

NIEDJA MARISTONE DE OLIVEIRA BARRETO QUEIROZ

**REGRESSÃO LOGÍSTICA – UMA ESTIMATIVA BAYESIANA
APLICADA NA IDENTIFICAÇÃO DE FATORES DE RISCO
PARA HIV, EM DOADORES DE SANGUE.**

RECIFE

2004

NIEDJA MARISTONE DE OLIVEIRA BARRETO QUEIROZ

**REGRESSÃO LOGÍSTICA – UMA ESTIMATIVA BAYESIANA
APLICADA NA IDENTIFICAÇÃO DE FATORES DE RISCO
PARA HIV, EM DOADORES DE SANGUE.**

**Dissertação apresentada ao colegiado do
Mestrado de Biometria da Universidade Federal
Rural de Pernambuco, para obtenção do título de
Mestre em Biometria.**

**Orientador: Prof. Dr. Eufrázio de Souza Santos.
Co-orientadores: Prof. Dr. Borko Stosic e Prof. Dr.
Gauss Moutinho Cordeiro.**

RECIFE

2004

Catálogo na Fonte
Setor de Processos Técnicos da Biblioteca Central – UFRPE

Q3r Queiroz, Niedja Maristone de Oliveira Barreto
Regressão logística – uma estimativa Bayesiana aplicada na identificação de fatores de risco para HIV, em doadores de sangue / Niedja Maristone de Oliveira Barreto Queiroz – 2004.
95f. : il.

Orientador: Eufrázio de Souza Santos
Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco. Departamento de Física e matemática.
Inclui bibliografia e anexo.

CDD 574.018 2

1. Regressão logística
2. Método Bayesiano
3. Máxima verossimilhança
4. Infecção HIV
5. Sangue – Doador
- I. Santos, Eufrázio de Souza
- II. Título

NIEDJA MARISTONE DE OLIVEIRA BARRETO QUEIROZ

**REGRESSÃO LOGÍSTICA – UMA ESTIMATIVA BAYESIANA
APLICADA NA IDENTIFICAÇÃO DE FATORES DE RISCO
PARA HIV, EM DOADORES DE SANGUE.**

Dissertação apresentada ao colegiado do
Mestrado de Biometria da Universidade
Federal Rural de Pernambuco, para obtenção
do título de Mestre em Biometria.

Aprovada em 26 de março de 2004.

BANCA EXAMINADORA:

Prof. Dr. Divaldo de Almeida Sampaio
Universidade de Pernambuco – UPE

Prof. Dr. Gauss Moutinho Cordeiro
Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Borko Stosic
Universidade Federal Rural de Pernambuco - UFRPE

Prof. Dr. Rinaldo Luiz Caraciolo Ferreira
Universidade Federal Rural de Pernambuco - UFRPE
Suplente

Prof.Dr. Eufrázio de Souza Santos – UFRPE
Universidade Federal Rural de Pernambuco
Orientador

*“ Não acrediteis em coisa alguma
pelo fato de vos mostrarem o
testemunho escrito
de algum sábio antigo;
Não acrediteis em coisa alguma
com base na autoridade de mestres
e sacerdotes;
Aquilo, porém, que se enquadrar na
vossa razão,
e depois de minucioso estudo
for confirmado pela vossa experiência,
conduzindo ao vosso próprio bem
e ao de todas as outras coisas vivas;
A isso aceitai como verdade;
E daí, pautai vossa conduta!”*

Buda

*Aos meus pais Barreto e Eunice, o Princípio.
A meu marido Marcelo, o Caminho.
Aos meus filhos Marcelo Jr., Tatyana e Bruno, o Prosseguir.
A Deus, a luz.*

“[...] talvez não tenhamos conseguido fazer o melhor, mas lutamos para que o melhor fosse feito [...] Não somos o que deveríamos ser, mas somos o que iremos ser. Mas graças a Deus, não somos o que éramos” (Martin Luther King).

AGRADECIMENTOS

A Deus, Ser Supremo, que através de sua força infinita tanto me ajudou nos momentos de angústia e desânimo, mostrando-me, sempre, a luz no final do túnel. Minha fonte de força para nunca desistir.

A minha mãe, Maria Eunice, pelo amor incondicional, e pelo esforço constante no objetivo de minha formação, sempre participando e me estimulando a ir mais além.

À Universidade Federal Rural de Pernambuco UFRPE, na pessoa do Magnífico Reitor, Prof. Emídio Cantídio de Oliveira Filho, em nome de que estendo meus agradecimentos aos professores e funcionários dessa instituição, pela oportunidade e apoio durante todo o período do mestrado.

Ao Prof. Dr. Eufrázio Santos, pela orientação na realização deste trabalho e pela dedicação de tanto tempo de sua vida ao programa de Pós-graduação do mestrado em Biometria, como coordenador, professor e orientador.

Aos meus Co-Orientadores Prof. Dr. Gauss Moutinho Cordeiro e Prof. Dr. Borko Stosic pela amizade e contribuições para a realização desta dissertação.

À Fundação de Hematologia e Hemoterapia de Pernambuco - HEMOPE, na pessoa do seu presidente, Dr. Aderson Araújo, em nome de que estendo meus agradecimentos aos diretores e demais colegas servidores dessa fundação, os quais, direta ou indiretamente contribuíram para a realização desse estudo.

Ao Dr. Divaldo Sampaio, Coordenador de Ações Estratégicas do HEMOPE; por toda a sua colaboração, interesse e esforço, não só pela coleta dos dados, mas por toda a construção desta dissertação.

Ao Dr. João Lemos (In Memoriam).

A Dr^a Paula Loureiro pela atenção e todo apoio dispensado para a realização dessa pesquisa.

À Secretaria Estadual de Saúde-SES/PE, na pessoa do Ilmo. Secretário Dr. Guilherme Robalinho, em nome de quem estendo meus agradecimentos ao Coordenador de AIDS, Sr. François Figueiroa, às colegas Goretti Godoy e Djair Sena, pela disponibilização dos dados referentes a AIDS em Pernambuco.

Aos amigos João Bosco Moraes e Maria de Fátima Valadares, por todo o carinho e contribuições na conquista dessa realização.

À Prof^a. Dr^a. Maria Cristina Falcão Raposo, pela amizade, carinho e contribuições na análise dos dados.

