Netherlands. **International Journal of Epidemiology**, Oxford, v. 19, p. 374-379, Jun. 1990.

DRESSLER, W. W. Culture and blood pressure: using consensus analysis to create a measurement. **Cultural Anthropology Methods**, Florida, v. 8, n. 3, p. 6-8, Aug. 1996.

DRUMMOND JR.; M.; BARROS, M. B. A. Desigualdades socioespaciais na mortalidade do adulto no Município de São Paulo. **Revista Brasileira de Epidemiologia**, São Paulo, v. 2, n.1/2, p. 34-49, abr./jun. 1999.

DUARTE, E. C., SCHNEIDER, M. C., PAES-SOUSA, R., SILVA, J. B., CASTILLO-SALGADO, C. Expectativa de vida ao nascer e mortalidade no Brasil em 1999: análise exploratória dos diferenciais regionais. **Revista Panamericana de Salud Publica**, Washinton, v. 12, n. 6, p. 436-444, 2002.

DULEEP, H. O. Measuring socioeconomic mortality differentials over time. **Demography**, Chicago, v. 26, n. 2, p.345-351, May. 1989.

DUNCAN, B. B. *et al.*, A. Social Inequality in Mortality in São Paulo State, Brazil. **International Journal of Epidemiology**, Oxford, v. 24, n. 2, Apr. 1995.

EFRON, B. Bootstrap Methods: another look at the Jackknife. **The Annals of Statistics**, Ann Arbor, v.7, n.1, p.1–26, Jan.1979.

ELO, I. T.; PRESTON, S. H. Educational differentials in mortality: United States, 1979-85. **Social Science and Medicine**, Oxford, v. 42, p. 47-57, Jan. 1996.

ELO, I. T. *et al.* Mortality among elderly hispanics in the United States: past evidence and new results. **Demography**, Chicago, v. 41, n. 1, p.109-128, Feb. 2004.

ELO, I. T.; PRESTON, S. H. Effects of early-life conditions on adult mortality: a review. **Population Index**, Princeton, v. 58, n. 2, p. 186-212, Jul./Aug. 1992.

ETCHES, J. **Economic inequality in adult mortality in Canadá**: analyses of the longitudinal administrative databank. 2009. 257 f. Tese (Doctoral Thesis in Epidemiology)- Graduate Department of Public Health Sciences, University of Toronto Faculty of Medicine, 2009.

FELDMAN, J. J. *et al.* National trends in educational differentials in mortality. **American Journal of Epidemiology**, Baltimore, v.129, n. 5, p.919-933, May 1989.

FERNANDES, D. M. Diferenciais de mortalidade segundo instrução: regiões metropolitanas. Brasil, 1970. In: ENCONTRO DE ESTUDOS POPULACIONAIS, 4., 1984, Águas de São Pedro. **Anais....** Águas de São Pedro: ABEP, 1984. v.2, p.643-660.

FERRIE, J. E. (Ed.) **Work, stress and health: the Whitehall II study**. London: Cabinet Office, 2004.

FINKELSTEIN, A., MCKNIGHT, R. **What did medicare do (and was it worth it)?** [S. I.]. National Bureau of Economic Research, 2005. (NBER Working Paper, 11609). Disponível em: <<http://www.nber.org/papers/w11609>>. Acesso em: 15 nov. 2009.

FOX, A. J.; GOLDBLATT, P. O. **Longitudinal study. socio-demographic mortality differentials**. London: Her Majesty's Stationery Office, 1982. (Series LS, 1).

FOX, A. J.; ADELSTEIN, A. M. Occupational mortality: work or way of life? **Journal of Epidemiology and Community Health**, London, v. 32, p. 73-78, Mar. 1978.

FUCHS, V. **Economic aspects of health**. Chicago: University of Chicago Press, 1982.

GLIED, S.; LLERAS-MUNEY, A. **Health inequality, education and medical innovation**. [S. I.]: National Bureau of Economic Research, 2003. (NBER Working Paper, 9738). Disponível em: <<http://www.nber.org/papers/w9738>>. Acesso em: 10 dec. 2008.

GOLDMAN D., SMITH J. Socioeconomic differences in the adoption of new medical technologies. **American Economic Review**, Nashville, v. 95, n. 2, p.234-237, May 2005.

GOLDMAN, N. Social inequalities in health, disentangling the underlying mechanisms. **Annals New York Academy of Sciences**, New York, v.954, n.1, p.118-139, Dec. 2001.

GOODMAN, L. A.; KEYFITZ, N.; PULLUM, T. W. **La formación de familia y la frecuencia con que se dan diversas relaciones de parentesco**. Santiago de Chile: CELADE, 1975.

GROOT, W., VAN DEN BRINK, H. M. The health effects of education. **Economics of Education Review**, Cambridge, v. 26, n. 2, p. 186-200, Apr. 2007.

GROSSMAN, M. On the concept of health capital and the demand for health. **The Journal of Political Economy**, Chicago, v. 80, n. 2, p. 223-55, mar.-apr. 1972.

GROSSMAN, M. The relationship between health and schooling. **Eastern Economic Journal**, Bloomsburg, v. 34, n.3, p. 281-292, Mar. 2008.

HENRY, L. Mesure indirecte de la mortalité des adultes. **Population**, Paris, v. 15, n. 3, p. 457-466, Jun./Jul. 1960.

HERNANDEZ, E. M.; WARREN, J. R. The effects of macro- and individual-level socioeconomic status on child mortality in Brazil, 1970 to 2000. In: POPULATION ASSOCIATION OF AMERICA. **Annual Meeting Program**. New Orleans, 2008. Disponível em: <http://paa2008.princeton.edu/abstractViewer.aspx?submissionId=80452>. Acesso em: 05 mar. 2010.

HILL, K.; TRUSSELL, J. Further developments in indirect mortality estimation. **Population Studies**, London, v. 31, n. 2, p. 313-334, Jul. 1977.

HIMES, C. L., PRESTON, S. H., CONDRAN, G. A. A relational model of mortality at older ages in low mortality countries. **Population Studies**, London, v. 48, n. 2, p. 269-291, Jul. 1994.

HOFFMAN, R. Do socioeconomic mortality differences decrease with rising age? **Demographic Research**, Rostock, v. 13, n. 2, p. 35-62, Aug. 2005.

HOFFMAN, R. **Does the socioeconomic mortality gradient interact with age? Evidence from US survey data and Danish register data**. Rostock: Max Planck Institute for Demographic Research, 2005. (WORKING PAPER, WP 2005-020).

HONAKER, J.; KING, G.; BLACKWELL, M. **Amelia II**: a program for missing data. users guide. Cambridge: Harvard University, 2009. Disponível em: <http://gking.harvard.edu/amelia>. Acesso em: 2 fev. 2010.

HORTA, C. J. G.; CARVALHO, J. A. M.; FRIAS, L. A. M. Recomposição da fecundidade por geração para Brasil e regiões: atualização e revisão. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 12., 2000, Caxambu. **Anais...** Caxambu: ABEP, 2000.

HOUSE, J. S. *et al.* The social stratification of aging and health. **Journal of Health and Social Behavior**, Washington, v. 35, p. 213-234, Sept. 1994.

HSIEH, C. C.; PUGH, M. D. Poverty, income inequality, and violent crime: a meta-analysis of recent aggregate data studies. **Criminal Justice Review**, Atlanta, v. 18, p. 182–202, Oct./Dec. 1993.

HU, Y.; GOLDMAN, N. Mortality differentials by marital status: an international comparison. **Demography**, Chicago, v. 27, n. 2, p. 233-250, May 1990.

HUISMAN, M. *et al.* Socioeconomic inequalities in mortality among elderly people in Europe. **Journal of Epidemiology and Community Health**, London, v. 58, p. 468–75, Jun. 2004.

HUMMER, R. A.; ROGERS, R. G.; EBERSTEIN, I. W. Sociodemographic differentials in adult mortality: a review of analytic approaches. **Population and Development Review**, New York, v. 24, n. 3, p. 553, Sept. 1998.

HUMMER, R. A. *et al.* Religious participation and U.S. adult mortality. **Demography**, Chicago, v. 36, n. 2, p. 273-85, May. 1999.

HURT, L. S.; RONSMANS, C.; SAHA, S. Effects of education and other socioeconomic factors on middle age mortality in rural Bangladesh. **Journal of Epidemiology and Community Health**, London, v. 58, n. 4, p. 315-320, Jun. 2004.

IBGE. **Anuário Estatístico do Brasil**. Rio de Janeiro, v. 51, jan. 1991.

IBGE. **Pesquisa Nacional por Amostra de Domicílios - (PNAD) 1996**: Microdados. Rio de Janeiro, 1996.

IBGE. **Pesquisa sobre padrões de vida 1996-1997**: microdados. Rio de Janeiro, 1999. 1 CD ROM.

ISHITANI, L. H. *et al.* Desigualdade social e mortalidade precoce por doenças cardiovasculares no Brasil. **Revista de Saúde Pública**, São Paulo, v. 40, n. 4, p. 684-91, ago. 2006.

JOHNSON, N., J.; SORLIE, P. D.; BACKLUND, E. The impact of specific occupation on mortality in the U.S.: national longitudinal mortality study. **Demography**, Chicago, v. 36, n. 3, p. 355-367, Aug. 1999.

