

SEÇÃO DE BIOESTATÍSTICA

ESTIMAÇÃO DA PROBABILIDADE DE MUTAÇÃO GERMINATIVA
ATRAVÉS DA HISTÓRIA FAMILIAR

ESTIMATION OF THE PROBABILITY OF MUTATION THROUGH THE FAMILY HISTORY

Silvana Schneider^{1,2,6}, Juliana Giacomazzi^{2,3,6}, Patrícia Ashton-Prolla¹⁻⁶, Suzi Alves Camey^{1,7}

RESUMO

Introdução: Cerca de 10% de todos os tumores são primariamente causados por mutações germinativas de alta penetrância em genes de predisposição ao câncer. Indivíduos portadores dessas mutações têm risco significativamente maior de desenvolver câncer que não portadores e estratégias de redução de risco podem ser implementadas se estas pessoas forem identificadas. Essa identificação pode ser feita pela estimativa da probabilidade de mutação através da história pessoal e familiar do probando.

Objetivo: Descrever um dos métodos utilizados para estimar a probabilidade de um indivíduo ser portador ou não de uma mutação germinativa em dois genes associados a uma determinada doença mendeliana, condicionada a sua história familiar (HF).

Metodologia: Foi realizada a estimativa da frequência populacional da mutação em dois genes de predisposição ao câncer de mama e ovário aplicando-se a fórmula de Bayes e o teorema da probabilidade total à história familiar do indivíduo, a probabilidade dele possuir mutação considerando determinada história familiar.

Resultados e conclusão: São apresentadas e analisadas as equações utilizadas no cálculo da probabilidade de uma mutação germinativa em dois genes (*BRCA1* e *BRCA2*) associados a um desfecho específico, que é a síndrome de predisposição ao câncer de mama e ovário. Os cálculos foram feitos para incorporar a informação de dois genes, mas as fórmulas podem ser estendidas para mais genes ou para apenas um. O modelo poderá ser aplicado à população brasileira quando a probabilidade de desenvolver câncer de mama ou ovário, dado o *status* genético de *BRCA1* e *BRCA2*, para esta população for conhecida.

Palavras-chave: *Estimação de probabilidade; genética; mutação*

ABSTRACT

Background: About 10% of all malignancies are primarily caused by germline, high-penetrance mutations in cancer predisposition genes. Carriers of such mutations are at an increased lifetime risk of developing multiple cancers, as compared to non-carriers and strategies for cancer risk reduction are available if these at-risk individuals are identified. This can be done by estimating the probability that a given individual is a carrier of a germline mutation, based on his/her personal and familial history of cancer.

Aim: Describe one of the methods used to estimate probability of carrying a germline mutation in two genes associated with a mendelian disease according to family history. Methods: Na estimate of the populational frequency of germline mutations in two breast and ovarian cancer predisposition genes was done using the Bayes formula and the probability theorem taking into account the cancer family history of the proband.

Results and conclusion: The equations used in the calculations of probability for a given mutation in two genes (*BRCA1* e *BRCA2*) are presented and analysed in relation to a specific outcome, the hereditary breast and ovarian cancer (HBOC) syndrome. Calculations were done to incorporate the information of both genes, but the same formulas may be used for inclusion of more genes or only one gene. This model may be applied to the Brazilian population once the probability of developing breast and ovarian cancer, given a specific genetic status of *BRCA1* and *BRCA2* (mutation penetrance) is determined for this population.

Keywords: *Probability estimation; genetic; mutation*

Rev HCPA 2011;31(1):99-110

Mutações genéticas são importantes determinantes de diversidade entre e dentro das espécies. Na espécie humana, muitas delas estão associadas a doenças genéticas herdadas. Em câncer, mutações germinativas estão associadas a síndromes de predisposição a diversas neoplasias e mutações somáticas são também

identificadas em praticamente todos os tumores humanos (1). Mutações podem causar doenças por meio de quatro efeitos diferentes na função protéica que exercem: perda ou ganho de função, aquisição de uma propriedade nova ou expressão anômala de um gene (seja do ponto de vista temporal ou espacial) (2).

1. Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brasil.

2. Núcleo Mama Porto Alegre, Associação Hospitalar Moinhos de Vento, Porto Alegre, Brasil

3. Programa de Pós-graduação em Medicina: Ciências Médicas; UFRGS, Porto Alegre, Brasil.

4. Departamento de Genética, UFRGS, Porto Alegre, Brasil.

5. Programa de Pós-graduação em Genética e Biologia Molecular, UFRGS, Porto Alegre, Brasil.

6. Laboratório de Medicina Genômica, Hospital de Clínicas de Porto Alegre, Brasil

7. Departamento de Estatística, Instituto de Matemática, UFRGS, Porto Alegre, Brasil.

Contato: Suzi Camey. E-mail: camey@mat.ufrgs.br (Porto Alegre, RS, Brasil).

“Em 2007, a versão *online* de *Mendelian Inheritance in Man* listou um pouco mais de 3.900 doenças com padrões mendelianos de herança. Dessas, 3.310, ou aproximadamente 85%, são conhecidas por serem causadas por mutações em mais de 1.990 genes” (2).

Sanches et al (3) encontraram uma relação entre funções de alguns genes e características da doença ao qual estes genes estão associados, como a idade de aparecimento de um determinado fenótipo e o modo da herança. A presença de mutações nesses genes proporciona um aumento do número de casos da doença em síndromes específicas e em determinados grupos fenotípicos (4). Através do padrão que segue a história familiar, pode-se observar se existe a presença de mutação no gene associado à doença, pois a característica singular da doença genética é sua tendência a recorrer dentro das famílias (2). Tomando como exemplo uma das causas mais comuns da predisposição genética ao câncer de mama, a síndrome de câncer de mama e ovário hereditários, vemos que uma parcela significativa das famílias com essa doença apresenta uma mutação germinativa em um de dois genes, *BRCA1* e *BRCA2*. A identificação da mutação germinativa nesses genes em uma pessoa com o fenótipo da doença permite a definição de condutas específicas de redução de risco e também, a identificação de outros familiares em risco (irmãos e filhos de um portador de mutação, por exemplo, tem chance de 50% de também serem portadores, considerando a herança autossômica dominante das mutações). Nesse sentido, ferramentas que permitem estimar a probabilidade de mutação em um desses genes otimizam a estimativa de risco e direcionam a indicação ou não do teste genético, que deve ser realizado apenas em casos selecionados pela complexidade e alto custo (5).

