

# Reconstrução de histórias de nascimentos a partir de dados censitários: uma análise comparativa de duas metodologias

Adriana de Miranda-Ribeiro\*  
Eduardo L. G. Rios-Neto\*\*  
José Alberto Magno de Carvalho\*\*\*

*Este trabalho tem como objetivo principal aplicar duas metodologias de reconstrução de histórias de nascimentos – uma desenvolvida no Cedeplar e outra de autoria de Luther e Cho (1988) – aos dados censitários brasileiros de 2000, a fim de comparar e testar a confiabilidade dos resultados e a praticidade de cada uma das metodologias. Pretende-se, com isso, determinar a metodologia mais adequada ao caso brasileiro e buscar um melhor aproveitamento das bases de dados disponíveis, no que concerne a formas alternativas de estimação da fecundidade. Para tanto, foram utilizados os microdados do Censo Demográfico de 2000. Os dados de mortalidade foram retirados de Carvalho (1974, 1978), Carvalho e Pinheiro (1986) e do relatório da pesquisa Pronex/Cedeplar (CEDEPLAR, 1999).*

**Palavras-chave:** Histórias de nascimentos. Fecundidade. Censo Demográfico.

## Introdução e objetivos

Uma história de nascimentos pode ser entendida como a história da fecundidade das mulheres ao longo de um período passado. As histórias de nascimentos baseiam-se em um conjunto de dados que são, em geral, coletados em uma única pesquisa e que possibilitam a alocação dos nascimentos no momento em que se realizaram. As informações sobre o momento do nascimento dos filhos e a idade da mulher na data da pesquisa permitem que se calcule a idade da mulher, quando do nascimento de cada filho, e, assim, se estimem medidas de fecundidade para períodos anteriores ao da pesquisa. A eficácia das estimativas depende da acurácia dos dados, de que a migração e a mortalidade não sejam seletivas por nível de fecundidade e de que os possíveis erros ocorram na mesma propor-

ção em todos os subgrupos, no caso de estudos estratificados (BRASS, 1974).

A ausência de boas histórias de nascimentos e a constatação de que o banco de dados formado pelas mães e seus filhos próprios – resultado da aplicação do Método dos Filhos Próprios (CHO; RETHERFORD; CHOE, 1986) aos dados do censo – é uma fonte preciosa de informações relativas ao nascimento desses filhos motivaram o desenvolvimento de metodologias de reconstrução de histórias de nascimentos. Este trabalho objetiva aplicar duas metodologias de reconstrução de histórias de nascimentos – uma desenvolvida no Cedeplar e outra elaborada por Luther e Cho (1988) – aos dados censitários brasileiros de 2000, a fim de comparar e testar a confiabilidade dos resultados e a praticidade de cada uma delas. Pretende-se, com isso, determinar a metodologia mais adequada ao caso bra-

\* Gestora de Ensino e Pesquisa da Fundação João Pinheiro/MG.

\*\* Professor titular do departamento de Demografia e do Cedeplar/UFMG.

\*\*\* Professor titular do departamento de Demografia e do Cedeplar/UFMG.

sileiro e buscar um melhor aproveitamento das bases de dados disponíveis, no que concerne a formas alternativas de estimação da fecundidade. Além das taxas de fecundidade específicas e totais, as histórias de nascimentos possibilitam o cálculo de uma série de outras medidas, de período e de coorte, o que permite um panorama mais detalhado das transformações na fecundidade.

As duas metodologias descritas neste trabalho têm como base a alocação de filhos às mães e representam um avanço em relação ao Método dos Filhos Próprios (MFP) no tratamento dado aos filhos omitidos (não-alocados ou não-sobreviventes). Enquanto o MFP redistribui os filhos não-alocados proporcionalmente entre as mulheres e recupera os não-sobreviventes por meio da projeção retrospectiva dos filhos sobreviventes, as duas metodologias de reconstrução de histórias de nascimentos recuperam os filhos omitidos sob a ótica das mães. Isso significa que as duas metodologias de reconstrução de histórias de nascimentos não “procuram uma mãe” para os filhos não-alocados ou não-sobreviventes, mas sim buscam atribuir uma idade aos filhos omitidos de cada mulher, sem vinculá-los a qualquer criança enumerada no censo.

O objetivo das duas metodologias é tornar completa a história de nascimentos das mulheres que têm entre 15 e 64 anos de idade, enumeradas no censo ou pesquisa.

A metodologia desenvolvida no Cedeplar, denominada HNM, é baseada em um processo de pareamento e cobre um período anterior ao censo de quinze anos. A metodologia elaborada por Luther e Cho (1988), denominada HNP, baseia-se em um processo probabilístico e cobre todo o período reprodutivo das mulheres; no entanto, serão analisados somente os quinze anos anteriores ao da realização do censo.

Assim como nas histórias de nascimentos tradicionais, parte-se do pressuposto de que migração e mortalidade não são seletivas segundo o nível de fecundidade e que eventuais erros não são seletivos, em relação às características analisadas.

As bases de dados utilizadas neste trabalho foram os microdados do Censo Demográfico de 2000 e os resultados da aplicação do Método dos Filhos Próprios publicados em Miranda-Ribeiro (2007). Os dados sobre mortalidade foram retirados de Carvalho (1974, 1978), Carvalho e Pinheiro (1986) e do relatório da pesquisa Pronex/Cedeplar (CEDEPLAR, 1999).

### **Metodologia de reconstrução de histórias de nascimentos baseada em um processo de pareamento (*matching*)**

A metodologia de reconstrução de histórias de nascimentos baseada no processo de pareamento (*matching*), desenvolvida no Cedeplar, tem como objetivo tornar completa a história de nascimentos das mulheres entre 15 e 64 anos de idade, para os 15 anos anteriores ao censo ou pesquisa utilizado para sua aplicação. Em linhas gerais, o procedimento consiste em buscar, no universo de histórias de nascimentos completas, aquela que mais se aproxima da história de nascimentos parcial (incompleta), com base na comparação de algumas variáveis. A história de nascimentos completa que tiver a maior associação com a parcial determinará as variáveis que serão imputadas, de modo a torná-la completa.

Para construir as histórias de nascimentos – completas e parciais –, utiliza-se o procedimento de alocação de filhos às mães.<sup>1</sup> A idade e o sexo de cada filho alocado definirão o ano em que a mulher teve aquele filho e o sexo do filho nascido naquele ano. O período de reconstrução das histórias de nascimentos é limitado em 15 anos, em virtude do mesmo pressuposto adotado no MFP: o de que a maioria dos filhos que têm até 14 anos de idade reside com a mãe ou é enumerada no mesmo domicílio. Esse pressuposto é importante, também, na definição do conceito de história de nascimentos completa, para mulheres com 30 anos ou mais de idade.

Para definir as histórias de nascimentos que irão compor o banco de dados das histórias de nascimentos completas e parciais,

<sup>1</sup> Ver Miranda-Ribeiro (2007).

as mulheres são divididas em dois grupos distintos. O primeiro é formado por aquelas com idades entre 15 e 29 anos e o segundo pelas mulheres de 30 a 64 anos. Supondo que o período reprodutivo tem início aos 15 anos de idade, as mulheres do primeiro grupo somente podem ter filhos de 0 a 14 anos de idade. Assim, suas histórias de nascimentos serão completas se todos os seus filhos nascidos vivos estiverem vivos na data da pesquisa e forem alocados. Para as mulheres do segundo grupo, a definição das histórias de nascimentos completas é diferente, porque elas podem ter filhos de 15 anos ou mais de idade e, portanto, de acordo com o pressuposto, filhos vivos que residam em outro domicílio.