Às Colegas da SES/PE, Carmem Clara, Cláudia Cunha, Elizabeth Caldas, Eliane Pinto, Enide Holanda, Graça Cavalcanti, Graça Pinto, Alda Justo pela amizade e incentivo.

Aos colegas de Biometria, Ana Patrícia Falcão, Cleto França, Dirac Cordeiro, Gilberto da Silva, José Carlos da Silva, José de Melo, Lázaro Araújo, Telmo Trajano, Wellington Lundgren, pelo estímulo, apoio e companheirismo durante todo curso.

Ao Sr. Valter da Silva, Setor de informática do HEMOPE, que contribuiu na preparação do programa gerador do banco de dados da pesquisa, junto ao SBS. Assim como às colegas Ana Cristina Bezerra e Jeanine Nóbrega, que me ajudaram na coleta dos dados sobre o processo de sorologia do HEMOPE.

À bibliotecária do HEMOPE Eliene Vasconcellos pela competência e presteza na revisão das referências bibliográficas.

Aos colegas de trabalho do HEMOPE, Fred Lucena, Bernadete Cavalcanti, Conceição Lima e Jorge Alberto Araújo pela compreensão e incentivo.

As minhas irmãs, Albernice e Cristiane, por ter me escutado nas horas de angústia e indicado o melhor caminho a seguir nos momentos de dúvidas.

A minha sogra, Maria Moyzette pelo apoio e dedicação.

E em fim a meu marido, Marcelo e filhos Marcelo Jr., Tatyana e Bruno, por ter me entendido nas ausências e apoiado na realização desse sonho.

RESUMO

Regressão logística tem aplicação em diversos campos como epidemiologia, pesquisa médica, bancos, pesquisa de mercado e pesquisa social. Uma de suas vantagens é que a interpretação da medida é possível através das “Odds Ratios” (OR), que são funções dos parâmetros do modelo. Neste estudo foi usado o modelo de regressão binária, com o objetivo de estimar a relação entre duas variáveis tendo em conta a presença de outros fatores. Utilizou-se para isso uma abordagem bayesiana para estimar essas medidas de risco, fazendo uma comparação com os resultados da abordagem clássica proveniente de um processo stepwise backward, utilizando o critério da razão de verossimilhança como exclusão da variável do modelo e o teste de Wald como análise de cada parâmetro do modelo final, ambos no nível de significância de 0,05. Realizou-se uma aplicação com dados reais proveniente de um estudo transversal de 106.203 doadores de sangue de 1ª doação aptos na triagem clínica no Hemocentro Recife da Fundação HEMOPE. Estimou-se medidas de associação “OR”, da infecção por HIV, com relação a algumas condições sócio-demográficas, marcadores sorológicos para outras Doenças Sexualmente Transmissíveis (DST) e tipo de doação. Para as análises no método clássico foi utilizado o pacote estatístico SPSS versão 10 e no método bayesiano o Winbugs 14. Os resultados indicaram que as OR estimadas, utilizando os dois métodos, foram bastante próximas, apesar do clássico utilizar o método de estimação por Máxima Verossimilhança, e o bayesiano utilizar os métodos de Monte Carlo Cadeia de Markov (MCMC), que são métodos diferentes. Concluiu-se, que os fatores independentemente associados ao risco de infecção por HIV entre doadores de sangue no período foram, pela estimativa bayesiana: idade 18 a 28 anos (2,45) e 29 a 39 anos (2,79); escolaridade: analfabeto (8,17), ensino fundamental (3,31) e médio (3,29); Anti-Hbc positivo (1,95); sífilis positivo (3,14); residir na Região Metropolitana do Recife (RMR) (2,41) e tipo de doação voluntária (11,94).

ABSTRACT

Logistic regression has application in several fields as epidemiology, medical research, banks, market research and social research. One of its advantages is that the interpretation of the measure is possible through the "Odds Ratios" (OR), that are functions of the parameters of the model. In this study the binary regression model was used, with the objective of estimating the relationship between two variables, taking into account the presence of other factors. For his purpose a Bayesian approach was used to estimate those risk measures, and these results were compared with the corresponding classical results obtained by application of a stepwise backward process, using the maximum likelihood as criterion for exclusion of the variable of the model, and the Wald test as analysis of each parameter of the final model, both at the level of significance of 0,05. An application was performed using real data from a transverse study of 106.203 blood donor candidates, found apt by the clinical screening process performed at the blood bank Recife of the HEMOPE foundation. Measures of HIV infection association "OR" were estimated in relation with certain socio-demographic conditions, serological markers for other Sexually Transmissible Diseases as well as the donation type. For the classical analysis the statistical package SPSS version 10 was used, and for the bayesian analysis the Winbugs 14. The results indicated that OR obtained using the two methods are rather similar, in spite of the fact that the classical approach used Maximum likelihood and the bayesian approach used the Markov Chain Monte Carlo(MCMC), which are quite different methods. It was concluded, that the factors independently associated to the HIV infection risk among donors of blood in the observed period, for the bayesian estimate, were: age 18 to 28 years (2,45) and 29 to 39 years (2,79); illiteracy (8,17), primary school (3,31) and secondary school (3,29); positive Anti-Hbc (1,95), positive syphilis (3,14), residence in the Metropolitan Area of Recife (2,41) and type of voluntary donation (11,94).