Levando-se em consideração a herança da mutação, um dos interesses é saber a probabilidade de um indivíduo carregar a mutação germinativa do gene associado à doença, dado um certo padrão de diagnósticos na sua história familiar (4). E, conhecendo-se a história familiar do indivíduo em relação à doença, pode-se calcular a sua probabilidade de possuir mutação no gene associado à doença. Portanto, a estimativa da probabilidade de mutação é realizada com base em dados da história familiar do probando, indivíduo para o qual é feito o cálculo da probabilidade (6).

A probabilidade inicial de um indivíduo ser portador da mutação no gene em estudo é a estimativa da frequência populacional da mutação no gene. E, condicionando aos genótipos dos membros da família, para todos os possíveis genótipos de cada um dos familiares (*locus* com alelos selvagens; *locus*

com apenas um alelo mutado e *locus* com os dois alelos mutados), a probabilidade de ter a mutação pode ser obtida (7). O cálculo da probabilidade do *status* genético envolve algumas informações da história familiar, tais como: número de pessoas afetadas com a respectiva idade de início da doença, número de pessoas não afetadas e idade atual ou do óbito. Essas informações podem ser representadas graficamente por um heredograma, que proporciona uma fácil visualização das relações de parentesco de cada membro e uma visão global do padrão de herança e da relação na família entre indivíduos que apresentam ou não determinado fenótipo (8).

A probabilidade de mutação pode ser utilizada em duas áreas: no aconselhamento clínico individual e na pesquisa. Na primeira, ela proporciona um suporte importante na decisão sobre fazer ou não o teste genético, planejamento familiar e intervenções redutoras de risco (exames de rastreamento do câncer, cirurgias redutoras de risco e quimioprevenção) (4). Na segunda, seu cálculo oferece uma abordagem flexível para a construção de modelos de risco, por exemplo o modelo de Tyrer-Cuzick para câncer de mama (9) que são utilizados como ferramentas complementares à avaliação clínica e análise de heredograma.

O cálculo da probabilidade de mutação é aplicado no caso da síndrome Câncer Mama-Ovário, pois é sabido que em mulheres com uma história familiar positiva para estas doenças o risco de câncer de mama e ovário é significativamente maior do que nas mulheres sem esse histórico e/ou da população geral (10). Além disso, aproximadamente dois terços das famílias com três ou mais casos de câncer de mama ou ovário apresentam uma mutação em *BRCA1* ou em *BRCA2*, sendo que as estimativas da frequência de um indivíduo da população geral ser portador da mutação em *BRCA1* descritas na literatura variam entre 0,056% a 0,24%. Um estudo recente com a população canadense encontrou a taxa mais alta já publicada, 0,32%, ou 1 em 312 (11). A estimativa da frequência de um indivíduo da população geral ser portador da mutação em *BRCA2* é 0,69% a 0,072%, (11). Algumas evidências apontam para uma maior mortalidade por câncer de mama entre mulheres com mutações germinativas em *BRCA1*, em comparação com mulheres com câncer de mama esporádico (12).

Neste trabalho, apresenta-se a metodologia do cálculo da probabilidade de um probando ter uma mutação em *BRCA1* ou *BRCA2* dada a sua história familiar. Neste cálculo é incorporada a probabilidade do genótipo dado o fenótipo, através das informações contidas no heredograma, da probabilidade do fenótipo dado o genótipo (penetrância), e da distribuição marginal dos genótipos (prevalência).

A motivação deste estudo foi descrever um dos métodos utilizados para estimar a probabilidade do probando possuir ou não mutação em dois genes associados a uma determinada doença, condicionada a sua história familiar (HF). O cálculo da probabilidade de mutação apresentado envolve os genes que estão mais frequentemente associados ao Câncer de Mama, *BRCA1* e *BRCA2*.

METODOLOGIA

A história familiar normalmente é representada por um heredograma que contém as informações acerca do número de indivíduos de uma família, sexo e idade dos mesmos, suas relações de matrimônio e descendência, dados do diagnóstico de uma determinada doença genética que segrega na família e idade ao diagnóstico desta doença (ou de aparecimento de determinado fenótipo) nos familiares afetados. O primeiro indivíduo afetado da família que é avaliado do ponto de vista genético é denominado probando. Sabe-se que os indivíduos herdam dois alelos de cada gene, e que habitualmente um vem do pai e outro da mãe, os quais podem ser ambos normais (selvagens), ou então um deles pode ser mutado, ou ainda os dois alelos podem ser mutados, apresentando uma mesma mutação ou duas mutações distintas no mesmo gene (uma em cada alelo). No caso de mutações nos genes *BRCA1* e *BRCA2* o fenótipo de predisposição hereditária ao câncer ocorre quando apenas um dos alelos de um destes genes está mutado na linhagem germinativa e por esse motivo a herança da predisposição genética na síndrome de câncer de mama e ovário

hereditários é considerada autossômica dominante. Como em outras doenças com essa característica, a chance de transmitir o alelo mutado é a mesma quando este é paterno ou materno, e tanto homens quanto mulheres apresentam o fenótipo de maior predisposição ao câncer.

Supondo-se que os pais não tenham nenhum vínculo biológico (casamentos não-consanguíneos), a herança da mutação é considerada independente, ou seja, herdar a mutação em um alelo da mãe não depende de herdar a mutação em um alelo do pai. Quando o indivíduo receber os dois alelos sem alteração terá um genótipo normal (selvagem), caso contrário terá um genótipo mutante. Outra suposição considerada é que a herança da mutação de um dos genes é independente da herança do outro gene, denotados por B_1 e B_2 . Clinicamente, a observação de indivíduos com os dois alelos mutantes (seja em *BRCA1* ou *BRCA2*) é extremamente rara, pois acredita-se que na vida embrionária este genótipo é letal.

Com base nessas suposições, pretende-se calcular a probabilidade conjunta do *status* genético de B_1 e B_2 para cada indivíduo, dada a sua história familiar (6), ou seja,

$$P[B_1 = b_1, B_2 = b_2 | HF] \quad (1)$$

onde b_i é o número de alelos mutantes em B_i , $i = 1, 2$, logo $b_i = 0, 1, 2$. Quando a herança da mutação dos genes B_1 e B_2 é condicionada à HF, perde-se a independência da herança. Dessa forma, pela fórmula de Bayes (13), pode-se reescrever a equação (1) da seguinte maneira,

$$P[B_1 = b_1, B_2 = b_2 | HF] = \frac{P[B_1 = b_1, B_2 = b_2]P[HF | B_1 = b_1, B_2 = b_2]}{P[HF]} \quad (2)$$

A probabilidade do *status* genético de B_1 e B_2 dada a história familiar pode ser pensada de forma Bayesiana como sendo uma atualização, onde a probabilidade $P[B_1 = b_1, B_2 = b_2]$ é a probabilidade *a priori* e a probabilidade $P[HF | B_1 = b_1, B_2 = b_2]$ é a verossimilhança.