JONES, M. P. Indicator and stratification methods for missing explanatory variables in multiple linear regression. **Journal of the American Statistical Association**, New York, v. 91, n. 433, p. 222-230, Mar. 1996.

KAGAMIMORI, S.; IIBUCHI, Y.; FOX, J. A comparison of socioeconomic differences in mortality between Japan and England and Wales. **World Health Statistics Quarterly**, Geneva, v. 36, n. 2, p. 119–28, Mar. 1983.

KAWACHI, I.; KENNEDY, B, P.; LOCHNER, K. Social capital, income inequality, and mortality. **American Journal of Public Health**, Washington, v. 87, n. 9, p. 1491–1498, Sept. 1997.

KESTENBAUM, B. A Description of the extreme aged population based on improved medicare enrollment data. **Demography**, Chicago, v. 29, n. 4, p. 565-80, Nov. 1992.

KING, G. *et al.* Analyzing incomplete political science data: an alternative algorithm for multiple imputation. **American Political Science Review**, Washington, v. 95, n. 1, p. 49-69, Mar. 2001.

KINGTON, R. S.; NICKENS, H. W. Racial and ethnic differences in health: recent trends, current patterns, future directions. In: SMELSER, N. J.; WILSON, W. J.; MITCHELL, F. (Ed.). **America becoming: racial trends and their consequences**. Washington, DC: National Research Council, 2001. v. 2.

KITAGAWA, E. M.; HAUSER, P. M. **Differential mortality in the United States: a study in socioeconomic epidemiology**. Cambridge: Harvard University Press, 1973.

KOCH, E. *et al.* Desigualdad educacional y socioeconómica como determinante de mortalidad en Chile: análisis de sobrevivencia en la cohorte del proyecto San Francisco. **Revista Médica de Chile**, Santiago de Chile, v. 135, n. 11, p. 1370-1379, nov. 2007.

KRIEGER, N. Theories for social epidemiology in the 21st century: an ecosocial perspective. **International Journal of Epidemiology**, Oxford, v. 30, p. 668-677, ago. 2001.

KRIEGER, N.; WILLIAMS, D. R.; MOSS, N. E. Measuring social class in US public health research: concepts, methodologies, and guidelines. **Annual Review of Public Health**, Palo Alto, v. 18, p. 341-7, May. 1997.

KUNST, A. E. *et al.* Monitoring of trends in socioeconomic inequalities in mortality: Experiences from a European project. **Demographic Research**, Rostock, special collectio 2, article 9, Apr. 2004.

KUNST, A. E.; GROENHOF, F.; MACKENBACH, J. P. Occupational class and cause specific mortality in middle aged men in 11 European countries: comparison of population based studies. **British Medical Journal**, London, v. 316, n. 7145, p.1636-1642, May 1998.

KUNST, A. E.; MACKENBACH, J. P. The size of mortality differences associated with education level in nine industrialized countries. **American Journal of Public Health**, Washington, v. 84, n. 6, p.932-937, Apr. 1994.

LAUDERDALE, D. S. Education and survival: birth cohort, period, and age effects. **Demography**, Chicago, v. 38, n. 4, p. 551-561, Nov. 2001.

LIANG, J. *et al.* Old age mortality in Japan: does the socioeconomic gradient interact with gender and age? **The Journal of Gerontology. Series B: Psychological Sciences and Social Sciences**, Washington, v. 57, n.5, p. 294-307, Sept. 2002.

LIANG, J. *et al.* Socioeconomic gradient in old age mortality in Whuhun, China. **Journal of Gerontology, Psychological Sciences, and Social Sciences**, Washington, v.55, n.4, p. 222-33, Jul. 2000.

LIMA-COSTA, M. F.; MATOS, D. L.; CAMARANO, A. A. Evolução das desigualdades sociais em saúde entre idosos e adultos brasileiros: um estudo baseado na Pesquisa Nacional por Amostra de Domicílios (PNAD 1998, 2003). **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 11, n. 4, p. 941-950, out./dez. 2006.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data**. Chichester: Wiley, 2002.

LLERAS-MUNEY, A. The relationship between education and adult mortality in the United States. **Review of Economic Studies**, Oxford, v. 72, n. 250, p. 189–221, Jan. 2005.

LONG, J. S.; FREESE, J. **Regression models for categorical dependent variables using stata**. Rev. ed. College Station, TX: Stata Press, 2003.

LOTKA, A. J. Orphanhood in relation to demographic factors: a study in population analysis. **Metron**, v. 9, n. 193, p. 37-109, Jan./Mar. 1939.

LUY, M. **Estimating mortality differentials in developed populations from survey information on maternal and paternal orphanhood**. Vienna: Vienna Institute of Demography, 2009. (European Demographic Research Papers, 3).

MACEDO, A. C. *et al.* Violência e desigualdade social: mortalidade por homicídios e condições de vida em Salvador, Brasil. **Revista de Saúde Pública**, São Paulo, v. 35, n. 6, p. 515-22, dez. 2001.

MACKENBACH, J. P., KUNST, A. E. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. **Social Science and Medicine**, Oxford, v. 44, n. 6, p.757-71, Mar. 1997.

MACKENBACH, J. P. *et al.* Widening socioeconomic inequalities in mortality in six Western European countries. **International Journal of Epidemiology**, Oxford, v. 32, n. 5, p. 830-837, Oct. 2003.

MANTON, K. G.; STALLARD, E. Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: explaining the black/white mortality crossover by a model of mortality selection. **Human Biology**, Detroit, v. 53, n. 1, p.47-67, Jan. 1981.

MARMOT, M. G.; MCDOWALL, M. E. Mortality decline and widening social inequalities. **Lancet**, v. 328, n. 8501, p. 274-276, Feb. 1986.

MARMOT, M. G.; SHIPLEY, M. J. Do socioeconomic differences in mortality persist after retirement? 25 year follow up of civil servants from the first Whitehall study. **British Medical Journal**, London, v. 313, n. 7066, p. 1177-1180, Nov. 1996.

MARMOT, M. The influence of income on health: views of an epidemiologist. **Health affairs**, Millwood, v. 21, n. 2, p. 31-46, Mar./Apr. 2002.

MARMOT, M.; WILKINSON, R. G. Psychosocial and material pathways in the relation between income and health: a response to Lynch et al. **British Medical Journal**, London, v. 322, n. 7296, p. 1233-1236, Jul. 2001.

MARTIKANIEN, P.; VALKONEN, T. Bias related to the exclusion of the economically inactive in studies on social class differences in mortality. **International Journal of Epidemiology**, Oxford, v. 28, p. 899-904, Oct. 1999.

MCDONOUGH, P. *et al.* Gender and the socioeconomic gradient in mortality. **Journal of Health and Social Behavior**, Washington, v. 40, n. 1, p. 17-31, mar. 1999.

MCEWEN, B. S. Allostasis and allostatic load: implications for neuropsychopharmacology. **Neuropsychopharmacology**, New York, v. 22, n. 2, Feb. 2000.

MCEWEN, B. S. Protective and damaging effects of stress mediators. **New England Journal of Medicine**, Royston, n. 338, p.171–179, Jan. 1998.

MERRICK, T. W. The effect of piped water on early childhood mortality in urban Brazil, 1970 to 1976. **Demography**, Washington, v. 22, n. 1, p. 1-24, Feb. 1985.

MESSIAS, E. Income inequality, illiteracy rate, and life expectancy in Brazil. **American Journal of Public Health**, Washington, v. 93, n. 8, p.1294-1296, Aug. 2003.

MIGUEL, E., KREMER, M. Worms: identifying impacts on education and health the presence of treatment externalities. **Econometrica**, Chicago, v. 72, n. 1, p. 159-217, Jan. 2004.

MOLLA, M. T.; MADANS, J. H.; WAGENER, D. K. Differentials in adult mortality and activity limitation by years of education in the United States at the end of the 1990s. **Population and Development Review**, New York, v. 30, n. 4, p. 625-646, Dec. 2004.

MULATU, M. S.; SCHOOLER, C. Causal connections between socio-economic status and health: reciprocal effects and mediating mechanisms. **Journal of Health and Social Behaviour**, Washington, v. 43, n. 1, p. 22-41, Mar 2002.

NAJMAN, J. M.; CONGALTON, A. A. Australian occupational mortality, 1965–67: cause specific or general susceptibility. **Sociology of Health & Illness**, Oxford, v. 1, n. 2, p. 158-176, Jun. 2008.

NIE, N. H. *et al.* **Statistical package for the social sciences**. 2nd ed. Mc-Graw Hill, 1975.

NORONHA, K. V. M. S.; ANDRADE, M. V. **O efeito da distribuição de renda sobre o estado de saúde individual no Brasil**. Natal: ANPEC, 2005.

PAES, N. A. **Elderly mortality in Brazil: trends, differentials, causes and links with socioeconomic indicators**. 2001. Trabalho apresentado na XXIVth IUSSP Population Conference, Salvador de Bahia, Brasil, 2001. Disponível em: <www.iussp.org/Brazil2001/s10/S15_02_Paes.pdf>. Acesso em: 12 dec. 2007.