LISTA DE TABELAS

	Página
Tabela 1. Análise univariada da infecção por HIV segundo gênero, idade, escolaridade, situação conjugal, região de residência, tipo de doação, status sorológico para: anti-Hbc, HbsAg, anti-HCV, anti-HTLV e sífilis.	64
Tabela 2. Codificação Dummy das variáveis categóricas: escolaridade, idade e tipo de doação, do modelo final no SPSS.	65
Tabela 3. Tabela de contingência para o teste de Hosmer-Lemeshow no SPSS	66
Tabela 4. Estimativas dos parâmetros β , seu desvio padrão(SE), teste de Wald,, OR e Intervalo de confiança(IC) para OR, referentes ao modelo (24), no SPSS.	70
Tabela 5. Distribuição da Função Eficiência por classe do Logit.	71
Tabela 6. Resultados das estimativas dos parâmetros, seu desvio padrão e IC 95%, para o modelo final múltiplo, realizando-se 1.400 iterações no Winbugs	73
Tabela 7. Comparação das estimativas das OR, seus IC 95% e amplitude do IC, para o modelo final múltiplo, entre os dois métodos de estimação.	74
Tabela 8. Tabela-padrão para a apresentação dos resultados em quatro tipos de estudo: ensaio clínico randomizado e estudos de coorte, de caso-controle e transversal.	88
Tabela 9. Marcador sorológico da sífilis versus infecção por HIV. jan98-nov/03, Recife-PE.	88
Tabela 10. Estimativas dos β e OR do modelo apenas com a variável sífilis, no SPSS.	89
Tabela 11. Estimativas dos parâmetros e OR do modelo apenas com a variável sífilis, no Winbugs, com 800 iterações.	90
Tabela 12. Classificação cruzada entre infecção HIV e o tipo de doação. Recife-PE. jan1998-nov2003.	92
Tabela 13. Especificação de variáveis dummy para tipo de doação usando autóloga/ específica como grupo de referência.	92
Tabela 14. Estimativas dos parâmetros e OR do modelo apenas com a variável tipo de doação transformada em dummy, com grupo de referência “autóloga ou específica”, no SPSS.	93
Tabela 15. Estimativas dos parâmetros e OR do modelo apenas com a variável tipo de doação, no Winbugs, com 1.400 iterações.	94

LISTA DE FIGURAS E QUADRO

	Página
Figura 1. Estrutura de um estudo transversal.	30
Quadro 1. Variáveis relevantes associadas à infecção HIV, que entraram no estudo.	31
Figura 2. Recepção de doadores do Hemocentro Recife.	38
Figura 3. Setor de Pré Triagem do Hemocentro Recife.	39
Figura 4. Setor de Triagem Clínica do Hemocentro Recife.	39
Figura 5. Sala de Coleta do Hemocentro Recife.	40
Figura 6. Laboratório de Sorologia do Hemocentro Recife.	41
Figura 7. Gráfico da função sigmóide.	45
Figura 8. Ponto que Maximiza o nº de doador e Minimiza o risco de Infecção por HIV.	71
Figura 9. Estimativas dos parâmetros segundo o número de iterações no WinBUGS, por variável independente.	72
Figura 10. gráficos dos quantis da cadeia para os parâmetros β_{10} e β_{11} com 12.000 iterações.	73
Figura 11. Gráficos do traço para cada parâmetro do modelo, realizando-se 1400 iterações, no WinBUGS 14.	76
Figura 12. Gráficos da densidade posteriori para cada parâmetro do modelo, realizando-se 1.400 iterações, no WinBUGS 14.	77
Figura 13. Gráficos dos quantis posteriori para cada parâmetro do modelo, realizando-se 1.400 iterações, no WinBUGS 14.	78
Figura 14. Gráficos de convergência (do traço, da densidade posteriori, dos quantis posteriori) das estimativas dos parâmetros, do modelo da infecção HIV, apenas com a variável sífilis, no Winbugs 14, com 800 iterações.	91
Figura 15. Gráficos de convergência (do traço, da densidade posteriori, dos quantis posteriori) das estimativas dos parâmetros, do modelo da infecção HIV, apenas com a variável tipo de doação, no Winbugs 14, com 1.400 iterações.	94

ABREVIATURAS E SIGLAS UTILIZADOS

AIDS	<i>Acquired Immune Deficiency Syndrome</i>
Anti - HCV	Anticorpo anti-vírus da hepatite C
Anti - HIV	Anticorpo anti-vírus Imunodeficiência Humana
Anti - HTLV	Anticorpo anti-vírus linfotrópicos de células T humanas
Anti-HBc	Anticorpo contra o antígeno de centro (core)
DST	Doenças Sexualmente Transmissíveis
ELISA	Enzima imunoensaio
HBs Ag	Antígeno de superfície do vírus da hepatite B
HCV	Vírus da hepatite C
HEMOPE	Fundação de Hematologia e Hemoterapia de Pernambuco
HIV	Vírus da Imunodeficiência Humana
HTLV	Vírus linfotrópicos de célula T humanas
IC a 95%	Intervalo de Confiança a 95%
MCMC	<i>Markov Chain Monte Carlo</i>
NIH	<i>National Institute of Health</i>
OMS	Organização Mundial de Saúde
OR	Odds Ratios
PR	Prevalência Relativa
RMR	Região Metropolitana do Recife
RR	Risco Relativo
SBS	Sistema de Banco de Sangue
SES/PE	Secretaria Estadual de Saúde de Pernambuco
SIDA	Síndrome da Imunodeficiência Adquirida
SUS	Sistema Único de Saúde
TC	Triagem Clínica
TSS	Teste Sorológico para Sífilis
VHB	Vírus da Hepatite B

SUMÁRIO

	Página	
1	INTRODUÇÃO E JUSTIFICATIVA	14
2	REVISÃO DE LITERATURA	16
2.1	REGRESSÃO LOGÍSTICA	16
2.2	AIDS E TRANSMISSÃO POR TRANSFUSÕES DE SANGUE	22
3	MATERIAIS E MÉTODOS	28
3.1	MATERIAIS	28
3.1.1	Área de estudo	28
3.1.2	População de estudo	28
3.1.3	Tamanho amostral	28
3.1.4	Desenho do estudo	29
3.1.5	Elenco das variáveis	30
3.1.5.1	Variável dependente	31
3.1.5.2	Variáveis independentes	32
3.1.6	Coleta de dados	37
3.1.7	Processamento dos dados	41
3.1.8	Aspectos éticos	41
3.2	MÉTODOS	42
3.2.1	Modelo de regressão logística binária	42
3.2.2	Transformação logit	46
3.2.3	Método clássico de estimação dos parâmetros	47
3.2.3.1	Modelo logístico simples	47
3.2.3.2	Modelo logístico múltiplo	50
3.2.3.3	Teste de hipótese sobre os coeficientes	53
3.2.3.4	Bondade do ajuste	54
3.2.4	Método bayesiano de estimação dos parâmetros	55
3.2.5	Medida de associação	59
3.2.6	Análise estatística	61
4	RESULTADOS E DISCUSSÕES	63
4.1	MÉTODO CLÁSSICO	64