A primeira probabilidade a ser detalhada é $P[B_1 = b_1, B_2 = b_2]$. Sem condicionar nas informações sobre a doença e a história familiar, a probabilidade de o indivíduo herdar um determinado número de alelos mutantes em B_i segue uma distribuição Binomial com parâmetros $n = 2$ e $p = fr_i$, onde fr_i é a frequência relativa da mutação do gene B_i na população.

Lembrando que a herança de mutação nos genes é independente, obtêm-se a probabilidade conjunta pelo produto das marginais,

$$P[B_1 = b_1, B_2 = b_2] = P[B_1 = b_1]P[B_2 = b_2] \quad (3)$$

Pelo teorema da probabilidade total (13), a probabilidade da história familiar é dada por,

$$\begin{aligned}
 P[HF] &= \sum_{b_1=0}^2 \sum_{b_2=0}^2 P[HF, B_1 = b_1, B_2 = b_2] \\
 &= \sum_{b_1=0}^2 \sum_{b_2=0}^2 P[HF | B_1 = b_1, B_2 = b_2] P[B_1 = b_1, B_2 = b_2]
 \end{aligned} \tag{4}$$

Para calculá-la, e, portanto, poder calcular a equação (2), é necessário determinar

$$P[HF | B_1 = b_1, B_2 = b_2] \tag{5}$$

O cálculo da equação (5) envolve toda a história familiar do probando, portanto, dependendo do número de familiares esse cálculo pode ser computacionalmente pesado. O cálculo depende basicamente de três probabilidades: (i) probabilidade do *status* genético do filho, dado o *status* genético dos pais; (ii) probabilidade do *status* genético dos pais, dado o *status* genético do filho e (iii) probabilidade para um indivíduo desenvolver ou não o defecho, dado o seu *status* genético de B_1 e B_2 . A probabilidade do *status* genético do filho, dado o *status* genético dos pais, $P[F_i = f_i | M_i = m_i, P_i = p_i]$, (i), é dada pela Lei de Mendel (14), onde m_i e p_i representam os possíveis *status* genéticos da mãe e do pai respectivamente, f_i = *status* genético do filho.

Pela fórmula de Bayes (13), tem-se que a probabilidade do *status* genético dos pais, dado o *status* genético do filho (ii), é dada por,

$$P[M_i = m_i, P_i = p_i | F_i = f_i] = \frac{P[F_i = f_i | M_i = m_i, P_i = p_i] P[M_i = m_i, P_i = p_i]}{\sum_{m_i=0}^2 \sum_{p_i=0}^2 P[F_i = f_i | M_i = m_i, P_i = p_i] P[M_i = m_i, P_i = p_i]}, \tag{6}$$

com $f_i, m_i, p_i \in \{0,1,2\}$ e para este cálculo é necessário a probabilidade conjunta do *status* genético dos pais, que pode ser obtida pela independência entre o *status* genético da mãe e o *status* genético do pai, pois é assumido que a união não é consanguínea. Logo,

$$P[M_i = m_i, P_i = p_i] = P[M_i = m_i] P[P_i = p_i] \tag{7}$$

As probabilidades $P[M_i = m_i]$ e $P[P_i = p_i]$ são obtidas pela frequência do gene B_i na população, logo têm-se todas as informações necessárias para calcular a equação (6).

Finalmente, caracteriza-se a probabilidade para um indivíduo desenvolver o defecho, dado o seu *status* genético de B_1 e B_2 (iii), denotada por $\rho_m^{b_1 b_2}(a)$. Inicialmente define-se $\rho_m^{X, b_1 b_2}(a)$, como a seguir:

$$\rho_m^{X, b_1 b_2}(a) = \begin{cases} r_X^{b_1 b_2}(a), & \text{quando o indivíduo já teve a doença} \\ 1 - R_X^{b_1 b_2}(a), & \text{caso contrário.} \end{cases} \tag{8}$$

onde $R_X^{b_1 b_2}(a)$ é a probabilidade de um indivíduo desenvolver a doença X até a idade a dado o seu *status* de B_1 e B_2 e $r_X^{b_1 b_2}(a)$ é a probabilidade de um indivíduo ter a doença na exata idade a dado o seu *status* de B_1 e B_2 (6).

Normalmente, $R_X^{b_1 b_2}(a)$ e $r_X^{b_1 b_2}(a)$ são obtidas através da literatura. Esses dados são baseados em pesquisas que encontraram a probabilidade de desenvolver a doença dada a presença da mutação, no caso de ausência da mutação a probabilidade é obtida pela incidência populacional da doença. Quando a informação sobre a idade ou sobre o *status* da doença não é conhecida $\rho_m^{X, b_1 b_2}(a) = 1$ (6).

Supondo-se que o defecho englobe duas doenças, denotadas por X e Y, supostamente associadas aos genes B_1 e B_2 , e que o *status* de uma doença não depende do *status* da outra, quando condicionados ao *status* genético de B_1 e B_2 , tem-se que a probabilidade de um indivíduo desenvolver o defecho é a multiplicação das probabilidades para cada doença, que é denotada por $\rho_m^{b_1 b_2}(a)$ e definida por:

$$\rho_m^{b_1 b_2}(a) = \rho_m^{X, b_1 b_2}(a) \rho_m^{Y, b_1 b_2}(a), \quad b_1, b_2 \in \{0,1,2\} \tag{9}$$

Obtêm-se $\rho_m^{Y,b_1b_2}(a)$ de forma similar à $\rho_m^{X,b_1b_2}(a)$.

Voltando a equação (5), ou seja, $P[HF | B_1 = b_1, B_2 = b_2]$, onde é necessário considerar a contribuição de cada membro da família. Essa contribuição pode ser dividida em duas partes: os descendentes do probando, representada pelas funções D e Ds ; e os ascendentes do probando, representada pelas funções Av e Ap .