PAES-SOUSA, R. Diferenciais intra-urbanos de mortalidade em Belo Horizonte, Minas Gerais, Brasil, 1994: revisitando o debate sobre transições demográfica e epidemiológica. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 18, n. 5, p.1411-1421, set./out. 2002.

PAGNANELLI, F. **Une synthese des informations sur les differences de mortalite em Italie selon certains facteurs socio-economiques**. In: FIFTH MEETING OF THE UNDP/WHO/CICRED NETWORK ON SOCIOECONOMIC DIFFERENTIAL MORTALITY IN INDUSTRIALIZED COUNTRIES, 1991. **Proceedings...** Paris, France: Committee for International Cooperation in National Research on Demography, 1991.

PALLONI, A.; ARIAS, E. Paradox lost: Explaining the Hispanic adult mortality advantage. **Demography**, Chicago, v. 41, n. 3, p. 385-415, Aug. 2004.

PALLONI, A., HELIGMAN, L. Re-estimation of structural parameters to obtain estimates of mortality in developing countries. **Population Bulletin of the United Nations**, New York, n 18, p. 10-33, Mar. 1985.

PALLONI, A.; MASSAGLI, M.; MARCOTTE, J. Estimating adult mortality with maternal orphanhood data: analysis of sensitivity of the techniques. **Population Studies**, London, v. 38, n. 2, p. 255-279, July. 1984.

PAMUK E. *et al.* **Health, United States, 1998 with socioeconomic status and health chartbook**. Maryland: National Center for Health Statistics, 1998.

PAPPAS, G. *et al.* The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. **The New England Journal of Medicine**, Boston, v. 39, n. 2, p.103-109, July 1993.

PEARCE, N. E. *et al.* Mortality and social class in New Zealand. I: overall male mortality. **New Zealand Medical Journal**, Wellington, v. 96, n. 730, p. 281-285, Apr. 1983.

PRESTON, S. H.; ELO, I. T. Are educational differentials in adult mortality increasing in the United States? **Journal of aging and health**, Newbury Park, v.7, n. 4, p.476-96, Nov. 1995.

PRESTON, S. H.; HEUVELINE, P.; GUILLOT, M. **Demography: measuring and modeling population processes**. Victoria: Blackwell, 2001.

PRESTON, S. H.; TAUBMAN, P. Socioeconomic differences in adult mortality and health status. In: MARTIN, L. G.; PRESTON, S. H., **Demography of aging**. Washington: Academy Press, 1994. p. 279-318.

RODRÍGUEZ, G. **General linear statistical models**. Princeton: Princeton University/Graduate Program, 2001.

ROFMAN, R. Diferenciales de mortalidad adulta en Argentina. **Notas de Población**, Santiago, ano 22, n. 59, p.103-140, jun. 1994.

ROGERS, R. G.; HUMMER, R. A.; KRUEGER, P. M. Adult mortality. In: POSTON, D.; MICKLIN, M. **Handbook of population**. New York, Springer, 2002. pt. 2, cap. 10.

ROGOT, E.; SORLIE, P. D.; JOHNSON, N. J. Life expectancy by employment status, income, and education in the national longitudinal mortality study. **Public Health Reports**, Bethesda, v. 107, n. 4, p.457-461, Jul./Aug. 1992.

ROMERO, D. E.; CUNHA, C. B. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de menores de um ano registrados no sistema de informações sobre mortalidade (1996/2001). **Cadernos de Saúde Pública**, Rio de Janeiro, v. 22, n. 3, p. 673-84, mar. 2006.

ROSETO-BIXBY, L. The decline in adult mortality in Costa Rica. In: CHACKIEL, J., RUZICKA, L.; TIMAEUS, I. (Ed.). **Adult mortality in Latin America**. Oxford: Clarendon Press, 1996. p. 166-196.

ROSETO-BIXBY, L.; DOW, W. H. Surprising SES Gradients in mortality, health, and biomarkers in a Latin American population of adults. **Journal of Gerontology B: Psychol Scientific and Social Science**, Washington, v. 64, n. 1, p. 105-117, Jan. 2009.

ROSS, C.; WU, C. Education, age, and the cumulative advantage in health. **Journal of Health and Social Behaviour**, Washington, v. 37, p. 104-120, Mar. 1996.

ROYSTON, P. Multiple imputation of missing values. **Stata Journal**, College Station, v. 4, n. 3, p. 227-241, June 2004.

SASTRY, N. Urbanization, development, and under-five mortality differentials by place of residence in São Paulo, Brazil, 1970-1991. **Demographic Research**, Rostock, special collection 2, article 14, p. 366-386, Apr. 2004. Disponível em: <<http://www.demographic-research.org/special/2/9/S2-9.pdf>>. Acesso em: 12 dec. 2009.

SASTRY, N. What explains rural-urban differentials in child mortality in Brazil? **Social Science and Medicine**, Oxford, v. 44, n. 7, p. 989-1002, Oct. 1997.

SCHAFER, J. L. **Analysis of incomplete multivariate data**. London: Chapman & Hall, 1997.

SHRYOCK, H. S.; SIEGEL, J. S. **The methods and materials of demography**. Washington, U.S Bureau of the Census, 1973. 2 v.

SILVA, L. M. V.; PAIM, J. S.; COSTA, M. C. N. Desigualdades na mortalidade, espaço e estratos sociais. **Revista de Saúde Pública**, São Paulo, v. 33, n. 2, p. 187-197, abr. 1999.

SIMÕES, C. C. S.; MONTEIRO, C. A. Tendência secular e diferenciais regionais da mortalidade infantil no Brasil. In: MONTEIRO, C. A. (Org.). **Velhos e novos males da saúde no Brasil: a evolução do país e suas doenças**. São Paulo: HUCITEC, 1995. p. 153-156.

SMITH, G. D.; BARTLEY, M.; BLANE, D. The black report on socioeconomic inequalities in health 10 years on. **British Medical Journal**, London, v. 301, n. 6748, p.18-25, Aug. 1990.

SMITH, J. P. Healthy Bodies and thick wallets: the dual relation between health and economic status. **Journal of Economic Perspectives**, Nashville, v. 13, n. 2, p. 145-166, Mar./May 1999.

SMITH, J. P.; KINGTON, R. S. Demographic and economic correlates of health in old age. **Demography**, Chicago, v. 34, n. 1, p. 159-170, Feb. 1997.

SMITH, K. V.; GOLDMAN, N. Socioeconomic differences in health among older adults in Mexico. **Social Science and Medicine**, Oxford, v. 65, n. 7, p. 1372-1385, Oct. 2007.

SON, M. Commentary: why the educational effect is so strong in differentials of mortality in Korea? **International Journal of Epidemiology**, Oxford, v. 33, n. 2, p. 308-310, Apr. 2004.

SON, M. **Occupational class and health**: the differentials in mortality, morbidity and workplace injury rates by occupation, education, income and working conditions in korea. Tese (PhD) - Faculty of Medicine of the University of London, London School of Hygiene & Tropical Medicine, London: 2001.

SORLIE, P.*et al.* Black-white mortality differences by family income. **The Lancet**, Boston, v. 340, n. 8815, p. 346-50, Aug. 1992.

SUBRAMANIAN, S. V. *et al.* The mortality divide in India: the differential contributions of gender, caste, and standard of living across the life course. **American Journal of Public Health**, Washington, v. 96, n. 5, p.818-825, May 2006.

SZWARCWALD, C. L. *et al.* Desigualdade de renda e situação de saúde: o caso do Rio de Janeiro. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 15, n. 1, p.15-28, jan./mar. 1999.

SZWARCWALD, C. L.; CASTILHO, E. A. Estimates of infant mortality in Brazil in the 80's: a proposal for a methodological procedure. **Revista de Saúde Pública**, São Paulo, v. 29, n. 6, p. 451-62, abr. 1995.

TIMÆUS, I. An assessment of methods for estimating adult mortality from two sets of data on maternal orphanhood. **Demography**, Chicago, v. 23, n. 3, p. 435-450, Aug. 1986.

TIMÆUS, I. M. Estimation of adult mortality from orphanhood before and since marriage. **Population Studies**, London, v. 45, n. 3, p. 455-472, Nov. 1991a.

TIMÆUS, I. M. Estimation of adult mortality from orphanhood in adulthood. **Demography**, Chicago, v. 28, n. 2, p. 213-227, May 1991b.

TIMÆUS, I.; NUUN, A. Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method. **Health transition Review**, Canberra, v. 7, n. 2, p. 23-43, Feb. 1997.

TOWNSEND, P.; DAVIDSON, N. (Ed.). **Inequalities in health: the black report & the health divide**. London: Penguin Books, 1988.

TURRA, C. M. **Contabilidade das gerações: riqueza, sistema de transferencias e consequencias de mudancas no padrao demografico**. 2000. 204 f. Dissertação (Mestrado)-Centro de Desenvolvimento e Planejamento Regiona, Universidade Federal de Minas Gerais, 2000.

UNITED NATIONS. **Manual X: indirect techniques for demographic estimation**. New York, 1983.

UNITED NATIONS. **Tablas modelo de mortalidad para países en desarrollo**. Nueva York, 1983. (Estudios Demográficos, 77).