4.2	MÉTODO BAYESIANO	72
5	CONCLUSÕES	80
6	RECOMENDAÇÕES	81
7	REFERÊNCIAS BIBLIOGRÁFICAS	82
	APÊNDICE	88
	ANEXOS:	
	ANEXO 1 FICHA DE DOAÇÃO	95
	ANEXO 2 DECLARAÇÃO DO COMITÊ DE ÉTICA	96
	ANEXO 3 PROGRAMA REGRESSÃO LOGÍSTICA NO WINBUGS	97

1. INTRODUÇÃO E JUSTIFICATIVA

A Análise de Regressão Logística tem sido mencionada por alguns autores, Cox (1969), Martins (1988), Hosmer e Lemeshow(1989), Ebrahim (2001), Amorim (2001) e Guimarães (2001), como uma ferramenta bastante poderosa de modelagem estatística, geralmente apropriada para variáveis categóricas. Ela consiste em relacionar, através de um modelo, a variável resposta categórica com as variáveis explanatórias que influenciam a ocorrência de determinado evento. A variável resposta é geralmente dicotômica, mas pode ser politômica, isto é, ter mais do que dois níveis de respostas, podendo ser nominal ou ordinal. Já as variáveis explanatórias podem ser categóricas ou contínuas.

A Regressão Logística tem aplicação em diversos campos como epidemiologia, pesquisa médica, bancos, pesquisa de mercado, pesquisa social, etc.

Neste estudo foi tratado o modelo de regressão logística binária, em que a variável resposta é dicotômica, e esta dicotomia foi representada por doador de sangue “soropositivo” ou “soronegativo” para a infecção pelo Vírus da Imunodeficiência Humana (HIV), com o objetivo de “estimar a associação de fatores de risco relacionados à infecção por HIV em doadores de sangue”.

Existem vários meios de transmissão do HIV, dos quais as vias: sexual, parenteral e perinatal são as principais. O presente estudo enfocou a via parenteral, especificamente a transmissão pela transfusão de sangue.

Atualmente, com as medidas de triagem clínica e sorológica de doadores, a prevenção da transmissão do HIV-1 através das transfusões sanguíneas se tornou extremamente eficiente, levando ao quase desaparecimento dessa forma de transmissão. Historicamente, vê-se uma prevalência baixa de testes HIV-positivos. Na região nordeste, a prevalência apresentou-se em crescimento chegando a 2,650 por mil doações em 1997 (BRASIL, 2003; KUPEK, 2000). Em Pernambuco, a proporção de casos por essa categoria de transmissão é da ordem de 0,05%, no período de 1998 a 2003 (BOLETIM INFORMATIVO DST/AIDS, 2003).

Mas, devido ao período de janela imunológica, no qual os teste sorológicos não detectam a infecção, existe um risco residual de transmissão, motivo pelo qual há necessidade de aprimorar a triagem clínica, pois a “janela” é o indicador dos resultados falsos-negativos da triagem sorológica, implicando na possibilidade de contaminação do receptor. A AIDS pós transfusional, embora rara, deve ser

preocupação permanente de todos os indivíduos que manipulam ou administram sangue ou derivados.

Na atualidade, os estudos sobre o processo de seleção de doadores de sangue, bem como sobre a doação de sangue e a segurança dos receptores de sangue, permanecem como temas de elevado interesse para a comunidade científica mundial. Tais estudos têm ampliado o conhecimento sobre novas formas de proteção de doadores e receptores de sangue em várias partes do mundo, suscitando a discussão de questões relevantes, sobretudo em relação à qualidade, à segurança e à quantidade dos produtos sanguíneos ofertados à população, as quais dependem, por sua vez, de conhecimento preciso acerca da população de candidatos à doação de sangue (SAMPAIO, 2001).

Objetivando contribuir para essa prevenção, desenvolveu-se o presente estudo, utilizando para estimar os parâmetros do modelo da regressão logística o enfoque Clássico e Bayesiano na detecção de fatores de riscos associados a soropositividade por HIV entre doadores de sangue, referentes ao período do estudo

Procurou-se avaliar a importância de variáveis sócio-demográficas definindo-as como: gênero, idade, escolaridade, situação conjugal e região de residência, e o tipo de doação como determinantes indiretos do risco de infecção por HIV na etapa da Triagem Clínica (TC). Avaliou-se, ainda, a utilidade dos testes sorológicos para pesquisa de anticorpos antígeno “core” do vírus da Hepatite B (anti-HBc), antígeno da Hepatite C (anti-HCV), antígeno linfotrófico de células T humanas tipos I e II (anti-HTLV I/II), testes sorológicos para sífilis e pesquisa do antígeno de superfície do vírus da hepatite B (HbsAg) como marcadores sorológicos indiretos do risco de infecção pelo HIV.

Comparou-se, também os resultados das duas metodologias de estimação, bastante utilizadas na literatura em estudos aplicados na área da saúde.

Assim, espera-se contribuir para minimizar esse risco residual, melhorando a qualidade do sangue a ser transfundido, facilitando uma terapêutica adequada para o receptor do sangue e subsidiar no maior controle da evolução da AIDS, via transfusão sanguínea.

2. REVISÃO DE LITERATURA

2.1. REGRESSÃO LOGÍSTICA

Ao longo do tempo o conhecimento científico vem sendo construído a partir de evidências concretas, produto da curiosidade e observação do homem.

Segundo Martins (1988) em muitas situações práticas o pesquisador vê-se envolvido com a necessidade de construir um modelo matemático, para estudar alguns fenômenos de observação. De acordo com ele um modelo científico pode ser definido como uma abstração de um sistema real, que possa ser utilizada com os propósitos de predição e controle. E, que para ser útil, tal modelo deve abranger elementos de dois atributos conflitantes: realismo e simplicidade. Se por um lado o modelo deve servir como uma aproximação razoavelmente precisa do sistema real, e conter a maior parte dos aspectos importantes do mesmo, por outro, não deve ser tão complexo que se torne impossível compreendê-lo e manipulá-lo.

A construção de um modelo é baseada em informações obtidas da realidade através das observações ou medidas. Em geral, é difícil afirmar, com certeza, se um modelo matemático idealizado é ou não adequado, antes que alguns testes de observações sejam realizados (MARTINS, 1988).