A função $D(m_1, m_2, p_1, p_2, n)$ incorpora as informações (idade e presença ou ausência da doença) dos filhos, dado o *status* genético dos pais, para qualquer casal da HF (6). Ela pode ser obtida através da equação:

$$D(m_1, m_2, p_1, p_2, n) = \prod_{m=1}^n \sum_{f_1=0}^2 \sum_{f_2=0}^2 \rho_m^{f_1f_2}(a) P[F_1 = f_1 | M_1 = m_1, P_1 = p_1] \times \quad (10)$$

$$\times P[F_2 = f_2 | M_2 = m_2, P_2 = p_2]$$

onde, f_1, f_2 é o *status* genético de B_1 e B_2 de cada filho do casal, m_1, m_2 é o *status* genético de B_1 e B_2 da mãe, p_1, p_2 é o *status* genético de B_1 e B_2 do pai, m é para identificar que a função $\rho_m^{f_1f_2}(a)$ refere-se ao m -ésimo filho do casal e n pode ser o número de filhos do probando (n_f), ou o número de irmãos do probando (n_i), ou o número de tios maternos do probando (n_a), ou o número de tios paternos do probando (n_A), ou, ainda, o número de filho do m -ésimo irmão do probando ($n(m)$), onde $m = 1, 2, 3, \dots, n_i$.

A função D pode ser interpretada como sendo a probabilidade dos filhos de um casal ter ou não a doença dado o *status* m_1, m_2 e p_1, p_2 do casal, independente do *status* genético dos filhos. Ela será usada para calcular a contribuição de cada filho da HF do probando. Se o casal não possui filhos, $n = 0$, assume-se que $\rho_m^{f_1f_2}(a) = 1$ e portanto $D = 1$ (6).

O termo $P[F_1 = f_1 | M_1 = m_1, P_1 = p_1]P[F_2 = f_2 | M_2 = m_2, P_2 = p_2]$ é calculado pela multiplicação das probabilidades, dadas pela Lei de Mendel e a $\rho_m^{f_1f_2}(a)$ é obtida pela equação (9).

A função $Ds(m_1, m_2, p_1, p_2)$ incorpora a informação dos irmãos do probando e seus respectivos cônjuges e filhos (6), é definida por:

$$Ds(m_1, m_2, p_1, p_2) = \prod_{m=1}^{n_i} \sum_{f_1=0}^2 \sum_{f_2=0}^2 \rho_m^{f_1f_2}(a) P[F_1 = f_1 | M_1 = m_1, P_1 = p_1] \times \quad (11)$$

$$\times P[F_2 = f_2 | M_2 = m_2, P_2 = p_2] \times$$

$$\times \sum_{k=0}^2 \sum_{l=0}^2 \rho_{q(m)}^{kl}(a) P[B_1 = k, B_2 = l] D(f_1, f_2, k, l, n(m))$$

onde, f_1, f_2 é o *status* genético dos genes B_1 e B_2 dos irmãos do probando, k, l é o *status* genético dos genes B_1 e B_2 do cônjuge dos respectivos irmãos do probando, m é para identificar que a função $\rho_m^{f_1f_2}(a)$ refere-se ao m -ésimo irmão, $q(m)$ é para identificar que a função $\rho_{q(m)}^{kl}(a)$ refere-se ao cônjuge do m -ésimo irmão, m_1 e m_2 é o *status* genético para B_1 e B_2 da mãe do probando, p_1 e p_2 é o *status* genético para B_1 e B_2 do pai do probando, $n(m)$ é o número de filhos do m -ésimo irmão.

A função $\rho_m^{f_1f_2}(a)$ é a probabilidade do irmão desenvolver ou não o desfecho condicionado ao seu *status* genético até a idade a , e $\rho_{q(m)}^{kl}(a)$ é a probabilidade do seu cônjuge desenvolver ou não a

doença até a idade a , ambas são obtidas pela equação (9). Quando a informação sobre a idade e *status* da doença dos cônjuges for desconhecida faz-se $\rho_{q(m)}^{kl}(a) = 1$.

A função Ds pode ser interpretada como a probabilidade dos irmãos do probando, seus respectivos cônjuges e seus filhos terem ou não a doença independente dos respectivos *status* genético, dado que os pais do probando têm *status* m_1, m_2 e p_1, p_2 .

A função $Av(f_1, f_2)$ incorpora a informação dos avôs e tios maternos (6), e pode ser obtida através da equação:

$$Av(f_1, f_2) = \sum_{m_1=0}^2 \sum_{m_2=0}^2 \sum_{p_1=0}^2 \sum_{p_2=0}^2 \rho_{c_1}^{m_1 m_2}(a) \rho_{c_2}^{p_1 p_2}(a) P[M_1 = m_1, P_1 = p_1 | F_1 = f_1] \times \quad (12)$$

$$\times P[M_2 = m_2, P_2 = p_2 | F_2 = f_2] D(m_1, m_2, p_1, p_2, n_a)$$

onde, f_1, f_2 é o *status* genético de B_1 e B_2 de cada filho dos avôs maternos, m_1, m_2 é o *status* genético de B_1 e B_2 da avó materna, p_1, p_2 é o *status* genético de B_1 e B_2 do avô materno, c_1 é para identificar que a função $\rho_{c_1}^{m_1 m_2}(a)$ refere-se a avó materna, c_2 é para identificar que a função $\rho_{c_2}^{p_1 p_2}(a)$ refere-se ao avô materno, n_a é o número de tios maternos.

A função Av pode ser interpretada como a probabilidade dos tios maternos e avôs maternos terem ou não a doença, independente do *status* genético dos avôs maternos dado o *status* genético da mãe do probando (f_1, f_2). Ela será usada para calcular a contribuição do lado materno da HF do probando.

A função $\rho_{c_1}^{m_1 m_2}(a)$ é a probabilidade para a avó desenvolver ou não o desfecho condicionado ao *status* genético e à idade a , e $\rho_{c_2}^{p_1 p_2}(a)$ é a probabilidade para o avô desenvolver ou não o desfecho condicionado ao *status* genético e a idade a , obtidas pela equação (9). A função $Av(f_1, f_2)$ pode ser usada para os avôs paternos, apenas substituindo n_a por n_A e as informações do lado paterno da HF.

A função $Ap(b_1, b_2)$ incorpora a informação de toda a HF familiar do probando com exceção do seu cônjuge e dos seus filhos (6), calculada pela expressão,

$$Ap(b_1, b_2) = \sum_{m_1=0}^2 \sum_{m_2=0}^2 \sum_{p_1=0}^2 \sum_{p_2=0}^2 \rho_{d_1}^{m_1 m_2}(a) \rho_{d_2}^{p_1 p_2}(a) P[M_1 = m_1, P_1 = p_1 | F_1 = b_1] \times \quad (13)$$

$$\times P[M_2 = m_2, P_2 = p_2 | F_2 = b_2] Ds(m_1, m_2, p_1, p_2) Av(m_1, m_2) Av(p_1, p_2)$$

onde, b_1, b_2 é o *status* genético de B_1 e B_2 do probando, m_1, m_2 é o *status* genético de B_1 e B_2 da mãe do probando, p_1, p_2 é o *status* genético de B_1 e B_2 do pai do probando, d_1 é para identificar que a função $\rho_{d_1}^{m_1 m_2}(a)$ refere-se a mãe do probando, d_2 é para identificar que a função $\rho_{d_2}^{p_1 p_2}(a)$ refere-se ao pai do probando.