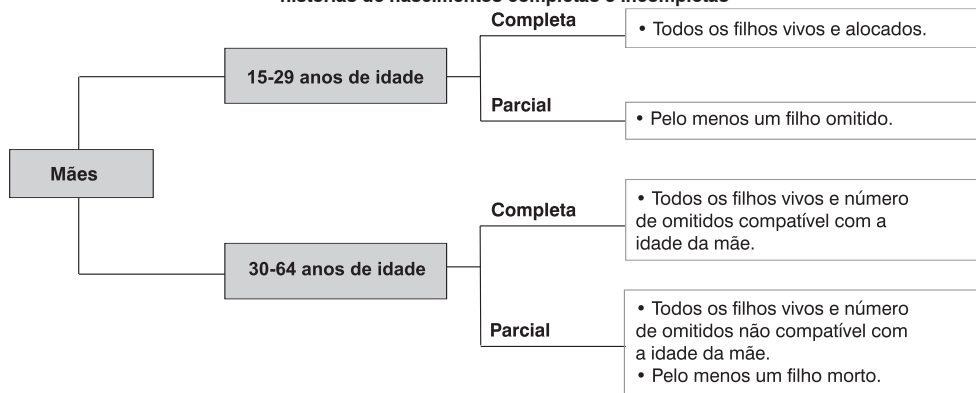
Desse modo, para o segundo grupo, são consideradas histórias de nascimentos completas aquelas das mulheres cujos filhos nascidos vivos estejam vivos na data da pesquisa, independente do número de filhos alocados, desde que o número de filhos não-alocados (filhos omitidos) seja compatível com a idade da mulher. Isso ocorre porque se supõe que os filhos omitidos tenham 15 anos ou mais de idade. Assim, uma mulher de 30 anos pode ter somente um filho omitido e este tem que ter 15 anos de idade; uma mulher de 31 anos pode ter dois filhos omitidos, um de 15 e outro de 16 anos; e assim sucessivamente.<sup>2</sup> Dessa forma, uma mulher de 30 anos que tenha dois ou mais filhos

omitidos tem pelo menos um filho menor de 15 anos omitido e, por isso, sua história de nascimentos não é considerada completa.

A definição das histórias de nascimentos parciais também é diferente para cada um dos dois grupos de idade considerados. Para mulheres do primeiro grupo, isto é, entre 15 e 29 anos de idade, serão parciais as histórias de nascimentos daquelas que tiverem o número de filhos alocados menor do que o de filhos nascidos vivos. Essa condição inclui as mulheres que têm, pelo menos, um filho morto e as que possuem todos os filhos vivos, mas para as quais não foi possível, por algum motivo, a alocação de todos.

As mulheres do segundo grupo com histórias de nascimentos parciais são aquelas que têm, pelo menos, um filho nascido vivo que já morreu ou aquelas que têm todos os filhos vivos e número de filhos omitidos incompatível com sua idade (conforme descrição anterior). Para completar o banco de dados das histórias de nascimentos parciais, acrescentam-se as mulheres que tiveram pelo menos um filho nascido vivo, mas para as quais não foi possível a alocação de filho algum, seja porque as relações de parentesco com o chefe da família não permitiram, seja porque filho e mãe não residiam no mesmo domicílio, seja porque a variável “idade do último filho nascido vivo” não existia ou era indeterminada. O Diagrama 1 apresenta essas relações de forma esquemática.

**DIAGRAMA 1**  
**Reconstrução de história de nascimentos baseada no processo de pareamento (*matching*): definição de histórias de nascimentos completas e incompletas**



Fonte: Elaboração própria.

<sup>2</sup> Na aplicação desta regra, é desconsiderada a ocorrência de nascimentos múltiplos.

Nos bancos de dados das histórias de nascimentos – completas e parciais – cada linha representa uma mulher e contém as variáveis censitárias disponíveis e selecionadas, além daquelas construídas a partir da idade e do sexo dos filhos alocados, que indicam se a mulher teve ou não um filho em determinado ano e o sexo do filho tido. Desse modo, os bancos de dados estão prontos para a comparação.

A comparação é feita utilizando-se o *software Reclink*, desenvolvido por Camargo Jr. e Coeli (2000). Basicamente, o *software* compara as duas bases de dados e classifica, para cada história de nascimentos parcial, as histórias de nascimentos completas, segundo um escore definido em função dos pareamentos corretos. A partir dos resultados encontrados pelo *Reclink*, determina-se a história de nascimentos com maior escore, utilizada para imputação das variáveis faltantes.

Para iniciar o procedimento de comparação, é necessário definir dois tipos de variáveis: as de blocagem e as de pareamento. As variáveis de blocagem têm por objetivo dividir os bancos de dados em blocos de registros lógicos e homogêneos, no sentido de otimizar o procedimento de comparação. Isto ocorre porque cada história de nascimentos completa do bloco é um possível par para a história de nascimentos parcial do mesmo bloco. Assim, o número total de possíveis pares dentro de cada bloco será igual ao produto entre o número de histórias de nascimentos parciais e o de histórias de nascimentos completas. A blocagem permite que as bases de dados sejam divididas em blocos mutuamente exclusivos, sendo as comparações limitadas aos registros pertencentes a um mesmo bloco. Os blocos são constituídos, também, de forma a aumentar a probabilidade de os registros neles contidos representarem pares verdadeiros (CAMARGO Jr.; COELI, 2000).

Já as variáveis de pareamento são utilizadas para a comparação entre cada história de nascimentos parcial e as histórias de nascimentos completas do mesmo bloco. Para cada campo acertado, é computado um escore. A soma (ponderada ou não – e neste caso, não ponderada) dos escores de todos

os campos fornece o valor total, que é o escore atribuído ao relacionamento entre cada par de registros. A não ponderação garante que todos os campos tenham o mesmo peso na determinação do escore final.

Uma vez efetuada a comparação entre as histórias de nascimentos, o programa *Reclink* gera um arquivo que apresenta, para cada história de nascimentos parcial, a lista de todos os possíveis pares e seus respectivos escores. Todas as histórias de nascimentos completas pertencentes ao mesmo bloco são possíveis pares. O par mais adequado – ou par verdadeiro – é aquele que tem o maior escore, ou seja, aquele no qual a checagem das variáveis de comparação gerou um maior número de pareamentos corretos. Há casos em que o escore máximo aparece mais de uma vez. Quando isso ocorre, vários procedimentos podem ser utilizados para a “escolha” do par. Uma possibilidade, adotada neste trabalho, é a seleção aleatória.