De acordo com Ebrahim (2001), na vida real, entretanto, muitos fatores estão atuando simultaneamente sobre a variável resposta. Alguns desses fatores podem influenciar a variável resposta, enquanto outros podem atuar simplesmente como fatores de confusão. Assim, em busca de se obter uma correta interpretação do que esteja realmente ocorrendo, devemos utilizar métodos que levem em consideração a ação conjunta de variáveis que estejam atuando simultaneamente. Desta forma os métodos bivariados são recomendados como etapa preliminar na análise de dados e a análise multivariada, como o segundo passo seqüencialmente mais lógico.

Gomes (2001) observa que o médico, na sua prática profissional, precisa reconhecer a doença (variável de efeito, o desfecho) e quais fatores estão envolvidos no seu aparecimento e na sua evolução (variáveis preditivas, fatores de exposição ou simplesmente exposição). Esses fatores podem representar associação entre exposição e doença, sendo, assim, fatores de risco ou de proteção.

Esse mesmo autor ainda refere que nos estudos transversais, a aferição de doença e exposição é feita simultaneamente em uma amostra da população ou

população de estudo, e a medida de frequência da doença é a prevalência (pessoas que estão doentes no momento da entrevista). Assim pode-se estabelecer associação, mas não causalidade. (GOMES, 2001).

Hulley *et. al.*(2003), porém, referem que, nesse caso, as variáveis preditoras e de desfecho são definidas apenas com base nas hipóteses de causa-efeito do investigador e não no delineamento do estudo. Essa definição é mais fácil para fatores constitucionais, como idade e raça, pois, como eles não podem ser alterados pelo efeito de outras variáveis, só podem ser considerados preditores. Para a maioria das variáveis, no entanto, a escolha se torna mais difícil. Qual a causa e qual a consequência?

Esses mesmos autores preconizam que em pesquisa médica deve-se entender como associação um evento em que duas ou mais características ocorrem juntas mais vezes do que seria esperado por chance ou por acaso. Quando encontramos associações entre possíveis fatores de risco e um resultado, devemos indagar: tal associação é verdadeira ou espúria? Se verdadeira, é causal ou não?(HULLEY *et. al.*,2003).

Nos últimos 15 anos, conforme relatos de Schiaffino *et al.* (2003), pode-se encontrar resultados de estudos epidemiológicos e clínicos com desenho transversal expressados em forma de OR, mas que utilizam a definição de Prevalência Relativa (PR). Este possível erro conceitual se viu favorecido pela dificuldade de ajustar a PR por múltiplas variáveis, enquanto por sua relativa simplicidade, uma alternativa comumente utilizada foi calcular o OR mediante modelos de regressão logística como aproximação da PR. Esta aproximação poderia ser correta quando a prevalência da doença ou condição de interesse que se está estudando é pequena (classicamente conhecido como suposição de doença rara).

Segundo Pereira (2000) quando a doença é pouco freqüente, isto é, menos de 5% para alguns autores, ou mesmo 10% para outros, o que é usualmente a situação encontrada na prática, para a maioria dos danos à saúde, o OR é uma boa estimativa do Risco Relativo (RR). As células “a” e “c” de uma tabela de contingência (Apêndice, tabela 8) pouco contribuem para o denominador dos coeficientes de incidência, pois $a/(a+b)$ resulta em praticamente o mesmo que a/b , como também $c/(c+d)$ resulta em c/d , desta maneira, omitindo-se “a” e “c” dos denominadores, chega-se à fórmula do RR.

Ainda segundo Pereira (2000), historicamente, o OR foi utilizado como aproximação do RR. No entanto, mais recentemente, constatou-se que o OR constitui também uma medida própria de risco, inclusive com vantagens sobre o RR. Uma delas é que, ao contrário do RR, o OR pode ser calculado em qualquer tipo de estudo epidemiológico.

Stromberg (1994) sugere que se utilize o OR calculado mediante regressão logística, tendo em conta sua verdadeira definição. Nesse sentido, de acordo com Concato (1993), os métodos multivariados se tornaram formas crescentemente populares de análise de dados em pesquisa médica, sendo a regressão logística uma das que aparece em 15% (96 de 660) de artigos publicados em dois jornais médicos gerais conduzidos durante 1989.

O modelo de regressão logística, segundo Schiaffino *et al.* (2003), proporciona o OR de contrair a doença naqueles indivíduos que sofreram uma exposição (X_E) em relação àqueles indivíduos que não a sofreram, ajustada por um conjunto de características (X_1, \dots, X_k). O modelo utiliza a transformação logit (logaritmo neperiano das odds) para evitar que a função obtida possa tomar valores negativos.

O problema de estimativas de risco instáveis e valores de p impróprios produzidos por dados de superestimação, segundo Concato (1993), pode ser prevenido assegurando um número adequado de eventos de resultado (número de casos de HIV-positivos). Uma baixa relação desses casos para cada variável independente (menos que 10:1), torna o risco estimado incerto. O assunto chave na superestimação é um amplo número de eventos de resultado, não só um tamanho de amostra grande. E o número de variáveis selecionadas para análise, deveria ser então, parcimonioso, baseado em sensibilidade clínica e qualidade de dados satisfatória.

Ainda de acordo com Concato (1993), em “circunstância de não linearidade” os riscos não deveriam ser calculados quantitativamente como um único valor, mas em zonas ou categorias dos dados. Esse problema foi evitado freqüentemente pela estratégia de usar variáveis independentes binárias – uma tática utilizada em 30 dos 44 artigos pertinentes. O verdadeiro impacto de variáveis contínuas ou ordinais, porém, pode ser mascarado quando são criadas duas zonas binárias. Logo, em lugar de usar uma classificação dicotômica, pode-se converter variáveis contínuas em uma ordem de zonas ordinais ou pode-se transformar em “variáveis dummy”

apropriadas para o contexto clínico de cada análise. Embora o número ideal de zonas não pode ser especificado com antecedência e frequentemente requer julgamento, clínico.

No que se refere aos estudos em doadores de sangue, McFarland *et. al.* (1998) verificaram que os fatores de risco para a soropositividade, identificados pelo estudo da análise de regressão logística clássica, eram consistentes com o conhecimento epidemiológico da infecção por HIV, em Zimbabwe e no restante do sub-Saharan, sendo implementado com sucesso um algoritmo para a exclusão de doadores, baseado nesses fatores.