A função Ap pode ser interpretada como a probabilidade de toda a família do probando ter ou não a doença independente do *status* genético de cada membro, dado o *status* genético do probando (b_1, b_2). Onde, a função $Ds(m_1, m_2, p_1, p_2)$ é obtida através da equação (11), a função $Av(m_1, m_2)$ é para os avôs maternos e a função $Av(p_1, p_2)$ é para os avôs paternos, obtidas através da equação (12), a função $\rho_{d_1}^{m_1 m_2}(a)$ é a probabilidade da mãe do probando desenvolver ou não o

desfecho condicionado ao seu *status* genético, e $\rho_{d_2}^{p_1 p_2}(a)$ é a probabilidade do pai do probando desenvolver ou não a doença. Tanto $\rho_{d_1}^{m_1 m_2}(a)$ como $\rho_{d_2}^{p_1 p_2}(a)$ são obtidas pela equação (9).

Enfim, incorporando as informações do próprio probando, do seu cônjuge e dos seus filhos obtém-se a probabilidade da sua história familiar condicionada ao seu *status* genético. Essa probabilidade pode ser obtida através da equação (6):

$$P[HF | B_1 = b_1, B_2 = b_2] = \rho_m^{b_1 b_2}(a) A p(b_1, b_2) \sum_{k=0}^2 \sum_{l=0}^2 \rho_{q(m)}^{kl}(a) P[B_1 = k, B_2 = l] D(b_1, b_2, k, l, n_f) \quad (14)$$

onde, b_1, b_2 é o *status* genético de B_1 e B_2 do probando, k, l é o *status* genético de B_1 e B_2 do cônjuge do probando, m é para identificar que a função $\rho_m^{b_1 b_2}$ refere-se ao probando, $q(m)$ é para indicar que a função $\rho_{q(m)}^{kl}(a)$ refere-se ao cônjuge do probando, n_f é o número de filhos.

A função $P[HF | B_1 = b_1, B_2 = b_2]$ pode ser interpretada como a probabilidade da história familiar do probando condicionada ao *status* b_1, b_2 . Será usada para calcular a contribuição de todos os membros da HF do probando.

A função $\rho_m^{b_1 b_2}(a)$ é a probabilidade de o probando desenvolver ou não o desfecho, condicionado ao *status* genético e à idade, e $\rho_{q(m)}^{kl}(a)$ é a probabilidade de seu cônjuge desenvolver ou não o desfecho dado o *status* genético e idade; ambas são obtidas através de (9).

A informação sobre a mutação nos genes do cônjuge é incorporada em $\sum_{k=0}^2 \sum_{l=0}^2 \rho_{q(m)}^{kl}(a) P[B_1 = k, B_2 = l] D(b_1, b_2, k, l, n_f)$, onde $P[B_1 = k, B_2 = l]$ é a probabilidade do *status* genético de B_1 e B_2 para o cônjuge, descrita em (3). E $\rho_{q(m)}^{kl}(a)$ é a probabilidade do cônjuge desenvolver ou não o desfecho, obtida por (9). Se o probando não tem filhos, a probabilidade do cônjuge não é levada em consideração, e, se a informação sobre a idade e *status* da doença do cônjuge for desconhecida, $\rho_{q(m)}^{kl}(a) = 1$ (6).

Usando a equação (14), pode-se finalmente calcular a equação (4) e, portanto, a equação (2). A partir desta equação, ou seja, a probabilidade conjunta do *status* genético de B_1 e B_2 do probando dada sua história familiar, diferentes resultados podem ser calculados. Por exemplo: a probabilidade do probando ser portador da mutação em pelo menos um dos genes dada a sua história familiar, π^* , que é dada por, (6),

$$\pi^* = 1 - P[B_1 = 0, B_2 = 0 | HF], \quad (15)$$

a probabilidade do probando carregar a mutação do gene B_1 , π_1^* , dada sua história familiar, que é dada por, (6),

$$\pi_1^* = P[B_1 = 1 \text{ ou } 2 | HF] = \sum_{b_1=1}^2 \sum_{b_2=0}^2 P[B_1 = b_1, B_2 = b_2 | HF] \quad (16)$$

e, analogamente, π_2^* , que é probabilidade do probando carregar a mutação do gene B_2 , dada sua história familiar.

Exemplo

Sabe-se que os genes *BRCA1* e *BRCA2*, denotados B_1 e B_2 , estão associados ao câncer de mama (CM) e ao câncer de ovário (CO). Mulheres portadoras de mutação germinativa em um desses genes (em um de seus alelos) têm uma chance de desenvolver câncer de mama ou ovário significativamente maior do que as mulheres que não são portadoras de mutação, e, geralmente, desenvolvem a doença em idade mais jovem (6). Mutações germinativas em *BRCA1* e *BRCA2* são responsáveis pelo alto risco para câncer de mama e/ou ovário em cerca de 2/3 das famílias com a síndrome (15).

A probabilidade de um indivíduo desenvolver câncer de mama ou ovário, condicionada ao *status* da mutação nos genes *BRCA1* e *BRCA2*, é estimada pelas incidências encontradas na literatura. Aqui, será utilizado o banco "BRCApenet.metaDSL.2008" que pode ser obtido através do pacote BayesMendel (16). Este banco contém as informações sobre as incidências para câncer de mama e ovário, condicionadas ao *status* genético de *BRCA1* e *BRCA2*, para cada sexo e idade em intervalos de um ano. O banco "BRCApenet.metaDSL.2008" foi construído baseado em várias pesquisas, nas quais as taxas de incidência femininas são obtidas pela combinação de nove estudos, com as melhores estimativas disponíveis em publicações reunidas através da técnica de metanálise (17). As estimativas para as penetrâncias do sexo masculino são baseados em uma das maiores coortes norte-americanas já coletadas, com base em dados de 1939 famílias, com 97 homens com câncer de mama que foram coletadas em oito centros do National Cancer Institute's Cancer Genetics Network (18). Essas incidências são as funções $\rho_m^{b_1, b_2}$ e estão distribuídas em quatro matrizes: câncer de mama nas mulheres; câncer de ovário nas mulheres; câncer de mama nos homens e câncer de ovário nos homens - porém, nesta última, as probabilidades obviamente são iguais a zero.