Por outro lado, existem casos em que não se encontra um possível par para a história de nascimentos parcial. Isso acontece quando ela não pertence a nenhum bloco criado pelas variáveis de blocagem. Nestas situações, repetiu-se o procedimento com novos parâmetros de blocagem. Se ainda assim não foi encontrado um possível par para alguma história de nascimentos incompleta, optou-se por não incorporá-la ao banco de dados final. Dessa forma, mulheres com história de nascimentos incompleta e não completada pelo procedimento adotado são retiradas da amostra, o que não influencia no resultado global das estimativas, pois representam uma parcela residual das histórias de nascimentos.

A disponibilidade de variáveis do Censo Demográfico de 2000 determinou os parâmetros de blocagem e pareamento utilizados no procedimento de reconstrução das histórias de nascimentos. Neste trabalho, optou-se pela utilização de variáveis demográficas básicas e de reprodução na determinação dos parâmetros de blocagem e pareamento, buscando um procedimento não enviesado, no que concerne às características socioeconômicas. Como parâmetros de blocagem, foram utilizadas: UF

de residência, idade, parturição e idade do último filho nascido vivo. Como parâmetros de pareamento, foram utilizadas as variáveis indicativas de nascimento de filho no ano (15 variáveis), aquelas indicativas do sexo do filho nascido no ano (15 variáveis) e o total de filhos nascidos mortos.

Além de otimizar o procedimento, os parâmetros de blocagem garantem que os possíveis pares sejam formados entre mulheres da mesma unidade da federação, de mesma idade, mesma parturição e que tiveram o último filho no mesmo ano. Quando é necessária a repetição do procedimento, no caso em que uma história incompleta não encontra algum par, é retirada a variável de blocagem "idade da mulher".

### **Metodologia de reconstrução de histórias de nascimentos baseada em um processo probabilístico**

Esta metodologia foi desenvolvida por Luther e Cho (1988) e trata da reconstrução da história de nascimentos dos filhos das mulheres, a partir das informações contidas em censos e pesquisas domiciliares. A metodologia é considerada pelos autores uma extensão do MFP, descrito anteriormente, porque está baseada na alocação dos filhos às respectivas mães, mas apresenta avanços no modo de tratar os filhos omitidos (filhos não-alocados ou que não sobreviveram à data da pesquisa).

A história de nascimentos é construída a partir das informações sobre as idades da mãe e dos filhos na data da pesquisa, por meio da alocação no tempo de todos os nascimentos. Se todos os filhos nascidos vivos da mulher estiverem vivos e residirem no mesmo domicílio que ela, a história de nascimentos é completa. Caso a mulher tenha pelo menos um filho nascido vivo que tenha morrido ou que não resida com ela (ou seja, um filho omitido), a história de nascimentos é parcial.

Em linhas gerais, a metodologia proposta – HNP – busca reconstruir, a partir da história de nascimentos parcial das mulheres que não têm todos os seus filhos vivos e residentes no seu domicílio, na data da pesquisa, uma história de nascimentos

completa. A reconstrução da história de nascimentos é feita por meio de um processo probabilístico, que atribui a cada mulher a probabilidade de ter tido um filho, não-alocado ou não-sobrevivente na data da pesquisa,  $j$  anos antes da pesquisa.

Para descrever a metodologia, é conveniente que se defina a idade do filho morto, na data do censo ou pesquisa, como sendo a idade completa que este filho teria se estivesse vivo. Assim, é possível satisfazer a ideia de uma história de nascimentos como um conjunto de idades de crianças, se nenhuma tiver morrido, no momento do censo ou pesquisa, o que permite determinar a idade das mães no momento do nascimento dessas crianças. Com essa definição, também conceitua-se um conjunto de probabilidades de nascimento, específicas por idade (probabilidades específicas de nascimento), como um conjunto de probabilidades de que uma dada criança, não-alocada ou morta, tenha (ou tivesse) completado uma determinada idade, na data da pesquisa.

A determinação das funções de probabilidade parte da definição das funções de fecundidade das coortes envolvidas na análise, que representam a fecundidade passada das mulheres de cada idade. As funções de fecundidade das coortes podem ser obtidas de fontes de dados existentes ou da aplicação do MFP aos dados da pesquisa. Neste caso, é necessário transformar as taxas de período em taxas de coorte. Como o MFP fornece uma série de taxas específicas de fecundidade por um período de 15 anos anteriores ao censo, é necessário completar a fecundidade das coortes que fizeram 15 anos de idade antes do início da série – ou seja, das coortes de mulheres que, na data do censo, tinham 30 anos ou mais. A forma mais simples de completar a fecundidade dessas coortes é mantendo constante a fecundidade do primeiro ano da série, durante um período de 20 anos anterior à série.

A probabilidade de uma mulher de determinada idade ter tido um filho aos  $x$  anos de idade é encontrada pela integração matemática da curva da função de fecundidade de sua coorte, entre as idades  $(x - 1/2)$  e  $(x + 1/2)$ . Essa probabilidade é ajustada segundo a condição de omissão do filho: não-aloca-

do, mas vivo, ou não-sobrevivente na data do censo ou pesquisa. Assim, apesar de baseadas na mesma função de fecundidade, a probabilidade de atribuir uma idade  $j$  a um filho não-alocado é diferente da de atribuir a mesma idade  $j$  a um filho morto.

As probabilidades devem levar em conta, também, a idade dos filhos próprios (alocados) de cada mulher. É pouco provável – embora não seja impossível – que os filhos não-alocados ou mortos de uma mulher tenham a mesma idade que seus filhos próprios (caso de gêmeos, trigêmeos, etc.). Partindo dessa premissa, a metodologia pressupõe que, durante um intervalo de 2,5 anos,<sup>3</sup> centrado na idade exata em que a mulher teve um filho próprio, a probabilidade de ela ter tido um outro filho (omitido) seja igual a zero.

Para a modelagem das curvas de fecundidade, Luther e Cho (1988) utilizam o modelo descrito por Luther (1982), que emprega uma forma particular da curva de Pearson tipo III, descrita pela equação 1:

$$\tilde{f}(x) = (x - s)^2 * \exp\left[\frac{-2(x-s)}{m}\right] \quad (1)$$

Os parâmetros  $s$  e  $m$  são calculados segundo as fórmulas das equações 2 e 3:

$$s = \frac{\{[12,9210 + 2,04642 * \ln(r - 0,535)] * \sqrt{r} + 7,20 - 0,5\bar{a}\}}{\sqrt{r} - 0,5} \quad (2)$$

$$m = \frac{2(\bar{a} - s)}{3} \quad (3)$$

Nestas equações,  $x$  representa a idade da mulher;  $s$  é a idade na qual a coorte desta mulher iniciou sua vida reprodutiva;  $r$  é a razão entre as taxas específicas de fecundidade dos grupos etários 20-24 anos e 15-19 anos, da coorte da mulher; e  $\bar{a}$  é a idade média da fecundidade da coorte. A idade  $x$  deve satisfazer a condição  $s \leq x \leq u + 0,5$ , sendo que  $u$  é o menor valor entre a idade da mulher na data do censo e 46 anos de idade.