VAGERÖ, D.; LUNDBERG, O. Socio-economic mortality differentials among adults in Sweden. In: LOPEZ, A. D.; CASELLI, G.; VALKONEN, T. **Adult mortality in developed countries: from description to explanation**. Oxford: Clarendon Press, 1995. p.222-242.

VALKONEN, T. Trends in regional and socioeconomic mortality differentials in Finland. **International Journal of Health Sciences**, Assen, v. 3, n. 3-4 p. 157-166, 1993.

VAN BUUREN, S.; BOSHUIZEN, H. C.; KNOOK, D. L. Multiple imputation of missing blood pressure covariates in survival analysis. **Statistics in Medicine**, Chichester, v. 18, n. 6, p. 681-694, Mar. 1999.

VICTORA, C. G.; BARROS, F. C.; VAUGHAN, J. P. **Epidemiologia da desigualdade: um estudo longitudinal de 6000 crianças brasileiras**. 2. ed. São Paulo: Hucitec, 1989.

WALDRON, I.; LYE, D. Employment, unemployment, occupation, and smoking. **American Journal of Preventive Medicine**, New York, v. 5, n. 3, p. 142-149, May/June 1989.

WARREN, J. R.; HERNANDEZ, E. M. Did socioeconomic inequalities in morbidity and mortality change in the United States over the course of the twentieth century? **Journal of Health and Social Behavior**, Washington, v. 48, n. 4, p. 335-351, Dec. 2007.

WEBER, M. **Economia e sociedade: fundamentos da sociologia compreensiva**. Brasilia: Ed. UnB, 1991.

WILKINSON, R. G. Socioeconomic determinants of health: Health inequalities: relative or absolute material standards? **British Medical Journal**, v. 314, p. 591-595, Feb. 1997.

WILMOTH, J.R.; DENNIS, M. Social differences in older adult mortality in the United States: questions, data, methods, and results. In: ROBINE, J. M., *et al.* (Ed.). **Human longevity, individual life duration, and the growth of the oldest-old population**. Dordrecht: Springer, 2006.

WOOD, C. H.; CARVALHO, J. A. M. **The demography of inequality in Brazil**. London: Cambridge University Press, 1988.

YUCEL, R. M.; ZASLAVSKY, A. Practical suggestions on rounding in multiple imputation. In: JOINT STATISTICAL MEETINGS., Toronto, 2004. [Proceedings...], 2004. Disponível em: <<http://www.amstat.org/sections/srms/proceedings/y2004/files/Jsm2004-000652.pdf>>. Acesso em: 5 fev.2010.

ZHU, H.; XIE, Y. Socioeconomic differentials in mortality among the oldest old in China. **Research on Aging**, Thousand Oaks, v. 29, n. 125, p. 125-143, Feb. 2007.

ZIMMER, Z.; MARTIN, L. G.; CHANG, M. C. Changes in functional limitation and survival among Older Taiwanese in 1993, 1996 and 1999. **Population Studies**, London, v. 56, n.3, p. 265-276, Nov. 2002.

ZLOTNIK, H., HILL, K. The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality. **Demography**, Chicago, v. 18, n. 1, p. 103-122, Feb. 1981.

ANEXOS

TABELA A 1. Pesos pertencentes ao método de orfandade de Brass (Brass e Hill, 1973) para converter as proporções de órfãos em estimativas de probabilidades condicionais de sobrevivência das mães

Idade do filho	Idade média à maternidade								
	22	23	24	25	26	27	28	29	30
10	0,420	0,470	0,517	0,557	0,596	0,634	0,674	0,717	0,758
15	0,418	0,489	0,556	0,618	0,678	0,738	0,800	0,863	0,924
20	0,404	0,500	0,590	0,673	0,756	0,838	0,921	1,004	1,085
25	0,366	0,485	0,598	0,704	0,809	0,913	1,016	1,118	1,218
30	0,303	0,445	0,580	0,708	0,834	0,957	1,080	1,203	1,323
35	0,241	0,401	0,554	0,701	0,844	0,986	1,128	1,270	1,412
40	0,125	0,299	0,467	0,630	0,791	0,950	1,111	1,274	1,442
45	0,007	0,186	0,361	0,535	0,708	0,884	1,063	1,250	1,447
50	-0,190	-0,017	0,158	0,334	0,514	0,699	0,890	1,095	1,318
55	-0,368	-0,220	-0,059	0,101	0,270	0,456	0,645	0,856	1,083
60	-0,466	-0,352	-0,217	-0,084	0,053	0,220	0,378	0,579	0,800

Fonte: Brass e Hill, 1973

TABELA A 2. Valores da função $Z(M+n)$ para alocar no tempo as estimativas de mortalidade adulta do método de orfandade

U	f(u)	U	f(u)	U	f(u)	U	f(u)	U	f(u)
26	0,090	36	0,092	46	0,149	56	0,274	66	0,452
27	0,090	37	0,093	47	0,160	57	0,289	67	0,473
28	0,090	38	0,095	48	0,171	58	0,305	68	0,495
29	0,090	39	0,099	49	0,182	59	0,321	69	0,518
30	0,090	40	0,104	50	0,193	60	0,338	70	0,542
31	0,090	41	0,109	51	0,205	61	0,356	71	0,568
32	0,090	42	0,115	52	0,218	62	0,374	72	0,595
33	0,090	43	0,122	53	0,231	63	0,392	73	0,622
34	0,090	44	0,130	54	0,245	64	0,411	74	0,650
35	0,091	45	0,139	55	0,259	65	0,431	75	0,678

Fonte: Brass e Bangboye (1981)

TABELA A 3. Variáveis presentes no questionário da PPV 1996 utilizadas na Tese

Código	Pergunta ou nome da variável	Respostas e Categorias
Identificação de controle	Situação	Urbana Rural
02A04	Sexo	Homen Mulher
02A05	Data de nascimento	Dia
02A06		Mês
02A07		Ano
02A08	Quantos anos completos tem?	
02A09	Cor ou Raça	Branca Preta Parda Amarela/Indígena
02B01	Seu pai está vivo?	Sim Não Não sabe
02B02	Seu pai reside neste domicílio	Sim Não
02B03	No. de ordem do pai	
02B04	Seu pai sabe(ia) ler e escrever?	Sim Não Não sabe
02B05	Seu pai frequenta(ou) escola?	Sim Não Não sabe
02B06	Qual o nível de instrução de seu pai?	Não sabe Nenhuma série completa Elementar incompleto ou até a 3a série do 1o grau Elementar completo ou a 4a série do 1o grau Médio 1o ciclo incompleto ou 5a a 7a série do 1o grau Médio 1o ciclo ou 1o grau completo Médio 2o ciclo ou 2o grau incompleto Médio 2o ciclo o 2o grau Superior incompleto Superior Mestrado ou doutorado

Código	Pergunta ou nome da variável	Respostas e Categorias
04A01	Sabe ler jornal, revista, bilhete, etc.?	Sim Não
04A02	Sabe escrever uma carta ou bilhete?	Sim Não
04A04	Frequenta estabelecimento de ensino?	Sim Não
04A07	Já frequentou estabelecimento de ensino?	Sim Não
04A08	Qual a série que frequenta?	
04A09	A que grau corresponde esta série?	Creche Pré-escolar 1o grau 2o grau Supletivo 1o grau Supletivo 2o grau Pré-vestibular Superior Mestrado ou doutorado
04A01	Última série concluída com aprovação	
04A02	Último grau concluído com aprovação	Creche Pré-escolar Elementar Médio 1o ciclo ou ginásial 1o grau Supletivo 1o grau Médio 2o ciclo ou 2o grau Supletivo 2o grau Vestibular Superior Mestrado ou doutorado

Fonte: IBGE, PPV 1996

TABELA A 4. Variáveis presentes no questionário da PNAD 1996 utilizadas na Tese

Código	Pergunta ou nome da variável	Respostas e Categorias
UF	Unidade da Federação	
V0301	Número de Ordem	
V0302	Sexo	Masculino Feminino
V3031	Dia de nascimento	
V3032	Mês de nascimento	
V3033	Ano de nascimento	
V8005	Idade do morador	
V0404	Cor ou raça	Branca Preta Amarela Parda Indígena Ignorada
V0405	Tem mãe viva	Sim Não Não sabe
V0407	Número de ordem da mãe	
v0501	Naceu neste município	Sim Não Ignorado
V0502	Nasceu neste estado	Sim Não Ignorado
V5030	Código da UF / País estrangeiro em que nasceu	
V0601	Sabe ler e escrever	Sim Não Ignorado
V0602	Frequenta escola ou creche	Sim Não Ignorado
V0603	Qual o curso que frequenta	Ignorado Regular de 1o grau Regular de 2o grau Supletivo de 1o grau Supletivo de 2o grau Superior Alfabetização de adultos Pré-escolar ou creche Pré-vestibular Mestrado ou doutorado