A distribuição da incidência do CM para as mulheres dado o *status* genético está representada na Figura 1 (A), CO para as mulheres dado o *status* genético na Figura 1 (B), e a distribuição da incidência do CM para os homens dado o *status* genético na Figura 1 (C).

Na Figura 1 (A) no quadro (a) encontra-se a distribuição da incidência de CM em mulheres que não são portadoras de uma mutação em *BRCA1* ou *BRCA2*, com incidência máxima de aproximadamente 4 novos casos em cada 1.000 aos 80 anos; no quadro (b) a incidência de CM em mulheres que carregam pelo menos uma mutação em *BRCA1*, com aproximadamente 20 novos casos em cada 1.000 em torno dos 45 anos; no quadro (c) a incidência de CM em mulheres portadoras de pelo menos uma mutação em *BRCA2*, com distribuição semelhante ao *BRCA1*; e, no quadro (d) a incidência de CM em mulheres que carregam pelo menos uma mutação em *BRCA1* e uma em *BRCA2*, onde estão as maiores incidências de CM.

Na Figura 1 (B) encontram-se as incidências em mulheres para câncer de ovário. Nota-se que a incidência para as mulheres que não possuem mutação em *BRCA1* e *BRCA2* atinge um máximo aos 85 anos de aproximadamente 0,7 novos casos a cada 1.000. E, as maiores incidências são para mulheres portadoras de mutação em *BRCA1* e em *BRCA2*, de aproximadamente 25 novos casos em cada 1.000 aos 60 anos.

Na Figura 1 (C) encontram-se as incidências para CM em homens, nota-se que o número de casos é menor, tanto para não portadores de mutação como para os com mutação em *BRCA1* e a incidência é praticamente zero. Para os portadores de pelo menos uma mutação em *BRCA2* a incidência máxima é aproximadamente 10 casos a cada 1.000 aos 80 anos.

Neste exemplo, serão apresentadas 4 variações de uma mesma HF. A primeira variação, que será chamada de HF1, é de uma probanda com diagnóstico de câncer de mama aos 47 anos e idade atual de 59 anos. Na sua história familiar constam dois irmãos ($n_i = 2$), duas tias maternas ($n_a = 2$), uma tia paterna ($n_A = 2$), um filho ($n_f = 1$), 4 filhos para a primeira irmã ($n(1) = 4$) e um filho para o segundo irmão ($n(2) = 1$). As idades e diagnóstico de câncer de mama ou de ovário (*status* da doença) para todos os membros da família estão no herograma da Figura 2.

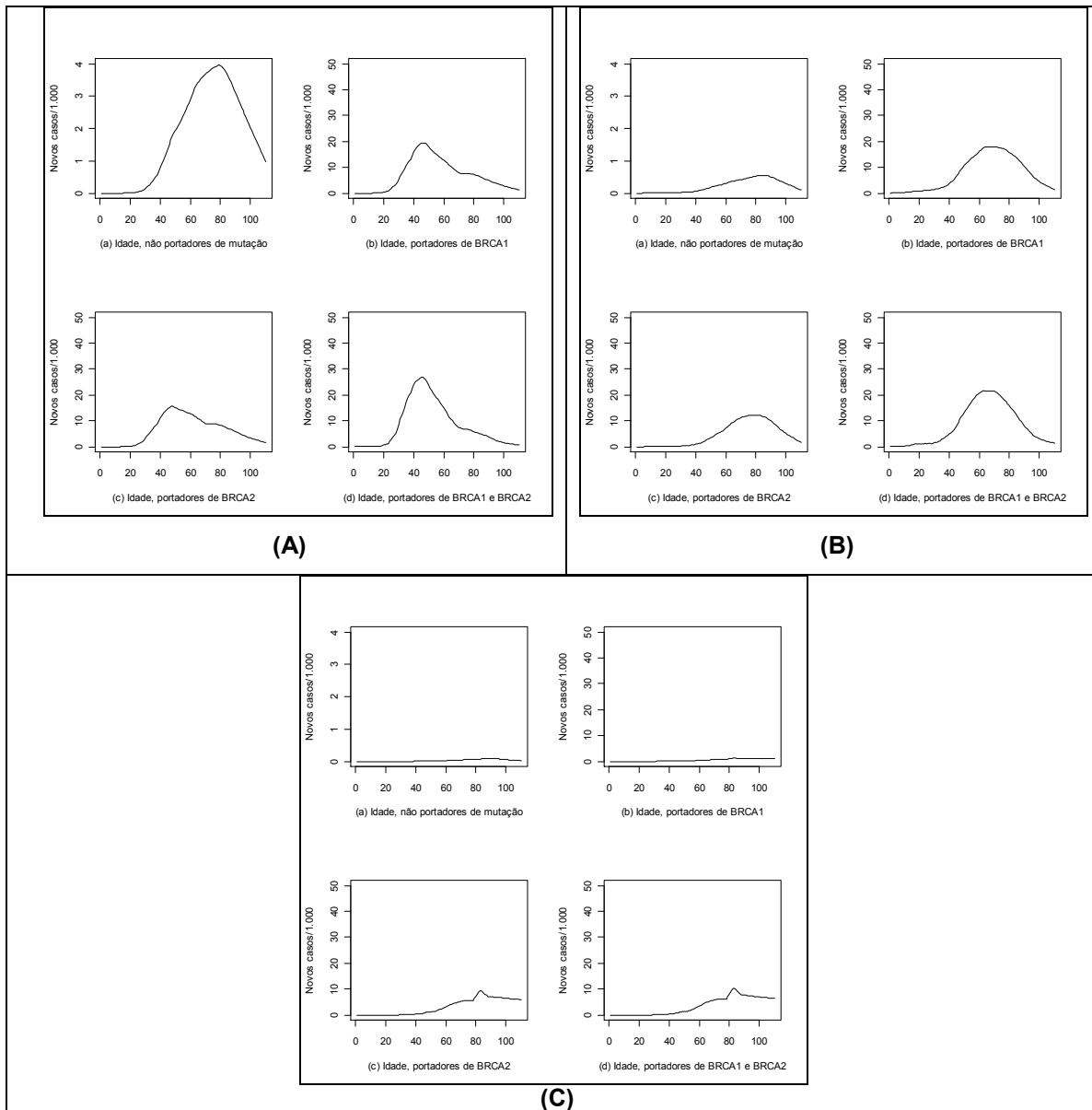


Figura 1 - (A) Incidência de câncer de mama (CM) para mulheres dado seu *status* genético de *BRCA1* e *BRCA2*, (B) Incidência de câncer de ovário (CO) para mulheres dado seu *status* genético de *BRCA1* e *BRCA2*, (C) Incidência de CM para homens dado seu *status* genético de *BRCA1* e *BRCA2*.