Por outro lado, Freire (2001) concluiu, considerando o problema de má classificação e utilizando dados do estudo de epidemiologia sobre meningite bacteriana, que o método bayesiano além de mais prático que os outros métodos estudados, apresentou-se como o de melhor performance para o exemplo utilizado.

Outros estudos vêm demonstrando o impacto do emprego dos métodos Bayesianos sobre a bioestatística (BRESLOW, 1990 apud FREIRE, 2001), seja na utilização dos modelos de regressão logística bayesianos no estudo do câncer de bexiga (GILKS; RICHARDSON, 1992 apud FREIRE, 2001), ou nas pesquisas de Marshall *et. al.* (1994), concluindo que a habilidade dos modelos Bayesianos de discriminar entre mortes e sobrevivências no ato cirúrgico é comparável com aquela da regressão logística clássica. Mais recentemente Kahn (1999), utilizou um modelo Bayesiano para predições clínicas de profundas trombozes venosas afetando membros inferiores, Viallfont *et.al.* (2001), que propuseram a utilização dos modelos bayesianos em estudos de caso-controle para investigar fatores de risco do câncer cervical. Andrieu *et.al.* (2000), que estimaram o risco familiar do câncer de mama associado a fatores reprodutivos e menstruais, através de um modelo de regressão logística com efeitos aleatórios bayesianos, e, finalmente Lee *et.al.* (2003), que introduziram uma aplicação bayesiana para análise de modelos de equação estrutural geral com variáveis dicotômicas à um estudo de não aderência para medicamentos em pacientes hipertensos.

Segundo Santana *et. al.* (2002), observou-se, recentemente, um crescente interesse pela utilização dos métodos bayesianos por parte da comunidade médica. Reflexo disso é o número de citações em Medline nos últimos 5 anos, que chegou a 1.212 aparições em títulos ou resumos. Em seu artigo ilustraram de maneira introdutória as enormes possibilidades que a análise bayesiana pode contribuir à

investigação biomédica.

Marshall *et. al.* (1994) estudaram 12.712 pacientes portadores de desvio de artéria coronária e verificaram que o poder preditivo do modelo bayesian-logit é comparado com o Modelo do teorema Bayes, regressão logística, e análise discriminante. Mostraram que o modelo pode ser reformulado usando a mesma equação logística empregada no modelo de regressão logístico clássico para produzir um modelo novo, o modelo bayesian-logit. E que estes modelos diferem somente na metodologia que serve para estimar os coeficientes da regressão, mas não no relacionamento estrutural entre os fatores de risco e o resultado. Produz, também, coeficientes para as covariáveis que indicam a força de seu relacionamento com a variável resposta, e é capaz de incorporar variáveis quantitativas da mesma forma como usado em modelos lineares ou de regressão logística clássica, evitando assim a necessidade para pontos arbitrários de corte em variáveis contínuas.

Schmid e Rosner (1993), propuseram uma aplicação Bayesiana para estimar os parâmetros em um modelo de regressão logística, quando os preditores estiverem sujeitos ao acaso ou erro de medida sistemático e calcularam a probabilidade logística verdadeira, através da distribuição a posteriori condicional do valor verdadeiro dos preditores dado seus valores observados. Afirmaram que as estimativas de risco resultantes diferem nitidamente dessas calculadas pela regressão logística padrão, que ignora erro de medida.

Na abordagem bayesiana, os principais conceitos envolvidos são a probabilidade a priori e a probabilidade a posteriori (ANTELMAN, 1977).

Antelman (1977); Bergerud e Reed(1998) e Santos (2001) esclarecem que a inferência bayesiana é constituída por três estruturas básicas: a distribuição a priori, que indica o estado atual de informação do pesquisador, ou seja, representa “tudo” o que é conhecido adicionalmente ao experimento antes da observação dos dados; a função de verossimilhança, que expressa todo o conhecimento do experimento contido nos dados, ou seja, codifica toda a informação relevante contida nos dados sobre o parâmetro; e a distribuição a posteriori, que representa o conhecimento sobre o experimento atualizado pelos dados, especificando o estado da informação sobre o parâmetro de interesse, após a observação dos dados.

Silva e Suárez (2000) e Santana *et. al.*(2002) destacam que a abordagem clássica é principalmente empírica utilizando somente a informação amostral como base para estimar e testar hipóteses a respeito de parâmetros populacionais,

enquanto que a abordagem bayesiana utiliza toda informação amostral além do julgamento pessoal para escolha de uma distribuição de probabilidade a priori adequada para o cálculo de estimativas.

Box e Tiao (1973) discutiram as idéias de Jeffreys (1961), sobre a distribuição a *priori* para representar o estado de ausência de informação ou ignorância a respeito do comportamento probabilístico dos parâmetros. O estudo abrangeu os casos uniparamétricos e multiparamétricos. É, talvez, a discussão mais difundida sobre *priori* não-informativa. Segundo Brasil (1991), a crítica mais freqüente à análise Bayesiana é que diferentes priores conduzem a diferentes respostas. Contudo, querendo encontrar “objetividade” pode-se usar priores não-informativas.

Se houver amostra pequena de dados, então é necessário fazer sérias considerações para a informação a priori. Escolher uma priori não informativa simplesmente porque ela é conveniente será insatisfatório. Conseqüentemente, quando amostra é grande, intervalos de confiança clássicos e intervalo de credibilidade Bayesiano serão quase idênticos numericamente (POLLARD, 1986).

Modelos de respostas dicotômicas são usados em muitos estudos médicos e de econometria para examinar a relação entre várias covariáveis e uma medida de resultado dicotômico. Zellner e Rossi(1984), analisaram modelos de resposta quantal do ponto de vista Bayesiano, usando distribuição de priores difusa e informativa. Quando informação anterior não está prontamente disponível uma distribuição de priori uniforme ou algumas outras priores não informativas tais como as priores de Jeffreys (1961), para os coeficientes da regressão é freqüentemente usado.

Ming-Hui Chen e Qi-Man Shao (2000) investigaram a propriedade da distribuição posterior para modelo de resposta dicotômica, usando uma distribuição priori difusa para os parâmetros da regressão.

Esses autores afirmam que quando informações prévias não estão prontamente disponíveis uma distribuição priori uniforme ou algumas outras priores não informativas tais como as priores de Jeffreys para os coeficientes de regressão, são freqüentemente usadas. Porém, uso de prior uniforme ou de prior não informativa conduz tipicamente a um problema desafiador. Ou seja, se as distribuições posteriores resultantes são formais.