Código	Pergunta ou nome da variável	Respostas e Categorias
V1218	A sua mãe sabe(ia) ler e escrever	Sim Não Não sabe informar Ignorado
V1219	Nível de instrução da sua mãe	Nunca frequentou escola ou não concluiu 1a série do 1o grau Elementar incompleto ou da 1a à 3a séries concluídas do 1o grau Elementar completo ou 4a série concluída do 1o grau Médio 1o ciclo incompleto ou da 5a à 7a série concluídas do 1o grau Médio 1o ciclo completo ou 8a série concluída do 1o grau Médio 2o ciclo incompleto ou 2o grau incompleto Médio 2o ciclo completo ou 2o grau completo Superior incompleto Superior completo Mestrado ou doutorado completo Não sabe informar Ignorado
V4721	Valor do rendimento mensal domiciliar - para todos os moradores	
V4722	Valor do rendimento mensal familiar I (inclusive agregado) - para todos os moradores	
V7424	Número de componentes da família (inclusive agregado)	

Fonte: IBGE, PNAD 1996

TABELA A 5. Distribuição dos valores faltantes da variável de escolaridade imputados em função das categorias de escolaridade e do tipo de imputação realizada, PPV e PNAD 1996

Base e tipo de imputação		Sem escolaridade	1 a 4 anos de estudo	5 a 8 anos de estudo	9 anos e mais de estudo
PPV	n=736				
	Imputação estocástica	39,65	36,24	12,5	11,62
	Imputação por algoritmo EM	39,65	36,24	12,5	11,62
	Imputação múltipla	39,02	36,62	12,75	11,62
	Imputação por hotdeck	37,75	35,61	15,03	11,62
PNAD	n=53.997				
	Imputação estocástica	29,86	38,79	18,56	12,78
	Imputação por algoritmo EM	29,9	38,69	18,58	12,84
	Imputação múltipla	29,29	37,22	19,79	13,7
	Imputação por hotdeck	29,45	38,47	19,44	12,63

Fonte: IBGE, PPV e PNAD 1996

TABELA A 6. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por regressão estocástica, PPV 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25-75}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25-75}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25-75}/l_{25}
0-4	0	1.782.101	9.234				6413541	20124				2278297	17679			
5-9	5	2.169.595	98.244				5939786	59967				1979136	26848			
10-14	10	2.754.409	133.572				6454029	133267				2250045	23039			
15-19	15	2.906.294	253.581	0,91975	0,854		6068648	191845	0,96936	0,688		1735216	28263	0,98397	0,732	
20-24	20	3.081.416	366.033	0,89382	0,992	0,91954	4332731	375042	0,92034	0,770	0,95806	1045258	39922	0,96321	0,829	0,98043
25-29	25	2.994.165	551.007	0,84458	1,131	0,90026	3156976	435760	0,87871	0,950	0,91826	651347	47013	0,93268	0,898	0,96010
30-34	30	2.462.114	823.486	0,74937	1,218	0,86536	2926186	676264	0,81228	1,001	0,87880	535866	52841	0,91024	0,939	0,93132
35-39	35	2.180.305	975.343	0,69092	1,288	0,76620	2578887	689220	0,78911	1,037	0,81314	334688	60308	0,84732	0,965	0,90807
40-44	40	1.911.120	1.476.827	0,56409	1,295	0,72839	2109155	718411	0,74593	1,008	0,78945	213086	62021	0,77456	0,927	0,84201
45-49	45	1.520.802	1.417.884	0,51751	1,275	0,57691	1235451	811547	0,60354	0,949	0,73860	176002	32553	0,84391	0,859	0,78437
50-54	50	839.522	1.577.850	0,34729	1,123	0,53852	569402	692606	0,45119	0,768	0,56818	69748	132700	0,34452	0,672	0,68021
55-59	55	452.066	1.886.229	0,19333	0,885	0,32957	316232	744224	0,29820	0,524	0,37839	37514	53966	0,41008	0,429	0,38195
60-64	60	265.753	1.572.727		0,607		159128	620519		0,277		15925	19765		0,196	
65-69	65	88.424	1.530.778				45091	584268				14412	43891			
70-74	70	26.071	1.124.503				9899	413552					54704			
75-79	75	35.340	741.381				12960	215946					44491			
80 e mais	80		738.492					282290					24223			

Fonte: IBGE, PPV 1996

TABELA A 7. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por imputação múltipla, PPV 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não órfãos (mãe viva)	S(n)	W(n)	l_{25+n}/l_{25}	Orfãos (mãe morta)	Não órfãos (mãe viva)	S(n)	W(n)	l_{25+n}/l_{25}	Orfãos (mãe morta)	Não órfãos (mãe viva)	S(n)	W(n)	l_{25+n}/l_{25}
0-4	0	1.779.383	9.234				6457580	21052				2236976	16751			
5-9	5	2.180.652	103.663				5905264	63163				2002601	18233			
10-14	10	2.756.713	133.572				6441851	131265				2259919	25041			
15-19	15	2.902.224	260.799	0,91755	0,854		6096661	184627	0,97061	0,688		1711273	28263	0,98375	0,732	
20-24	20	3.069.819	360.424	0,89493	0,992	0,91737	4321460	374245	0,92030	0,770	0,95902	1068126	46328	0,95843	0,829	0,97943
25-29	25	2.962.880	549.465	0,84356	1,131	0,90164	3186352	438316	0,87907	0,950	0,91824	653256	45999	0,93422	0,898	0,95596
30-34	30	2.457.591	822.371	0,74927	1,218	0,86414	2924827	677379	0,81195	1,001	0,87916	541748	52841	0,91113	0,939	0,93281
35-39	35	2.195.493	984.243	0,69046	1,288	0,76622	2571905	660683	0,79562	1,037	0,81256	326482	79945	0,80330	0,965	0,90740
40-44	40	1.892.191	1.473.366	0,56222	1,295	0,72835	2145611	721872	0,74826	1,008	0,79600	195559	62021	0,75922	0,927	0,80008
45-49	45	1.519.123	1.409.402	0,51873	1,275	0,57419	1235185	807778	0,60460	0,949	0,74086	177947	44804	0,79886	0,859	0,76483
50-54	50	840.475	1.562.285	0,34980	1,123	0,53958	568449	695154	0,44986	0,768	0,56868	69748	145717	0,32371	0,672	0,64311
55-59	55	454.274	1.862.129	0,19611	0,885	0,33211	314024	768324	0,29013	0,524	0,37385	37514	53966	0,41008	0,429	0,37302
60-64	60	265.753	1.583.086		0,607		159128	610160		0,277		15925	19765			0,196
65-69	65	89.332	1.518.312				44183	596734				14412	43891			
70-74	70	26.071	1.125.388				9899	412667				0	54704			
75-79	75	35.340	741.381				12960	220185				0	40252			
80 e mais	80	0	722.905				0	296487				0	25613			

Fonte: IBGE, PPV 1996

TABELA A 8. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por *hotdeck*, PPV 1996

Idade	n	Mães sem escolaridade				Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo					
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2,}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2,}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2,}$
0-4	0	1.782.819	10.162				6404402	20124				2286718	16751			
5-9	5	2.161.393	103.663				5900553	34728				2026571	46668			
10-14	10	2.732.027	133.572				6468962	135177				2257494	21129			
15-19	15	2.911.178	258.502	0,91845	0,854		6054165	194865	0,96882	0,688		1744815	20322	0,98849	0,732	
20-24	20	3.068.037	366.646	0,89325	0,992	0,91824	4364637	377159	0,92046	0,770	0,95768	1026731	37192	0,96504	0,829	0,98449
25-29	25	2.987.796	568.750	0,84008	1,131	0,90020	3170030	422358	0,88243	0,950	0,91856	644662	42672	0,93792	0,898	0,96227
30-34	30	2.463.470	827.959	0,74845	1,218	0,86009	2927846	669833	0,81382	1,001	0,88252	532850	54799	0,90675	0,939	0,93602
35-39	35	2.186.245	962.964	0,69422	1,288	0,76407	2588654	695479	0,78823	1,037	0,81477	318981	66428	0,82764	0,965	0,90401
40-44	40	1.933.561	1.465.569	0,56884	1,295	0,73126	2090885	729669	0,74130	1,008	0,78861	208915	62021	0,77109	0,927	0,82351
45-49	45	1.495.415	1.390.118	0,51825	1,275	0,58276	1281209	831658	0,60638	0,949	0,73436	155631	40208	0,79469	0,859	0,77443
50-54	50	823.653	1.498.670	0,35467	1,123	0,53843	582705	771786	0,43020	0,768	0,56549	72314	132700	0,35273	0,672	0,64981
55-59	55	452.645	1.868.691	0,19499	0,885	0,33629	315653	760164	0,29341	0,524	0,36511	37514	55564	0,40304	0,429	0,38145
60-64	60	275.284	1.577.851		0,607		142959	615395		0,277		22563	19765		0,196	
65-69	65	78.432	1.528.282				55083	581485				14412	49170			
70-74	70	26.071	1.123.675				9899	414380				0	54704			
75-79	75	35.340	741.571				12960	215925				0	44322			
80 e mais	80	0	754.413				0	276430				0	14162			

Fonte: IBGE, PPV 1996

TABELA A 9. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada com o algoritmo EM, PPV 1996