As probabilidades utilizadas na atribuição das idades no momento da pesquisa são derivadas destas funções  $\tilde{f}(x)$ . A curva da função é utilizada para encontrar a probabilidade relativa de ocorrência do evento  $B(j)$ ,

isto é, de que a mulher tenha tido um filho  $j$  anos (completos) antes da data da realização da pesquisa. Sendo  $g$  a idade completa da mulher na data da pesquisa, essa probabilidade é encontrada para todas as idades  $j$  do filho, ou seja, para todas as idades  $g-j$  da mulher, nas quais  $[s] \leq g-j \leq u$ ,  $u = \text{Min}(g, 46)$  e  $[s]$  é o maior inteiro menor que  $s$ . Assim, para cada mulher, são determinadas as probabilidades  $P[B(j)]$ ,  $j = g-u, g-u+1, \dots, g-[s]$  de ela ter tido um filho que, na data da pesquisa, teria  $j$  anos de idade. O valor de  $P[B(j)]$  é obtido pela integração da função  $\tilde{f}(x)$ , no intervalo  $(g-j-1/2)$  e  $(g-j+1/2)$ , sendo que  $\tilde{f}(x)$  é nula no intervalo  $(g-j-1, 25)$  a  $(g-j+1, 25)$  para cada idade  $j$  de um filho próprio da mulher; o mesmo é feito para cada idade  $j$  atribuída a um filho omitido da mulher. Assim, para cada idade  $j$  de um filho,  $j = g-u, g-u+1, \dots, g-[s]$ , a área sob a curva de  $\tilde{f}(x)$  é integrada entre  $g-j-1/2$  e  $g-j+1/2$ , para se obter a probabilidade relativa  $P[B(j)]$  de que um filho tenha nascido  $j$  anos antes do censo ou pesquisa. Formalmente, escreve-se  $P[B(j)]$  em função dos parâmetros da função de fecundidade:

$$P[B(j)] = H(g-j+1/2) - H(g-j-1/2) \quad (4)$$

para a qual

$$H(x) = 0,5m \left[ 0,5m^2 - \exp\left[\frac{-2(x-s)}{m}\right] \{ [x-s]^2 + m(x-s) + 0,5m^2 \} \right] \quad (5)$$

Uma vez encontradas as probabilidades de nascimento, podem-se derivar as probabilidades que serão utilizadas na atribuição da idade dos filhos omitidos de cada mulher, na data do censo ou pesquisa. Suponha-se um evento  $D$ , em que uma criança (de qualquer idade) tenha morrido antes da data da realização do censo ou pesquisa. Generalizando (sem suposição de independência), a probabilidade relativa de ter nascido  $j$  anos antes do censo ou pesquisa e ter morrido antes da sua realização é dada por:

$$P[B(j) \cap D] = P[B(j)]P[D|B(j)] \quad (6)$$

para  $j = g-u, g-u+1, \dots, g-[s]$ , na qual  $P[D|B(j)]$  é a probabilidade condicional de que uma criança nascida  $j$  anos antes do censo ou pesquisa não tenha sobrevivido à sua realização.

<sup>3</sup> Segundo os autores, este valor corresponde a um intervalo que não difere significativamente de um intervalo típico entre nascimentos.

Suponha-se, agora, o evento N, em que uma criança (de qualquer idade) seja não-própria (isto é, sobrevivente, mas não alocada) no momento do censo ou pesquisa. E seja D o evento definido anteriormente. Novamente generalizando, a probabilidade relativa de ter nascido  $j$  anos antes da realização do censo ou pesquisa e ser um filho não-próprio é dada por:

$$\begin{aligned} P[B(j) \cap N] &= P[B(j) \cap \bar{D} \cap N] \\ P[B(j) \cap N] &= P[B(j)]P[\bar{D}|B(j)]P[N|\bar{D} \cap B(j)] \\ P[B(j) \cap N] &= P[B(j)](1 - P[D|B(j)])P[N|\bar{D} \cap B(j)] \quad (7) \end{aligned}$$

na qual  $j = g - u, g - u + 1, \dots, g - [s]$ ;  $\bar{D}$  é o complemento de  $D$ , ou seja, a probabilidade de a criança sobreviver desde o nascimento até a data do censo ou pesquisa. Nesta equação,  $P[N|\bar{D} \cap B(j)]$  é a probabilidade condicional de que uma criança de  $j$  anos de idade no momento do censo ou pesquisa seja um filho não-próprio. As probabilidades  $P[B(j)]$  e  $P[D|B(j)]$  são calculadas conforme descrito anteriormente. A probabilidade condicional  $P[N|\bar{D} \cap B(j)]$  é obtida do censo ou pesquisa pela razão entre o total de filhos não-próprios e o total de filhos sobreviventes (filhos não-próprios mais filhos próprios), para cada idade  $j$ .

Para cada mulher, conforme descrito anteriormente, um processo randômico é utilizado para atribuir uma idade a cada filho omitido, utilizando as probabilidades calculadas pela equação 6, se a omissão for por morte, e pela equação 7, se a omissão for por não-alocação. O primeiro filho é escolhido aleatoriamente e, utilizando as funções de probabilidade, a ele é atribuída uma idade  $j$ . Em seguida, as probabilidades são recalculadas, considerando-se a idade  $j$ , atribuída a esse primeiro filho, já que a probabilidade de a mulher ter outro filho no intervalo  $g - j \pm 5/4$  é igual a zero. Isso implica uma diminuição da probabilidade de a mulher ter um outro filho de idades  $j - 1$  e  $j + 1$ . Na verdade, exceto para  $P[B(j)]$ ,  $P[B(j - 1)]$  e  $P[B(j + 1)]$ , as probabilidades permanecem as mesmas, quando da atribuição de uma idade a um filho omitido. Recalculadas as probabilidades, o processo se repete, até que uma idade tenha sido atribuída a todos os filhos omitidos da mulher.

### Comparação entre HNM e HNP: em busca de uma metodologia de reconstrução de histórias de nascimentos a partir de dados censitários

As duas metodologias de reconstrução de histórias de nascimentos foram aplicadas aos dados do Censo Demográfico de 2000. Além de ter como objetivo testar a metodologia de reconstrução de histórias de nascimentos desenvolvida no Cedeplar, a partir da comparação com uma metodologia já testada em outras bases de dados (LUTHER; CHO, 1988), buscou-se determinar a metodologia mais adequada ao caso brasileiro no que diz respeito tanto à exequibilidade quanto à qualidade dos resultados. Embora as metodologias de reconstrução de histórias de nascimentos não tenham como objetivo principal a construção de séries de TFT – para isso, o MFP seria mais que suficiente –, a determinação destas séries deve ser o primeiro passo para a definição da qualidade dos dados. Se as TFT produzidas pelas histórias de nascimentos estiverem próximas dos valores esperados, é maior a chance de que outras medidas calculadas pelas histórias de nascimentos também estejam.

Para testar as séries de TFT produzidas pelas duas metodologias de reconstrução de histórias de nascimentos, foi feita a comparação com os resultados alcançados pela aplicação do MFP (retirados de MIRANDA-RIBEIRO, 2007). As séries de TFT obtidas pela aplicação do MFP foram utilizadas como padrão – ou valores esperados – porque produzem resultados confiáveis. Essa confiabilidade advém, principalmente, do fato de as estimativas serem geradas no nível agregado e, em contextos de baixa mortalidade, de as estimativas tenderem a se aproximar do valor real da fecundidade, já que a correção é pequena. Além da comparação gráfica visual, foi aplicado um teste de médias, que avaliou as médias das séries de TFT dos pares HNM-MFP e HNP-MFP, determinando se a diferença entre elas era ou não estatisticamente significativa. Foi utilizado o teste “t” de Student, que compara médias de amostras independentes, com nível de significância de 5%.