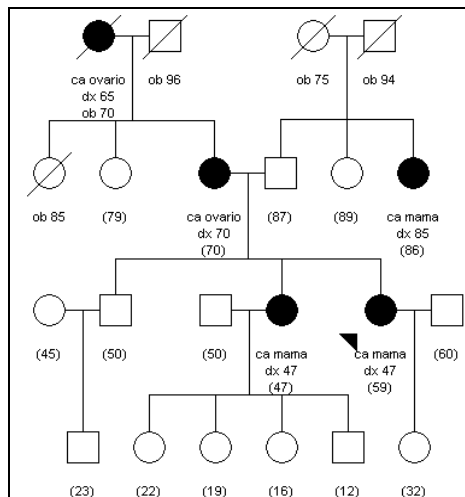


Figura 2 - Heredograma com a história familiar da probanda de 59 anos.

Utilizando-se a frequência dos genes encontrada na população canadense (maiores taxas de *BRCA1* e de *BRCA2* já descritas: $f_1 = 0,0032$ e $f_2 = 0,0069$, respectivamente) nota-se que a combinação dos *status* que maximiza a probabilidade da probanda possuir essa história familiar é *BRCA1* = 0 e *BRCA2* = 1, com probabilidade igual a $1,750 \times 10^{-12}$. Como consequência direta, temos que dada a história familiar, HF1, essa é combinação mais provável de ocorrer, pois observa-se (ver Tabela 1) uma probabilidade aproximada de 0,667 da probanda não ter mutação em *BRCA1* e ter um alelo mutado em *BRCA2* sabendo-se que a sua história familiar é HF1.

Tabela 1 - Probabilidades para o cenário 4 e HF1.

Número de alelos com mutação		Probabilidade do nº de alelos com mutação			
BRCA1	BRCA2	P(BRCA1)	P(BRCA2)	P(HF1 BRCA1, BRCA2)	P(BRCA1, BRCA2 HF1)
0	0	0,99361	0,98625	$3,113 \times 10^{-15}$	0,17071
1	0	0,00319	0,98625	$8,924 \times 10^{-13}$	0,15712
2	0	0,00001	0,98625	$8,276 \times 10^{-13}$	0,00047
0	1	0,99361	0,00685	$1,750 \times 10^{-12}$	0,66691
1	1	0,00319	0,00685	$1,370 \times 10^{-12}$	0,00168
2	1	0,00001	0,00685	$5,840 \times 10^{-13}$	0,00000
0	2	0,99361	0,00005	$1,176 \times 10^{-12}$	0,00311
1	2	0,00319	0,00005	$5,838 \times 10^{-13}$	$4,961 \times 10^{-06}$
2	2	0,00001	0,00005	$1,874 \times 10^{-13}$	$5,114 \times 10^{-09}$

A segunda variação da HF, HF2, considera que as informações sobre a idade e sobre o *status* genético da doença para o cônjuge da probanda e para os cônjuges de seus irmãos são desconhecidas. Na terceira variação da HF, HF3, tem-se que a informação dos cônjuges são desconhecidas, e que não há nenhum diagnóstico de câncer de mama ou de ovário na família, somente a probanda possui câncer de mama. E, a quarta variação da HF, HF4, é a mesma da HF3, mas a probanda também não possui câncer de mama.

Através da equação (15), pode-se obter a probabilidade da probanda ter pelo menos uma mutação germinativa em *BRCA1* e/ou *BRCA2*, e estas probabilidades (π) estão descritas na Tabela 2, para todos os cenários e variações da história familiar. As probabilidades de ter pelo menos uma mutação em *BRCA1* e de ter pelo menos uma mutação em *BRCA2* são obtidas pela equação (16) e também estão descritas na Tabela 2.

Considerando-se a HF1 e a frequência populacional para *BRCA1*=0,00080 e *BRCA2*=0,00030 (cenário 1), a probabilidade de ter pelo menos uma mutação em um dos genes é aproximadamente 0,549, de ter pelo menos uma mutação em *BRCA1* é aproximadamente 0,315 e de ter pelo menos uma mutação em *BRCA2* é aproximadamente 0,234.

Considerando-se, agora, a HF3, para as frequências do cenário 1, a probabilidade da probanda ter pelo menos uma mutação germina-

tiva é 0,00138, de ter pelo menos uma mutação em *BRCA1* é 0,00107 e pelo menos uma mutação em *BRCA2* é 0,00031. Percebe-se que se a probanda não tem uma história familiar de câncer de mama ou ovário, a probabilidade dela ter mutação é baixa. Essa probabilidade diminui ainda mais se a probanda não tem câncer de mama (HF4).

De acordo com a tabela 2, observa-se que nos três primeiros cenários independente história familiar, pode-se ver que a probabilidade de ter pelo menos uma mutação em *BRCA1* é maior do que ter pelo menos uma mutação em *BRCA2*. Isso ocorre porque a frequência populacional de mutações germinativas é maior para o gene *BRCA1*. Porém, no cenário 4, ocorre o inverso: a probabilidade de ter pelo menos uma mutação em *BRCA2* é maior do que a probabilidade de ter pelo menos uma mutação em *BRCA1*. Na HF1 e na HF2 a razão entre as probabilidades é quase quatro. A inversão da relação entre as probabilidade ocorre no cenário 4, pois nele a frequência populacional de *BRCA2* é maior do que a frequência de *BRCA1*.

Também pela tabela 3, percebe-se que as probabilidades de mutação dada a história familiar HF1 e HF2 são praticamente iguais, ou seja, possuir ou não a informação dos cônjuges não exerce grande influência no valor da probabilidade final. Porém, quando se supõe que não há nenhum caso de câncer de mama ou ovário na família, as probabilidades encontradas para a HF3 e HF4 são bem diferentes das probabilidades encontradas para a HF1 e HF2.

Tabela 2 - Probabilidade de mutação em BRCA1 e/ou BRCA2 em quatro diferentes cenários de frequência de mutação, e quatro histórias familiares diferentes.