Idade	n	Mães sem escolaridade				Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo					
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2+}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2+}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2,++}/l_{2+}$
0-4	0	1.787.374	10.162			6444683	35720				2241882	1155				
5-9	5	2.190.028	118.579			5914761	37925				1983728	28555				
10-14	10	2.727.966	163.058			6479159	118205				2251358	8615				
15-19	15	2.894.937	271.748	0,91419	0,854	6086903	190864	0,96960	0,688		1728318	11077	0,99363	0,732		
20-24	20	3.074.892	402.316	0,88430	0,992	0,91395	4367686	309755	0,93378	0,770	0,96134	1016827	68926	0,93652	0,829	0,98389
25-29	25	2.952.454	592.652	0,83283	1,131	0,89103	3211620	385724	0,89278	0,950	0,93173	638414	55404	0,92015	0,898	0,93485
30-34	30	2.473.848	849.753	0,74433	1,218	0,85214	2875498	660714	0,81316	1,001	0,89288	574820	42124	0,93172	0,939	0,92085
35-39	35	2.188.285	1.012.293	0,68372	1,288	0,76179	2571869	644101	0,79972	1,037	0,81366	333726	68477	0,82975	0,965	0,92820
40-44	40	1.895.814	1.451.027	0,56645	1,295	0,71836	2141078	725226	0,74698	1,008	0,80014	196469	81006	0,70806	0,927	0,82086
45-49	45	1.485.520	1.415.214	0,51212	1,275	0,58139	1261570	796289	0,61305	0,949	0,74009	185165	50481	0,78578	0,859	0,71906
50-54	50	809.460	1.547.147	0,34349	1,123	0,53293	598265	723309	0,45269	0,768	0,57582	70947	132700	0,34838	0,672	0,64240
55-59	55	453.127	1.846.389	0,19705	0,885	0,32663	315171	762440	0,29247	0,524	0,37645	37514	75590	0,33168	0,429	0,33884
60-64	60	264.110	1.546.869		0,607		160771	638699		0,277		15925	27443		0,196	
65-69	65	78.432	1.492.545				45091	628399				24404	37993			
70-74	70	26.071	1.123.329				9899	406162				0	63268			
75-79	75	36.603	715.536				11697	233861				0	52421			
80 e mais	80	0	721.566				0	309277				0	14162			

Fonte: IBGE, PPV 1996

TABELA A 10. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por regressão estocástica, PNAD 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}
40-44	40	2.782.432	1.704.822	0,62007	1,295	0,76378	3137840	1384174	0,693903	1,008	0,78930	296715	72300	0,80407	0,927	0,85137
45-49	45	2.050.811	2.077.676	0,49675	1,275	0,65400	1945539	1360788	0,588429	0,949	0,68848	183693	86222	0,68056	0,859	0,78660
50-54	50	1.311.062	2.175.664	0,37602	1,123	0,51164	1031228	1264501	0,449194	0,768	0,55611	104674	75805	0,57998	0,672	0,64759
55-59	55	766.301	2.312.719	0,24888	0,885	0,36138	551474	1158416	0,32252	0,524	0,38892	40435	74435	0,35201	0,429	0,44982
60-64	60	367.910	2.421.167		0,607		239385	1080873		0,277		22520	68625		0,196	
65-69	65	151.966	2.140.825				83747	867324				6351	68422			
70-74	70	48.352	1.640.384				21446	577509				1949	58098			
75-79	75	17.234	1.112.510				3997	355729				0	33930			
80 e mais	80	10.656	1.041.826				4468	338196				0	32397			

Fonte: IBGE, PNAD 1996

TABELA A 11. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por imputação múltipla, PNAD 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	l_{25n}/l_{25}
40-44	40	2.745.928	1.652.194	0,62434	1,295	0,76726	3166870	1432836	0,688494	1,008	0,78518	304189	76266	0,79954	0,927	0,84841
45-49	45	2.047.080	2.041.671	0,50066	1,275	0,65836	1946521	1391546	0,583128	0,949	0,68307	186442	91469	0,67087	0,859	0,78133
50-54	50	1.302.922	2.146.668	0,37770	1,123	0,51583	1037420	1289539	0,445827	0,768	0,55126	106622	79763	0,57205	0,672	0,63848
55-59	55	757.020	2.288.631	0,24856	0,885	0,36284	560079	1182238	0,321456	0,524	0,38664	41111	74701	0,35498	0,429	0,44812
60-64	60	364.088	2.416.508		0,607		243060	1084715		0,277		22667	69442		0,196	
65-69	65	153.299	2.132.845				81569	875847				7196	67879			
70-74	70	47.450	1.649.927				22103	571804				2194	54260			
75-79	75	17.568	1.138.817				3663	331073				0	32279			
80 e mais	80	10.868	1.105.721				4256	281302				0	25396			

Fonte: IBGE, PNAD 1996

TABELA A 12. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada por *hotdeck*, PNAD 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$\frac{I_{s+1}}{I_s}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$\frac{I_{s+1}}{I_s}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$\frac{I_{s+1}}{I_s}$
40-44	40	2.718.764	1.625.985	0,62576	1,295	0,76496	3230922	1472352	0,686952	1,008	0,78703	267301	62959	0,80937	0,927	0,85049
45-49	45	2.032.980	2.017.493	0,50191	1,275	0,65983	1978031	1432370	0,58	0,949	0,68145	169032	74823	0,69317	0,859	0,79292
50-54	50	1.288.502	2.120.318	0,37799	1,123	0,51720	1062308	1333235	0,443452	0,768	0,54830	96154	62417	0,60638	0,672	0,66472
55-59	55	762.148	2.233.445	0,25442	0,885	0,36377	559739	1250243	0,309251	0,524	0,37959	36323	61882	0,36987	0,429	0,47134
60-64	60	362.831	2.370.117		0,607		247015	1147347		0,277		19969	53201		0,196	
65-69	65	151.314	2.084.367				84759	938479				5991	53725			
70-74	70	45.059	1.592.326				25464	639240				1224	44425			
75-79	75	16.970	1.111.559				4261	368033				0	22577			
80 e mais	80	10.223	1.058.281				4901	339979				0	14159			

Fonte: IBGE, PNAD 1996

TABELA A 13. Proporções de órfãos e probabilidades condicionais, para a variável de escolaridade imputada com o algoritmo EM, PNAD 1996

Idade	n	Mães sem escolaridade					Mães com 1 a 8 anos de estudo					Mães com 9 e anos e mais de estudo				
		Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2..}/l_{2.}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2..}/l_{2.}$	Orfãos (mãe morta)	Não orfãos (mãe viva)	S(n)	W(n)	$l_{2..}/l_{2.}$
40-44	40	2.753.097	1.652.311	0,62494	1,295	0,76397	3158708	1426057	0,688957	1,008	0,78869	305182	82928	0,78633	0,927	0,83868
45-49	45	2.035.690	2.028.355	0,50090	1,275	0,65906	1954914	1406281	0,581613	0,949	0,68343	189439	90050	0,67780	0,859	0,77097
50-54	50	1.291.269	2.114.721	0,37912	1,123	0,51593	1048415	1319527	0,442754	0,768	0,54938	107280	81722	0,56761	0,672	0,64168
55-59	55	746.886	2.243.273	0,24978	0,885	0,36423	568434	1225140	0,316928	0,524	0,38288	42890	77157	0,35728	0,429	0,44752
60-64	60	359.938	2.336.412		0,607		248449	1161696		0,277		21428	72557		0,196	
65-69	65	147.142	2.040.344				87468	969520				7454	66707			
70-74	70	46.577	1.549.596				23221	668845				1949	57550			
75-79	75	14.810	1.044.903				6421	423057				0	34209			
80 e mais	80	10.019	967.746				4215	412375				890	32298			

Fonte: IBGE, PNAD 1996

TABELA A 14. Pesos de Karup-King para obter probabilidades de sobrevivência de idades intermédias de grupos quinquenais

	N1.0	N2.0	N3.0	N4.0
N1.0	1	0	0	0
N1.2	0,656	0,552	-0,272	0,064
N1.4	0,408	0,856	-0,336	0,072
N1.6	0,232	0,984	-0,264	0,048
N1.8	0,104	1,008	-0,128	0,016
N2.0	0	1	0	0
N2.2	-0,064	0,912	0,168	-0,016
N2.4	-0,072	0,696	0,424	-0,048
N2.6	-0,048	0,424	0,696	-0,072
N2.8	-0,016	0,168	0,912	-0,064
N3.0	0	0	1	0
N3.2	0,016	-0,128	1,008	0,104
N3.4	0,048	-0,264	0,984	0,232
N3.6	0,072	-0,336	0,856	0,408
N3.8	0,064	-0,272	0,552	0,656
N4.0	0	0	0	1