## Resultados

Os Gráficos 1 a 6 apresentam as séries de TFT, para o Brasil e grandes regiões, geradas pelas três metodologias – HNM, HNP e MFP. Para os três casos, as séries foram obtidas pela soma de nascimentos e mulheres das unidades da federação que compõem as regiões. O resultado do teste de médias é mostrado na Tabela 1.

No Gráfico 1, são apresentadas as séries de TFT para o Brasil, no período 1986-2000. Observa-se que a série da HNM se ajusta à do MFP durante praticamente todo o período, principalmente a partir de 1989. A série da HNP, ao contrário, não se ajusta à do MFP durante grande parte do período de análise, permanecendo em um nível inferior às demais curvas até o final da década de 1990. A análise visual é comprovada pelo teste-t (Tabela 1), que indica que a diferença entre as médias das séries é estatisticamente significativa para o par HNP-MFP e não é para o par HNM-MFP.

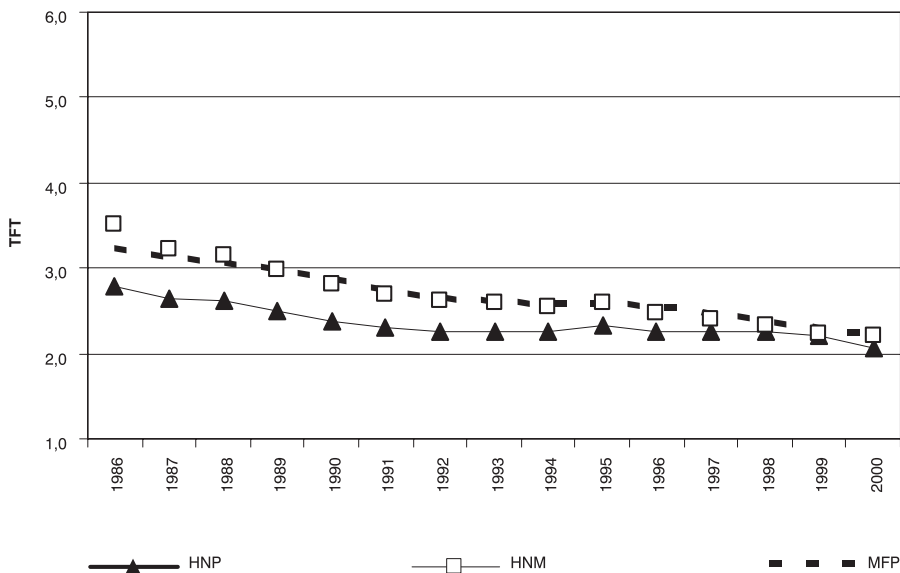
Para a Região Norte (Gráfico 2), a diferença entre a série da HNP e as outras duas é mais visível do que para o Brasil. Enquanto as séries HNM e MFP se ajustam bem, exce-

to no primeiro ano do período, a diferença para a série da HNP é de praticamente um filho nos primeiros anos, diminui ao longo do período e se aproxima bastante nos três últimos anos da série. Do mesmo modo que observado para o Brasil, o teste-t (Tabela 1) indica que a diferença entre as médias do par HNM-MFP não é estatisticamente significativa, ao contrário do que ocorre com o par HNP-MFP.

Resultado semelhante ao da Região Norte foi encontrado para o Nordeste, conforme mostra o Gráfico 3. Observam-se grande diferença entre a série da HNP e as da HNM e MFP no início do período e diminuição desta diferença ao longo do tempo. Neste caso, no entanto, não ocorre uma sobreposição das curvas ao final do período. A curva da HNM não se ajusta bem à do MFP nos primeiros anos, mas há uma sobreposição a partir de 1989. O teste-t (Tabela 1) mostra o mesmo resultado obtido para o Brasil e Região Norte.

Para a Região Sudeste (Gráfico 4), embora também não se observe sobreposição das curvas durante o período completo, as diferenças entre a série da HNP e as demais são menores durante todo o período de

GRÁFICO 1  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Brasil – 1986-2000



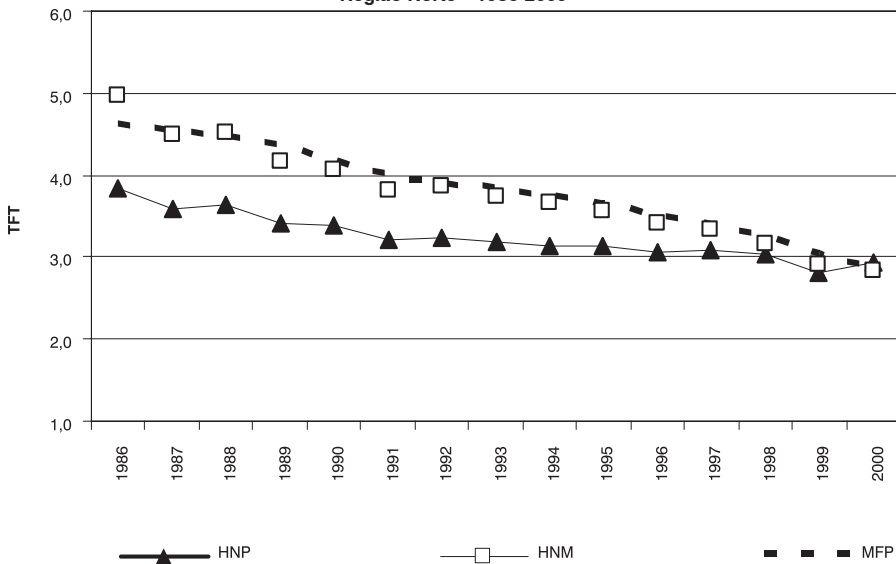
Fonte: Censo Demográfico de 2000.



análise. Assim como ocorre para as regiões analisadas anteriormente, a diferença diminui ao longo do tempo, mas, neste caso, há praticamente uma sobreposição a partir de

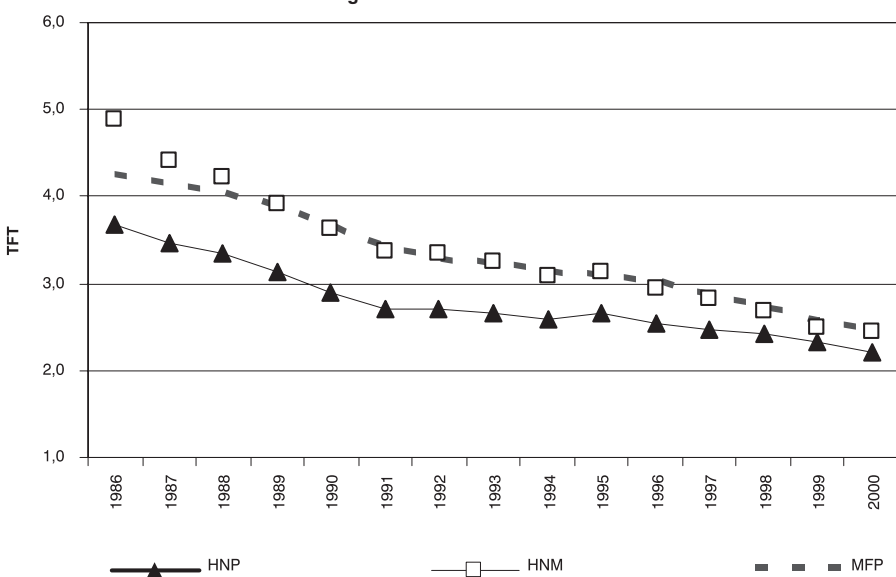
1996. As séries HNM e MFP se sobrepõem, exceto no primeiro ano da análise. O teste-t indica que a diferença entre as médias do par HNP-MFP é estatisticamente significan-

**GRÁFICO 2**  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Região Norte - 1986-2000



Fonte: Censo Demográfico de 2000.