Segundo Leandro (2001), o intensivo desenvolvimento computacional aliado ao uso de métodos de Monte Carlo Cadeia de Markov (MCMC) têm tornado as

técnicas Bayesianas mais acessíveis. O uso de MCMC é provavelmente o fator mais importante no desenvolvimento deste campo e tem tornado computacionalmente viável a aplicação de Métodos Bayesianos.

Atualmente, um dos tópicos mais ativos em Estatística Computacional é a inferência através de simulação iterativa, utilizando, especialmente, o amostrador de Gibbs e o algoritmo de Metropolis-Hastings (HASTINGS, 1970; GILKS, 1996; GAMERMAN, 1997; ROBERT; CASELLA, 1999 ; CHEN *et al.*, 2000). Esta área teve um desenvolvimento acelerado na segunda metade do século XX devido a grande disponibilidade de computadores de alta velocidade. A idéia essencial da simulação iterativa é gerar valores de uma variável aleatória, utilizando uma seqüência de distribuições que converge iterativamente para a distribuição origem e desempenha um papel extremamente marcante em inferência Bayesiana no momento de calcular resumos de distribuições a *posteriori* extremamente complicadas (LEANDRO, 2001).

Em uma rápida demonstração do amostrador de Gibbs, Albert e Chib (1993) mostraram como amostras normais truncadas poderiam ser usadas para implementar o amostrador de Gibbs para um modelo probit para respostas binárias. Como também, Pilote *et. al.* (2000) para estimarem os parâmetros do seu modelo regressão bayesiano-logit, usaram o amostrador de Gibbs, e estabeleceram fatores de risco para doenças da artéria coronária.

Segundo Santana *et. al.* (2002) a abordagem bayesiana pode ser considerada uma extensão da clássica, pois ambas as abordagens podem levar aos mesmos resultados, entretanto, estes resultados são diferentes em termos de interpretação. Ou melhor, é um enfoque alternativo para os métodos frequentista da estatística clássica (BERGERUD E REED, 1998 e SILVA e SUÁREZ, 2000).

2.2. AIDS E TRANSMISSÃO POR TRANSFUSÕES DE SANGUE

O HIV foi identificado pelo cientista francês Luc Montagnier em 1983. É um tipo de retrovírus da família *Retroviridae*, especificados em dois tipos: o vírus HIV-1 e o HIV-2, que são distintos dos pontos de vista geográfico e sorológico. O HIV-1 é encontrado nas Américas, Europa, África Subsaariana e na maioria dos outros países; o HIV-2 foi encontrado primeiramente na África Ocidental e alguns casos em outros países. O espectro clínico da infecção por HIV é contínuo, indo de portador

assintomático até SIDA (Síndrome de Imunodeficiência Adquirida, em inglês AIDS – Acquired Immune Deficiency Syndrome) avançada (ROUQUAYROL, 2003).

Apesar dos avanços verificados na medicina mundial, ainda não foi possível se obter a cura para a AIDS. Tendo como agente etiológico o HIV, a transmissão da síndrome pode ocorrer via relacionamentos sexuais, contactos com o sangue contaminado, e pela amamentação da mãe à criança.

O perfil epidemiológico da AIDS pode ser verificado graças as estatísticas oficiais por ser ela uma doença de notificação compulsória. Já o estudo da infecção pelo HIV ocorre, principalmente, de procedimentos matemáticos ou levantamentos de soroprevalência a infecção em populações-sentinelas (ROUQUAYROL, 1999).

Segundo dados da Secretaria de Saúde de Pernambuco (BOLETIM INFORMATIVO DST/AIDS, 2003), a epidemia da AIDS em Pernambuco teve início em 1983 e vem apresentando tendências crescentes. As taxas de detecção, entre 83 e 86, eram de 0,03 a 0,25 casos/100.000 habitantes, chegando a 7,96 casos/100.000 no ano 2000.

O número de casos de AIDS entre mulheres vem crescendo, sobretudo quando se analisa a proporção entre homens, pois esta razão que era de até 13 homens para cada mulher em 1987, nos últimos seis anos, está em torno de dois casos em homens para cada um caso em mulher (ROUQUAYROL, 1999 e BOLETIM INFORMATIVO DST/AIDS, 2003)

No Brasil, desde o começo da epidemia da AIDS, o grupo etário mais atingido é o de 25 a 39 anos de idade, perfazendo 60% do total de casos notificados até agosto/1998 (ROUQUAYROL, 1999). Em Pernambuco, ela vem também afetando segmentos cada vez mais jovens da população. Cerca de 65,9 % dos casos encontram-se na faixa etária de 20 a 39 anos, (BOLETIM INFORMATIVO DST/AIDS, 2002b), o que significa que estas pessoas vêm se infectando na adolescência.

Nos primeiros anos a epidemia em Pernambuco, atingiu mais acentuadamente a Região Metropolitana do Recife (RMR), que até hoje ainda é responsável por cerca de 83,0% dos casos, mas já se pode constatar a sua presença na maioria das cidades interioranas (BOLETIM INFORMATIVO DST/AIDS, 2002a).

A pauperização é outra tendência atual da epidemia, observada a partir do nível de escolaridade da população atingida nos últimos anos. Apresentando maior número de casos em pessoas analfabetas e com ensino fundamental,

representando cerca de 47% do total de casos no período de 1983/2003 (BOLETIM INFORMATIVO DST/AIDS, 2003).

Os procedimentos médicos, em geral, podem trazer consigo, além do potencial benefício, um risco inerente.

Segundo Canutti Júnior. (1998) em 1983, com o advento da AIDS, a preocupação com a segurança do sangue transfundido se traduziu, entre outras ações, na inclusão de um número maior de questões na entrevista para triagem de doadores, no intuito de identificar doadores com maior potencial de infecção com base em características comportamentais.