Fonte: Shryock e Siegel, 1973

TABELA A 15. Tabela de sobrevivência padrão de Himes et al., (1994) para estender uma tabela de sobrevivência até os 100 anos

x	lx	x	lx
0	1,00000	50	0,51060
1	0,84992	51	0,50105
2	0,80696	52	0,49115
3	0,78758	53	0,48086
4	0,77622	54	0,47014
5	0,76906	55	0,45904
6	0,76419	56	0,44744
7	0,76009	57	0,43541
8	0,75643	58	0,42292
9	0,75317	59	0,40994
10	0,75019	60	0,39652
11	0,74767	61	0,38254
12	0,74509	62	0,36814
13	0,74254	63	0,34507
14	0,73962	64	0,33787
15	0,73618	65	0,32209
16	0,73275	66	0,30589
17	0,72873	67	0,28930
18	0,72408	68	0,27242
19	0,71886	69	0,25529
20	0,71304	70	0,23801
21	0,70686	71	0,22060
22	0,70048	72	0,20320
23	0,69436	73	0,18588
24	0,68839	74	0,16879
25	0,68261	75	0,15205
26	0,67638	76	0,13576
27	0,67036	77	0,12005
28	0,66432	78	0,10504
29	0,65836	79	0,09086
30	0,65249	80	0,07763
31	0,64657	81	0,06542
32	0,64056	82	0,05433
33	0,63455	83	0,04440
34	0,62841	84	0,03565
35	0,62227	85	0,02807
36	0,61605	86	0,02166
37	0,60969	87	0,01632
38	0,60320	88	0,01201
39	0,59657	89	0,00859
40	0,58981	90	0,00597
41	0,58293	91	0,00402
42	0,57591	92	0,00261
43	0,56861	93	0,00163
44	0,56114	94	0,00098
45	0,55345	95	0,00056
46	0,54542	96	0,00031
47	0,53718	97	0,00016
48	0,52867	98	0,00008
49	0,51979	99	0,00004

Fonte: Himes et al.(1994)

TABELA A 16. Brasil – Taxas de fecundidade por período e taxas de fecundidade ponderadas pela mortalidade, 1933 – 1988

Idade	1933	1938	1943	1948	1953	1958	1963	1968	1973	1978	1983	1988
Taxas de fecundidade observadas												
15-19	0,072	0,075	0,078	0,077	0,077	0,071	0,066	0,071	0,075	0,078	0,081	0,080
20-24	0,276	0,274	0,272	0,269	0,267	0,271	0,276	0,252	0,229	0,209	0,189	0,184
25-29	0,316	0,308	0,299	0,300	0,300	0,305	0,310	0,286	0,262	0,230	0,198	0,174
30-34	0,259	0,247	0,234	0,238	0,242	0,242	0,242	0,225	0,208	0,170	0,132	0,099
35-39	0,195	0,182	0,170	0,173	0,175	0,180	0,185	0,162	0,138	0,107	0,076	0,054
40-44	0,100	0,092	0,085	0,086	0,088	0,092	0,095	0,080	0,065	0,048	0,031	0,021
45-49	0,028	0,026	0,024	0,024	0,024	0,025	0,027	0,022	0,017	0,012	0,008	0,005
Taxas de fecundidade ponderadas pela mortalidade												
15-19	0,037	0,039	0,041	0,040	0,035	0,032	0,029	0,031	0,033	0,036	0,040	0,041
20-24	0,203	0,203	0,203	0,201	0,166	0,168	0,169	0,152	0,142	0,134	0,128	0,132
25-29	0,288	0,283	0,278	0,277	0,251	0,253	0,256	0,225	0,211	0,192	0,175	0,162
30-34	0,281	0,269	0,257	0,261	0,255	0,253	0,251	0,235	0,222	0,189	0,155	0,123
35-39	0,242	0,228	0,214	0,217	0,242	0,248	0,253	0,232	0,203	0,163	0,122	0,092
40-44	0,145	0,136	0,126	0,128	0,166	0,172	0,179	0,162	0,134	0,103	0,071	0,051
45-49	0,049	0,046	0,042	0,043	0,059	0,061	0,063	0,061	0,049	0,037	0,023	0,016

Fonte: Horta, Carvalho e Frias (2000)

TABELA A 17. Brasil – Probabilidades de sobrevivência, 1965, 1975, 1985

Idade	1960/70	1970/80	1980/1990
	Carvalho, 1974	Carvalho e Pinheiro, 1988	IBGE, 1985
0	100000	100000	100000
1	92084	92058	95004
2	89273	90852	
3	87838	90298	
5	86474	89646	94060
10	85183	89171	93769
15	84480	88870	93507
20	83565	88377	93119
25	82259	87697	92581
30	80638	86782	91880
35	78713	85632	90948
40	76507	84120	89665
45	73912	82046	87863
50	70796	79406	85364
55	66909	75931	81910
60	61814	71022	77104
65	55185	64485	70418
70	46730	55683	61066
75	36364	43859	48171
80 e mais	24944	29047	31353
e₀	57,98	63,29	68

Fonte: Carvalho, 1974; Carvalho e Pinheiro, 1988, IBGE, 1985

TABELA A 18. Brasil – Óbitos acumulados, 1965

1960/70			
Idade	óbitos acum.	Idade	óbitos acum.
0	0	50	29215
1	7916	51	29935
2	10727	52	30687
3	12162	53	31493
4	13393	54	32327
5	13526	55	33211
6	13711	56	34126
7	13978	57	35096
8	14281	58	36109
9	14576	59	37189
10	14817	60	38317
11	14992	61	39506
12	15130	62	40741
13	15252	63	42062
14	15375	64	43449
15	15520	65	44907
16	15683	66	46442
17	15852	67	48029
18	16031	68	49704
19	16224	69	51461
20	16435	70	53294
21	16666	71	55210
22	16914	72	57182
23	17177	73	59256
24	17453	74	61397
25	17741	75	63601
26	18040	76	65851
27	18352	77	68142
28	18677	78	70439
29	19013	79	72760
30	19362	80	75073
31	19723	81	77365
32	20097	82	79593
33	20482	83	81799
34	20879	84	83924
35	21287	85	85940
36	21704	86	87828
37	22130	87	89583
38	22569	88	91185
39	23022	89	92632
40	23493	90	93930
41	23979	91	95064
42	24478	92	96052
43	24994	93	96900
44	25529	94	97617
45	26088	95	98195
46	26666	96	98648
47	27260	97	99014
48	27877	98	99292
49	28528	99	99503

Fonte: Carvalho, 1974

TABELA A 19. Brasil – Óbitos acumulados, 1975

1970/1980			
Idade	óbitos acum.	Idade	óbitos acum.
0	0	50	20591
1	7942	51	21215
2	9148	52	21873
3	9702	53	22588
4	10306	54	23336
5	10354	55	24138
6	10421	56	24978
7	10518	57	25879
8	10629	58	26832
9	10738	59	27863
10	10829	60	28952
11	10897	61	30117
12	10953	62	31341
13	11004	63	32671
14	11060	64	34087
15	11130	65	35597
16	11213	66	37210
17	11304	67	38899
18	11403	68	40709
19	11509	69	42636
20	11623	70	44675
21	11743	71	46838
22	11870	72	49093
23	12005	73	51502
24	12149	74	54022
25	12303	75	56650
26	12467	76	59363
27	12641	77	62157
28	12824	78	64982
29	13016	79	67863
30	13218	80	70753
31	13427	81	73632
32	13644	82	76435
33	13871	83	79219
34	14111	84	81897
35	14368	85	84423
36	14638	86	86769
37	14920	87	88925
38	15217	88	90862
39	15536	89	92579
40	15880	90	94084
41	16250	91	95362
42	16642	92	96441
43	17057	93	97333
44	17494	94	98055
45	17954	95	98607
46	18432	96	99018
47	18929	97	99329
48	19451	98	99551
49	20002	99	99708

Fonte: Carvalho e Pinheiro, 1988

TABELA A 20. Brasil – Óbitos acumulados, 1985

1980/1990			
Idade	óbitos acum.	Idade	óbitos acum.
0	0	50	14621
1	4996	51	15220
2	5497	52	15856
3	5671	53	16551
4	5813	54	17280
5	5940	55	18068
6	5980	56	18896
7	6037	57	19790
8	6103	58	20740
9	6170	59	21775
10	6231	60	22875
11	6283	61	24058
12	6332	62	25308
13	6381	63	26676
14	6433	64	28142
15	6493	65	29716
16	6560	66	31407
17	6632	67	33189
18	6709	68	35110
19	6792	69	37170
20	6881	70	39365
21	6976	71	41707
22	7078	72	44163
23	7185	73	46805
24	7299	74	49585
25	7419	75	52499
26	7545	76	55522
27	7677	77	58649
28	7815	78	61820
29	7962	79	65066
30	8120	80	68328
31	8286	81	71581
32	8459	82	74746
33	8643	83	77887
34	8839	84	80901
35	9052	85	83729
36	9278	86	86338
37	9515	87	88715
38	9768	88	90828
39	10040	89	92677
40	10335	90	94274
41	10651	91	95605
42	10985	92	96708
43	11341	93	97598
44	11724	94	98300
45	12137	95	98820
46	12577	96	99195
47	13041	97	99470
48	13534	98	99659
49	14058	99	99787