**GRÁFICO 3**  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Região Nordeste - 1986-2000



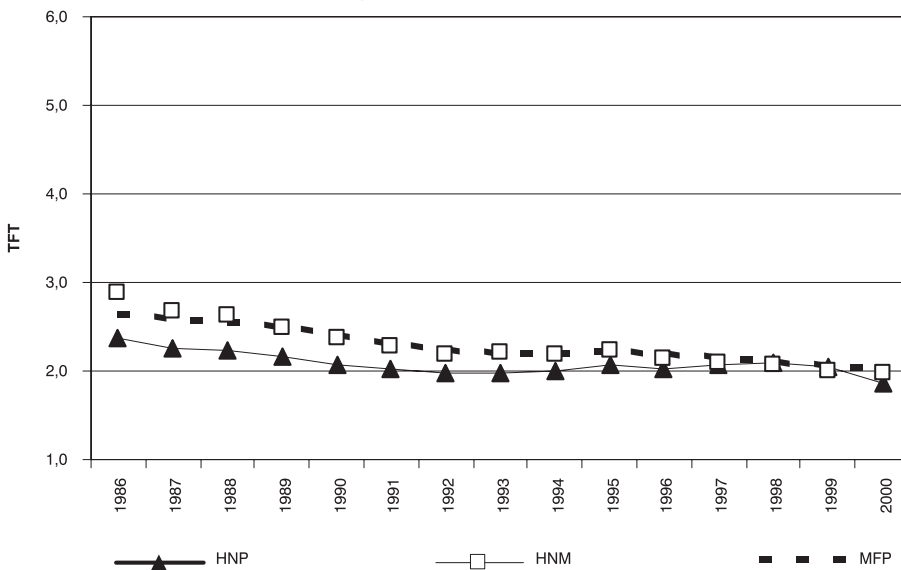
Fonte: Censo Demográfico de 2000.

te, ao contrário do que ocorre com o par HNM-MFP.

O Gráfico 5 mostra que, na Região Sul, a série gerada pela aplicação da HNM aos

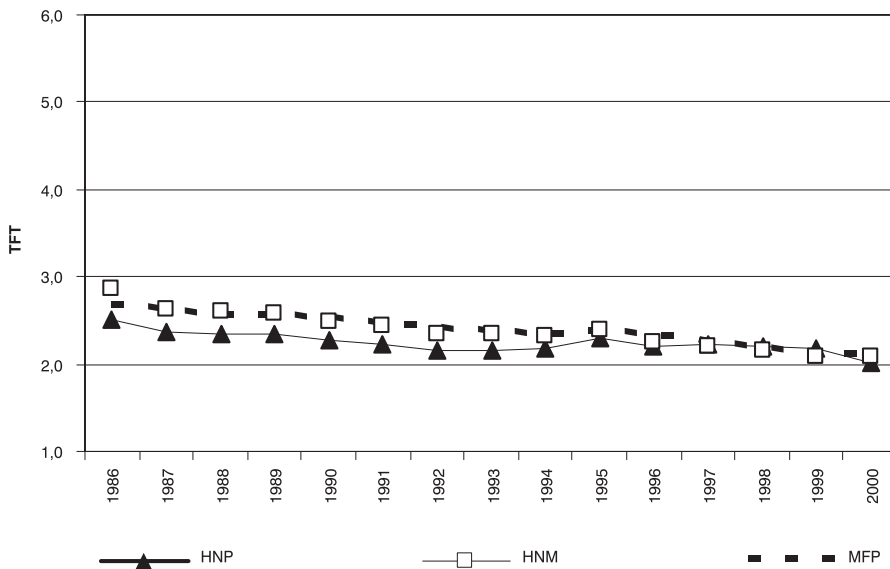
dados do Censo Demográfico de 2000 se ajusta bem à série gerada pela aplicação do MFP, como ocorre nas demais regiões analisadas. Além disso, a diferença entre a série

**GRÁFICO 4**  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Região Sudeste – 1986-2000



Fonte: Censo Demográfico de 2000.

**GRÁFICO 5**  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Região Sul – 1986-2000



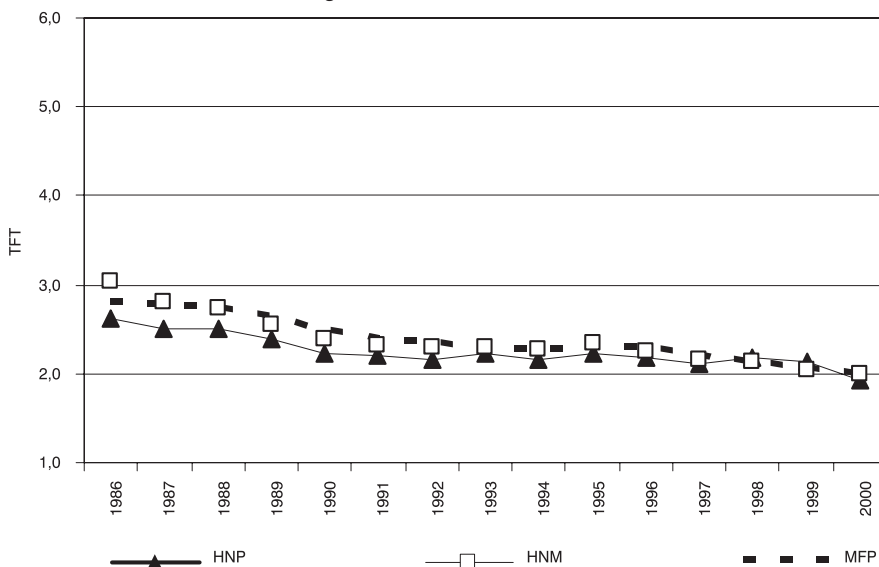
Fonte: Censo Demográfico de 2000.

gerada pela aplicação da HNP e as outras duas é pequena e as curvas praticamente se sobrepõem a partir de 1995. Essa pequena diferença, no entanto, não faz com que o teste de médias mostre um resultado diferente, ou seja, a diferença entre as médias do par HNP-MFP é estatisticamente significativa.