História Familiar	Cenário	Frequências		Probabilidades condicionadas a HF		
		f ₁	f ₂	π	BRCA1	BRCA2
HF1	1	0,000800	0,000300	0,54911	0,31533	0,233930
	2	0,000600	0,000220	0,49622	0,28761	0,208703
	3	0,000450	0,000165	0,44151	0,25587	0,185717
	4	0,003200	0,006900	0,82929	0,15926	0,671707
HF2	1	0,000800	0,000300	0,54911	0,31533	0,233930
	2	0,000600	0,000220	0,49622	0,28761	0,208703
	3	0,000450	0,000165	0,44152	0,25587	0,185717
	4	0,003200	0,006900	0,82929	0,15926	0,671707
HF3	1	0,000800	0,000300	0,00138	0,00107	0,00031
	2	0,000600	0,000220	0,00103	0,00080	0,00023
	3	0,000450	0,000165	0,00077	0,00060	0,00017
	4	0,003200	0,006900	0,00083	0,00037	0,00045
HF4	1	0,000800	0,000300	0,00011	9,372 x10 ⁻⁰⁵	1,967 x10 ⁻⁰⁵
	2	0,000600	0,000220	8,470 x10 ⁻⁰⁵	7,027 x10 ⁻⁰⁵	1,442 x10 ⁻⁰⁵
	3	0,000450	0,000165	6,351 x10 ⁻⁰⁵	5,270 x10 ⁻⁰⁵	1,082 x10 ⁻⁰⁵
	4	0,003200	0,006900	0,00083	0,00038	0,00046

CONSIDERAÇÕES FINAIS

Pelo fato de os genes *BRCA1* e *BRCA2* serem os genes mais frequentemente associados ao câncer de mama e ao câncer de ovário hereditários, o cálculo da probabilidade de mutação baseado na história familiar foi feito para englobar esses dois genes, porém as fórmulas podem ser estendidas para mais genes ou para apenas um.

Todas as funções apresentadas foram desenvolvidas por Giovanni PARMIGIANI, Donald A. BERRY e Omar AGUILAR e estão introduzidas no programa BayesMendel (16). Este programa serve para calcular a probabilidade de um probando ter mutação em *BRCA1* e/ou *BRCA2* dada a sua história familiar e a estimativa de risco para o probando desenvolver ou não câncer de mama em uma determinada idade.

As probabilidades de mutação poderão ser calculadas para a população brasileira quando for conhecida a probabilidade de desenvolver câncer de mama ou ovário, dado o *status* genético de *BRCA1* e *BRCA2*, para essa população. Essa informação estará disponível no momento em forem realizados testes genéticos para identificação de mutações germinativas nesses genes de forma sistemática e em grande número de famílias com câncer de mama e ovário hereditários em diversas regiões brasileiras. É importante levar em consideração que o resultado do teste genético, para que seja informativo, deverá ser realizado dentro das mais estritas condições técnicas levando em consideração boas práticas de laboratório e as metodologias laboratoriais de mais alta sensibilidade e especificidade (se-

quenciamento genômico completo da região codificadora de ambos genes e pesquisa de rearranjos, especialmente em *BRCA1*). Com base nesses dados e conhecendo-se os fatores de risco pessoais para câncer de mama e ovário, poderá ser desenvolvido um modelo de risco que calcule a probabilidade de probandos brasileiros serem portadores de mutação germinativa em genes *BRCA* e de desenvolver câncer de mama e ovário.

Agradecimentos

Silvana Schneider recebeu bolsa de iniciação científica (UFRGS) e apresentou esse trabalho como monografia de conclusão do curso de Estatística da UFRGS. Juliana Giacomazzi recebeu bolsa de doutorado (CAPES).

REFERÊNCIAS

1. Griffiths AJ, Wessler SR, Lewontin RC, Carroll SB. Introdução à Genética, Guanabara Koogan; 2008. P. 467-9.
2. Nussbaum RL, Mcnnes RR, Willard HF. Genética Médica. 7nd ed. Elsevier; 2008. P. 243-5 and P. 439-43.
3. Jimenez-Sanchez G, Childs B, and Valle D. Human disease genes. Nature. 2001;409(15):853-5
4. Chen S, Wang W, Broman K, Katki HA, Parmigiani G. BayesMendel: An R Environment for Mendelian Risk Prediction. Johns Hopkins University, Dept. of Biostatistics Working Papers. Working 2004; Paper 39.

5. Laboratório Fleury Medicina e Saúde. [acessado em 29.11.10]; Disponível em, URL: <http://www.fleury.com.br/Pages/Default.aspx>
6. Parmigiani G, Berry DA and Aguilar O. Determining Carrier Probabilities for Breast Cancer-Susceptibility Genes BRCA1 and BRCA2. *Am J Hum Genet.* 1998;62(1):145-58.
7. Ott J. Estimation of the Recombination Fraction in Human Pedigrees: Efficient Computation of the Likelihood for Human Linkage Studies. *Am J Hum Genet.* 1974;26:588-97.
8. Regateiro FJ. Manual de Genética Médica. Coimbra; 2007. P. 98-102.
9. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in Medicine* 2004;23:1111-30.
10. Easton DF, Ford D, Bishop DT. Breast and ovarian cancer incidence in BRCA1-mutation carriers. *Am J Hum Genet* 1995; 56(1):265-71.
11. Lindor NM, McMaster ML, Lindor CJ and Greene MH. Concise Handbook of Familial Cancer Susceptibility Syndromes. *Journal of the National Cancer Institute Monographs* 2008; 38, P. 23-8.
12. Cortesi L, Masini C, Cirilli C, Medici V, Marchi I, Cavazzini G, et al. Favourable ten-year overall survival in a Caucasian population with high probability of hereditary breast cancer. *BMC Cancer* 2010;10:90.
13. Mood A M. Introduction To The Theory of Statistics. 3rd Mcgraw-Hill Uk; 1974.P.34-5.
14. Paulino WR. *Biologia*, 9nd ed. Atica; 2006. P. 316.
15. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet.* 1998;62(3):676-89.
16. Parmigiani G, Chen S, Wang W, Katki H and Blackford A. BayesMendel: Determining Carrier Probabilities for Cancer Susceptibility Genes. R package version 2.0-5, 2010. [acessado em 08.11.10] Disponível em, URL: <http://astor.som.jhmi.edu/BayesMendel/>
17. Chen S, Parmigiani G. Meta-Analysis of BRCA1 and BRCA2 Penetrance. *J Clinical Oncology* 2007; 25(11):1329-33.
18. Tai YC, Domchek S, Parmigian G and Chen S. Breast Cancer Risk Among Male BRCA1 and BRCA2 Mutation Carriers. *JNCI J Natl Cancer Inst* 2007;99(23):1811-4.

Recebido: 07/02/2011

Aceito: 09/03/2011