Fonte: IBGE, 1985

TABELA A 21. Coeficientes da regressão Poisson em função da idade e da escolaridade por tipo de imputação, da variável de escolaridade PPV 1996 (entrevistados com 15 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,622	0,084 ***	0,626	0,084 ***	0,627	0,084 ***	0,626	0,084 ***
50 a 59 anos	1,284	0,080 ***	1,292	0,080 ***	1,293	0,080 ***	1,292	0,080 ***
60 a 69 anos	2,026	0,078 ***	2,037	0,078 ***	2,038	0,078 ***	2,038	0,077 ***
Escolaridade								
Sem escolaridade								
1 a 8 anos de estudo	-0,383	0,048 ***	-0,335	0,045 ***	-0,335	0,045 ***	-0,367	0,045 ***
9 anos e + de estudo	-0,820	0,122 ***	-0,810	0,117 ***	-0,763	0,115 ***	-0,750	0,112 ***
Sem informação	0,341	0,078 ***						
Constante	-5,876	0,073 ***	-5,863	0,072 ***	-5,865	0,072 ***	-5,852	0,072 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996

TABELA A 22. Coeficientes da regressão Poisson em função da idade e da escolaridade por tipo de imputação da variável de escolaridade, PNAD 1996 (entrevistados com 40 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,515	0,028 ***	0,516	0,028 ***	0,516	0,028 ***	0,517	0,028 ***
50 a 59 anos	1,131	0,026 ***	1,132	0,026 ***	1,132	0,026 ***	1,132	0,026 ***
60 a 69 anos	1,866	0,025 ***	1,865	0,025 ***	1,865	0,025 ***	1,865	0,025 ***
Escolaridade								
Sem escolaridade								
1 a 8 anos de estudo	-0,218	0,015 ***	-0,207	0,013 ***	-0,180	0,013 ***	-0,172	0,013 ***
9 anos e + de estudo	-0,473	0,046 ***	-0,445	0,040 ***	-0,510	0,045 ***	-0,412	0,039 ***
Sem informação	0,208	0,015 ***						
Constante	-5,746	0,024 ***	-5,683	0,024 ***	-5,689	0,024 ***	-5,692	0,024 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PNAD 1996

TABELA A 23. Coeficientes da regressão Poisson em função da idade e da escolaridade por tipo de imputação da variável de escolaridade, PPV 1996 (entrevistados com 40 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,514	0,115 ***	0,514	0,115 ***	0,514	0,115 ***	0,514	0,115 ***
50 a 59 anos	1,134	0,106 ***	1,134	0,106 ***	1,134	0,106 ***	1,134	0,106 ***
60 a 69 anos	1,869	0,102 ***	1,869	0,102 ***	1,869	0,102 ***	1,869	0,102 ***
Escolaridade								
Sem escolaridade								
1 a 8 anos de estudo	-0,263	0,057 ***	-0,252	0,054 ***	-0,241	0,054 ***	-0,257	0,054 ***
9 anos e + de estudo	-0,533	0,158 ***	-0,562	0,155 ***	-0,532	0,154 ***	-0,506	0,147 ***
Sem informação	0,129	0,090						
Constante	-5,671	0,097 ***	-5,654	0,097 ***	-5,658	0,097 ***	-5,651	0,097 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996

TABELA A 24. Coeficientes da regressão Poisson em função da idade, da escolaridade e da interação entre idade e escolaridade, por tipo de imputação da variável de escolaridade, PPV 1996 (entrevistados com 15 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,572	0,116 ***	0,572	0,112 ***	0,571	0,112 ***	0,570	0,111 ***
50 a 59 anos	1,205	0,109 ***	1,208	0,106 ***	1,205	0,106 ***	1,203	0,105 ***
60 a 69 anos	1,945	0,106 ***	1,949	0,103 ***	1,944	0,103 ***	1,939	0,102 ***
Escolaridade								
Sem escolaridade								
1 a 8 anos de estudo	-0,558	0,152 ***	-0,473	0,144 ***	-0,480	0,144 ***	-0,521	0,145 ***
9 anos e + de estudo	-1,353	1,317	-1,239	0,376 ***	-1,195	0,368 ***	-1,162	0,357 ***
Sem informação	0,491	0,247						
Idade x Escolaridade								
40 a 49 x 1 a 8 a. estudo	0,116	0,185	0,101	0,175	0,103	0,175	0,106	0,176
40 a 49 x 9 e + a. estudo	0,375	1,400	0,317	0,454	0,306	0,446	0,304	0,432
40 a 49 x sem inf.	-0,063	0,300						
50 a 59 x 1 a 8 a. estudo	0,201	0,175	0,162	0,165	0,169	0,165	0,176	0,166
50 a 59 x 9 e + a. estudo	0,597	1,376	0,490	0,434	0,489	0,426	0,475	0,412
50 a 59 x sem inf.	-0,140	0,284						
60 a 69 x 1 a 8 a. estudo	0,232	0,171	0,171	0,161	0,183	0,161	0,199	0,162
60 a 69 x 9 e + a. estudo	0,693	1,365	0,540	0,425	0,559	0,417	0,518	0,404
60 a 69 x sem inf.	-0,251	0,277						
constante	-5,810	0,095 ***	-5,792	0,092 ***	-5,789	0,093 ***	-5,774	0,092 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996

TABELA A 25. Coeficientes da regressão Poisson em função da idade, da escolaridade e da interação entre idade e escolaridade, por tipo de imputação da variável de escolaridade, PNAD 1996 (entrevistados com 40 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,513	0,042 ***	0,505	0,035 ***	0,504	0,036 ***	0,506	0,036 ***
50 a 59 anos	1,126	0,039 ***	1,117	0,033 ***	1,115	0,033 ***	1,118	0,034 ***
60 a 69 anos	1,859	0,038 ***	1,852	0,031 ***	1,849	0,032 ***	1,853	0,032 ***
Escolaridade								
Sem escolaridade								
1 a 8 anos de estudo	-0,245	0,058 ***	-0,243	0,050 ***	-0,219	0,049 ***	-0,202	0,049 ***
9 anos e + de estudo	-0,504	0,174 **	-0,465	0,153 **	-0,549	0,170 **	-0,440	0,149 **
Sem informação	0,214	0,057 ***						
Idade x Escolaridade								
40 a 49 x 1 a 8 a. estudo	0,025	0,069	0,032	0,060	0,032	0,059	0,027	0,059
40 a 49 x 9 e + a. estudo	0,043	0,210	0,036	0,184	0,051	0,204	0,044	0,179
40 a 49 x sem inf.	-0,017	0,069						
50 a 59 x 1 a 8 a. estudo	0,034	0,064	0,042	0,056	0,043	0,055	0,035	0,055
50 a 59 x 9 e + a. estudo	0,036	0,195	0,025	0,171	0,046	0,189	0,034	0,167
50 a 59 x sem inf.	-0,012	0,064						
60 a 69 x 1 a 8 a. estudo	0,027	0,062	0,039	0,053	0,043	0,052	0,032	0,052
60 a 69 x 9 e + a. estudo	0,023	0,186	0,011	0,163	0,032	0,181	0,019	0,160
60 a 69 x sem inf.	-0,001	0,061						
constante	-5,741	0,035 ***	-5,670	0,029 ***	-5,675	0,030 ***	-5,681	0,030 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PNAD 1996

TABELA A 26. Coeficientes da regressão Poisson em função da idade, da escolaridade e da interação entre idade e escolaridade, por tipo de imputação da variável de escolaridade, PPV 1996 (entrevistados com 40 anos e mais)

Variável	Educação sem imputar		Imputação estocástica		Imputação por hotdeck		Imputação com algoritmo EM	
	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.	Coef.	D. P.
Idade								
30 a 39 anos								
40 a 49 anos	0,504	0,150 ***	0,505	0,143 ***	0,503	0,144 ***	0,504	0,144 ***
50 a 59 anos	1,124	0,139 ***	1,126	0,133 ***	1,124	0,133 ***	1,125	0,134 ***
60 a 69 anos	1,862	0,133 ***	1,864	0,127 ***	1,862	0,128 ***	1,866	0,128 ***
Escolaridade								
Sem escolaridade	-0,294	0,216	-0,279	0,205	-0,272	0,204	-0,281	0,204
1 a 8 anos de estudo	-1,324	31,770	-1,244	26,208	-1,190	22,959	-1,020	18,536
9 anos e + de estudo	0,068	0,348						
Sem informação								
Idade x Escolaridade								
40 a 49 x 1 a 8 a. estudo	0,037	0,260	0,033	0,247	0,036	0,246	0,035	0,246
40 a 49 x 9 e + a. estudo	0,685	32,861	0,582	27,388	0,588	23,088	0,440	18,660
40 a 49 x sem inf.	0,031	0,419						
50 a 59 x 1 a 8 a. estudo	0,036	0,241	0,031	0,229	0,036	0,228	0,032	0,228
50 a 59 x 9 e + a. estudo	0,769	31,838	0,661	26,275	0,634	23,026	0,505	18,600
50 a 59 x sem inf.	0,053	0,388						
60 a 69 x 1 a 8 a. estudo	0,027	0,231	0,022	0,219	0,029	0,219	0,017	0,219
60 a 69 x 9 e + a. estudo	0,780	31,810	0,670	26,247	0,648	22,998	0,501	18,573
60 a 69 x sem inf.	0,066	0,372						
Constante	-5,664	0,124 ***	-5,648	0,119 ***	-5,650	0,119 ***	-5,646	0,119 ***

*** P>0,01; ** P>0,05; * P>0,1

Fonte: IBGE, PPV 1996