Assim como ocorre na Região Sul, o Gráfico 6, que apresenta os resultados para o Centro-Oeste, indica que a diferença entre

a série gerada pela HNP e as duas outras é pequena até os primeiros anos da década de 1990 e praticamente inexistente a partir de 1995. As curvas das séries HNM e MFP se sobrepõem durante o período, exceto no primeiro ano. O teste-t mostra o mesmo resultado observado para todas as regiões analisadas, ou seja, a diferença entre as médias do par HNP-MFP é estatisticamente significativa, o que não ocorre para o par HNM-MFP.

**GRÁFICO 6**  
Séries de TFT resultantes da aplicação da HNM, HNP e MFP  
Região Centro-Oeste – 1986-2000



Fonte: Censo Demográfico de 2000.

**TABELA 1**  
Teste-t para os pares HNP-MFP e HNM-MFP  
Brasil e Grandes Regiões

Regiões		Diferença entre as médias	Sig (2-tailed)
Brasil	HNP-MFP	-0,319	0,000
	HNM-MFP	0,011	0,665
Norte	HNP-MFP	-0,575	0,000
	HNM-MFP	-0,054	0,110
Nordeste	HNP-MFP	-0,525	0,000
	HNM-MFP	0,061	0,223
Sudeste	HNP-MFP	-0,199	0,000
	HNM-MFP	0,012	0,566
Sul	HNP-MFP	0,149	0,000
	HNM-MFP	-0,012	0,464
Centro-Oeste	HNP-MFP	-0,133	0,000
	HNM-MFP	-0,012	0,542

Fonte: Censo Demográfico de 2000.

Os resultados apresentados indicam que, para o caso brasileiro, a metodologia de reconstrução de histórias de nascimentos baseada no processo de pareamento (*matching*) – HNM – oferece resultados melhores do que os obtidos pela aplicação da metodologia baseada no processo probabilístico – HNP, porque se aproximam mais dos resultados alcançados pela aplicação do MFP.

O caso brasileiro, bem representado pelos resultados expostos, caracteriza-se pela heterogeneidade dos níveis de fecundidade e da qualidade da informação. Nítidamente, nas regiões que registram maior nível de fecundidade e pior qualidade dos dados – Norte e Nordeste –, os resultados da HNP foram menos satisfatórios do que os obtidos nas regiões com níveis mais baixos de fecundidade e melhor qualidade das informações – Sudeste, Sul e Centro-Oeste. Os resultados das séries da HNM parecem ser menos influenciados por esses dois aspectos, embora a curva da HNM da Região Nordeste apresente um ligeiro descolamento em relação à do MFP, nos primeiros anos do período de análise.

### Considerações finais

Este trabalho descreveu e apresentou os resultados da aplicação de duas metodologias de reconstrução de histórias de nascimentos a partir de dados censitários: HNM – baseada em processo de pareamento (*matching*); e HNP – baseada em processo probabilístico. A comparação foi feita pelo confronto entre as séries de TFT geradas pelos dois bancos de dados de histórias de nascimentos e uma terceira série de TFT, gerada pela aplicação do Método dos Filhos Próprios (MFP), utilizando os dados do Censo Demográfico de 2000. Um dos objetivos do trabalho era determinar a metodologia de reconstrução de histórias de nascimentos que produz melhores resultados para o caso brasileiro, caracterizado pela heterogeneidade na qualidade dos dados e nos níveis de fecundidade. Essa metodologia é a HNM, que produziu séries de TFT mais próximas daquelas elaboradas pelo MFP e dos resultados obtidos pela

aplicação da técnica P/F de Brass. Outro objetivo era avaliar as metodologias quanto à exequibilidade e adaptação aos dados brasileiros. Novamente, a HNM pode ser definida como a mais satisfatória, tanto pelas dificuldades de aplicação, quanto pela necessidade de obtenção de funções externas exigidas pela HNP.

A primeira diferença entre as duas metodologias está no objetivo principal de cada uma. Enquanto a HNM reconstrói a história de nascimentos das mulheres que têm entre 15 e 64 anos de idade nos 15 anos anteriores aos censos, a HNP cobre um período anterior de 50 anos. Isso seria uma vantagem comparativa a favor da HNP, caso a metodologia tivesse se mostrado eficiente na aplicação aos dados brasileiros. No entanto, se os resultados para o período de 15 anos não foram satisfatórios, muito menos satisfatórios foram para períodos maiores. A segunda diferença, desta vez favorecendo a HNM, diz respeito ao fato de ela não depender, em momento algum, da determinação de funções externas, ao contrário da aplicação da HNP, que depende de funções de mortalidade e de fecundidade de coorte para a determinação das funções de probabilidade. A vantagem da HNP em relação à HNM é o fato de a primeira não necessitar de *softwares* alternativos, o que torna o processo automático, uma vez imputados os parâmetros.

Supondo resultados satisfatórios na aplicação da HNP para o caso brasileiro, sua maior limitação residiria na complexidade, uma vez que a determinação das funções de fecundidade de coorte, dos parâmetros e das probabilidades utilizadas para atribuição da idade aos filhos omitidos não são tarefas simples. No caso do Brasil, a ausência de funções de coorte obriga a aplicação do MFP para sua obtenção. Além disso, na prática, a repetição do procedimento para cada filho omitido torna o processo demorado, apesar de não haver a necessidade de utilização de *softwares* alternativos na aplicação da HNM.

Em relação aos resultados, é possível observar que foram piores nas regiões de fecundidade mais alta, mas, de um modo geral, foram ruins em todas as regiões. Uma

hipótese para explicar as falhas no caso do Brasil pode estar relacionada ao pressuposto do modelo que assume baixa probabilidade de uma mulher ter filhos de idades próximas. Isso pode forçar a intervalos entre nascimentos maiores do que os reais, em se tratando de um contexto de fecundidade mais alta (em relação aos países desenvolvidos), como é o caso do Brasil.

Na lógica do procedimento, na medida em que as idades são atribuídas e as probabilidades recalculadas, pode chegar um momento em que as probabilidades estejam todas zeradas e ainda haja filhos para serem alocados. Se isso ocorre, não há como alocar esses filhos e as perdas são inevitáveis. Uma forma de adaptar a metodologia ao caso brasileiro seria reavaliar o cálculo das probabilidades após a alocação de cada filho, com base nos intervalos entre nascimentos obtidos no banco de dados da HNM. Entretanto, qualquer tentativa de melhorar a metodologia HNP seria mais um requinte do que uma necessidade, pois a metodologia HNM mostrou resultados bastante satisfatórios.

Ao contrário da HNP, a aplicação da HNM aos dados brasileiros mostrou-se simples e eficiente, apesar da necessidade de várias etapas e da utilização de *software* específico para o relacionamento das bases de dados de histórias de nascimentos completas e parciais. A utilização exclusiva de variáveis de reprodução na comparação entre as bases de dados é uma vantagem da metodologia, porque produz uma base de

dados final não enviesada e permite estudos comparativos levando em conta outras variáveis, como educação, renda, etc.

Acima da questão das vantagens e desvantagens, está o desenvolvimento de metodologias dessa natureza, que permitem a estimação de medidas alternativas – além das tradicionais – e auxiliam no entendimento do processo de transição da fecundidade. As histórias de nascimentos reconstruídas a partir de censos demográficos têm, como principal vantagem, a utilização de uma amostra de elevada representatividade. Outras bases de dados de histórias de nascimentos, como as PNDS (Pesquisa Nacional sobre Demografia e Saúde), são limitadas, principalmente devido ao tamanho da amostra. Entre medidas alternativas, podem ser citadas aquelas baseadas em probabilidades de nascimento: PDTFR (taxa de fecundidade total controlada por duração e parturição); PADTFR (taxa de fecundidade total controlada por idade, duração e parturição); e PPR (taxas de progressão por parturição). Além dessas medidas, a possibilidade de cálculo das funções de intensidade, que representam o risco de uma mulher de determinada idade ter um filho de determinada ordem, permite a aplicação de um modelo que estima os efeitos tempo e parturição da fecundidade. Dessa forma, é possível inserir o Brasil e outros países que possuam censos demográficos de qualidade no atual debate sobre o futuro e sobre níveis extremamente baixos de fecundidade.

## Referências

BRASS, W. **Methods for estimating fertility and mortality from limited and defective data**. Chapel Hill, North Carolina: The North Carolina Center, 1975.

\_\_\_\_\_. Cohort and time period measures of quantum fertility: concepts and methodology. In: BECKER, H. A. (Ed.). **Life histories and generations**. Utrecht: ISOR, 1991, p. 455-476.

\_\_\_\_\_. Demographic data analysis in less developed countries: 1946-1996. **Population Studies**, v. 50, n. 3, p. 451-467, 1996.

CAMARGO Jr., K. R.; COELI, C. M. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método *probabilistic record linkage*. **Cadernos de Saúde Pública**, v.16, n. 2, p. 439-447, jun. 2000. Disponível em: <[www.scielo.br/pdf/csp/v16n2/2093.pdf](http://www.scielo.br/pdf/csp/v16n2/2093.pdf)>. Acesso em: 04 mar. 2007.

CARVALHO, J.A.M. **Tendências regionais de fecundidade e mortalidade no Brasil**. Belo Horizonte: UFMG/Cedeplar, 1974.

\_\_\_\_\_. **Fecundidade e mortalidade no Brasil: 1960/1970**. Belo Horizonte: UFMG/Cedeplar, 1978. 102p.

CARVALHO, J.A.M.; PINHEIRO, S.M.G. **Fecundidade e mortalidade no Brasil: 1970/80**. Belo Horizonte: UFMG/Cedeplar, 1986. 151p.

UFMG/CEDEPLAR. **Dados preliminares para a projeção populacional por sexo e grupos de idades quinquenais, das Unidades da Federação: Brasil, 1990/2020**. Pronex, Relatório de atividades (uso restrito), 1999.

CHO, L-J.; FEENEY, G. **Fertility estimation by the own-children method: a methodological elaboration**. Chapel Hill: University of North Carolina at Chapel Hill, 1978.

CHO, L-J.; RETHERFORD, R.D.; CHOE, M.K. **The own-children method of fertility estimation**. Honolulu, Hawaii: University of Hawaii, 1986.

COALE, A.J.; DEMENY, P. **Regional model life tables and stable populations**. 2. ed. New York: Academic Press, 1983.

FERNANDEZ, R.E.; CARVALHO, J.A.M. A evolução da fecundidade no Brasil, período 1957-1979: aplicação da técnica dos filhos próprios para se estimar a fecundidade ano a ano. **Revista Brasileira de Estudos de População**, v. 3, n. 2, p. 67-86, 1986.

FEENEY, G. A simpler matrix approach to polynomial interpolation. Honolulu: East-West Population Institute, 1974. (mimeogr), apud CHO, L-J.; RETHERFORD, R.D.; CHOE, M.K. **The own-children method of fertility estimation**. Honolulu, Hawaii: University of Hawaii Press, 1986.

GRABILL, W.H.; CHO, L-J. Methodology for the measurement of current fertility

from population data on young children. **Demography**, v. 2, n. 1/2, p. 50-73, 1965.

GOLDMAN, N.; HOBBCRAFT, J. T. **Birth histories**. Voorburg, Netherlands: International Statistical Institute, 1982. (WFS, Comparative Studies, 17).

HAJNAL, J. The analysis of birth statistics in the light of the recent international recovery of the birth-rate. **Population Studies**, v. 1, n. 2, p. 137-64, Sept. 1947.

HINDE, A. **Demographic methods**. London: Hodder Arnold, 1998.

LUTHER, N. Y.; CHO, L-J. Reconstruction of birth histories from census and household survey data. **Population Studies**, v. 42, n. 3, p. 451-472, Nov. 1988.

MACHADO, C.J. **Early infant morbidity and infant mortality in the city of Sao Paulo, Brazil: a probabilistic record linkage approach**. 2002. Tese (Doutorado) – Johns Hopkins University, 2002.

MIRANDA-RIBEIRO, A. **Reconstrução de histórias de nascimentos a partir de dados censitários: aspectos teóricos e evidências empíricas**. Tese (Doutorado) – Cedeplar/UFMG, 2007.

WONG, L.R. Fecundidade no Brasil (urbano e rural): aplicação do método dos filhos próprios ao censo de 1970. **Informe Demográfico**, n. 9, p. 53-97, 1983a.

WONG, L.R. Níveis e tendências da fecundidade nas diversas regiões do Brasil: aplicação do método dos filhos próprios ao Censo de 1970. **Informe Demográfico**, n. 9, p. 101-139, 1983b.

## Resumen

*Reconstrucción del historial de nacimientos a partir de datos censales: un análisis comparativo de dos metodologías*

Este trabajo tiene como objetivo principal aplicar dos metodologías de reconstrucción del historial de nacimientos – una desarrollada en el Cedeplar y otra de autoría de Luther y Cho (1988) – a los datos censales brasileños de 2000, a fin de comparar y comprobar la confiabilidad de los resultados y la practicidad de cada una de las metodologías. Con esto se pretende determinar la metodología más adecuada al caso brasileño y buscar un mejor aprovechamiento de las bases de datos disponibles, en lo que concierne a las formas alternativas de estimación

de la fecundidad. Para ello, fueron utilizados los microdatos del Censo Demográfico de 2000. Los datos de mortalidad fueron tomados de Carvalho (1974, 1978), Carvalho y Pinheiro (1986) y del informe de la investigación Pronex/Cedeplar (CEDEPLAR, 1999).

**Palabras-clave:** Historial de nacimientos. Fecundidad. Censo Demográfico

### **Abstract**

*The reconstruction of birth histories based on census data: a comparative analysis of two methodologies*

The main objective of this article is to apply two methodologies for reconstructing birth histories – one developed by Cedeplar and another by Luther and Cho (1988) – based on 2000 Brazilian census data. The methodologies are compared and the results tested in order to verify the reliability of the findings and the practicality of each method. With this, the intention is to determine which methodology is best for Brazilian data, so that the available information can be fully explored. Most of the data are from the 2000 demographic census, but some are from Carvalho (1974, 1978), Carvalho and Pinheiro (1986), PRONEX/The Cedeplar Report (Cedeplar, 1999).

**Key-words:** Birth histories. Fertility. 2000 Brazilian Demographic Census.

Recebido para publicação em 22/08/2008.

Aceito para publicação em 12/03/2009.