Testes para detecção de anticorpos contra o HIV foram introduzidos em 1985, nos EUA. Após essa introdução, acreditou-se que as transfusões tornar-se-iam isentas de risco. Entretanto, apesar de melhorias de resultados, no final dos anos 80 foram relatados casos de transmissão do HIV através de transfusões com sorologia negativa. Estas transfusões correspondem a doações realizadas durante o período de janela imunológica (COVAS, 1998). Período de janela imunológica é o período no qual o desenvolvimento de anticorpos contra HIV-1 permanece em níveis não detectáveis pelos testes sorológicos. Gerando a presença de um risco, que representa a possibilidade remanescente de uma unidade hemoterápica transmitir doença, mesmo após triagem clínica (entrevista e exame físico) e sorológica adequada (CANUTTI JÚNIOR, 1998).

Até o desenvolvimento dos testes de identificação da presença do HIV, numerosos indivíduos, receptores habituais de transfusões, como os hemofílicos, foram contaminados pelo vírus e desenvolveram a doença. Na atualidade, o grande alerta mundial para o problema da AIDS e o teste sistemático de todos os candidatos a doador, faz com que apenas raramente, a AIDS seja transmitida por transfusões.

Os indivíduos com maiores riscos de contaminação pelas transfusões são os portadores de distúrbios hereditários da coagulação, como a hemofilia. Os hemofílicos recebem freqüentes transfusões de fatores de coagulação, principalmente os hiper concentrados, para prevenir ou cessar hemorragias, tornando-se mais expostos. Além do risco de contaminação determinado pela quantidade de transfusões, o fato dos fatores serem concentrados, ou seja, providos pela coleta de sangue de várias centenas de indivíduos, também aumenta o risco da contaminação. Um único doador infectado, portanto, pode contaminar um lote inteiro

de produtos concentrados e, em conseqüência, transmitir a doença a diversos receptores.

Segundo Mcfarland (1998) e Kupek (2001) a prevenção da contaminação viral pela transfusão sanguínea requer as seguintes medidas: evitar transfusão de sangue desnecessária; exclusão dos doadores de sangue pertencentes aos grupos de comportamento de risco (captação de maior número de doadores com baixo risco de infecção), detecção sistemática de anticorpos anti-HIV-1 em todas as doações.

Os dados referentes à prevalência de infecção pelo HIV entre doadores de sangue são considerados por vários investigadores como sendo no início da epidemia, superestimados (em relação à população geral), porque admite-se que pessoas sob risco demandavam aos serviços de doação de sangue em busca, na verdade, de resultados da sorologia para HIV (ROUQUAYROL, 1999).

Como indicador de infecção, para extrapolar as informações a toda a população, os dados de hemocentros têm limitações, pela seletividade da amostra. No entanto, a padronização dos procedimentos, em especial, o método de triagem e as técnicas laboratoriais empregadas, tornaria os hemocentros locais privilegiados para estudo, em indivíduos adultos, da tendência secular destas infecções. Os achados, em doadores de sangue, tendem a estar bem correlacionados com os resultados de inquéritos populacionais, o que faz com que sejam postulados como indicadores epidemiológicos, de obtenção rápida, em substituição aos inquéritos, em locais onde tais investigações não possam ser realizadas (PEREIRA, 2000).

Diferentes estudos utilizando métodos estatísticos foram realizados considerando variáveis sócio-demográficas e marcadores sorológicos de doenças sexualmente transmissíveis, procurando associa-las ao risco de infecção pelo HIV, em doadores de sangue.

No Hemocentro Recife (PE), serviço em que a presente pesquisa foi realizada, Sampaio (2001) correlacionou a positividade sorológica com idade de candidatos à doação de sangue, verificando que aqueles com idade de 50 a 60 anos apresentaram predominância de positividade sorológica global sobre os candidatos de 18 a 28 anos, porém os mais idosos não revelaram resultados positivos para o anti-HIV, indicando uma tendência dessa faixa etária de apresentar menor risco para o receptor, podendo constituir o público alvo preferencial para a doação de sangue.

McFarrland *et. al.* (1998) concluíram em seu estudo, que viver numa área urbana de alta densidade, ser casado ou ter sido casado e ter idade acima de 22,5

anos foram considerados como preditores independentes da sorologia por HIV entre doadores de sangue em Zimbábue.

Segundo Souza (1998) observou que, indivíduos com escolaridade até o 1º grau apresentaram risco 2,2 (IC95% 1,1 – 4,8) vezes maior de ser infectados quando comparados com os de maior escolaridade.

Zachariah *et. al.* (2002), estudando doadores de sangue em Malawi, concluíram que os fatores de risco significativos associados com a prevalência do HIV para doadores homens e mulheres incluíram viver em uma área semi-urbana ou urbana e ser do grupo etário 25 - 29 anos para mulheres e 30 - 34 anos para homens.

Andrade Neto *et. al.* (2002) concluíram em pesquisa, realizada no hemocentro de Curitiba, quanto à prevalência de vírus HIV por idade, que não houve uma faixa etária mais acometida que as outras. E, ainda, que a prevalência por estado civil foi estatisticamente superior entre indivíduos solteiros e desquitados ou divorciados do que entre pessoas casadas ou viúvas.

Após a descoberta do HCV (Vírus da Hepatite C), a manutenção do anti-HBc foi questionada. Em 1995, o *National Institute of Health* (NIH) dos EUA recomendou a manutenção do teste anti-HBc porque com o seu potencial para prevenir alguns casos de hepatite HBV pós-transfusional, ele atua como um marcador indireto para infecção de HIV em doadores de sangue e poderia diminuir o risco residual da infecção de HIV transmitida por transfusão de sangue. Na ausência de estudos consistentes que permitiriam a exclusão dos testes anti-HBc, muitos países como EUA, França e Brasil têm mantido esses testes na triagem sorológica de doadores de sangue (SÁEZ-ALQUÉZAR *et. al.*, 2001 apud REICHE, 2003).

Alguns estudos como o de Reiche *et. al.* (2003) confirmaram os anti-HBc e anti-HCV como úteis marcadores indiretos para a infecção de HIV e reforçam a sua manutenção na triagem sorológica dos doadores de sangue contribuindo para a prevenção do pequeno número de casos em que o HIV é ainda transmitido por transfusão.

Em 1972, instituições americanas recomendaram a eliminação de Teste Sorológico para Sífilis (TSS) na triagem de doadores de sangue por causa de seu valor limitado em saúde pública, mas a recomendação não foi implementada. Em 1983, o TSS foi reavaliado e a Administração de Alimentos e Medicamentos decidiu conservá-lo por conta do seu valor potencial como marcador de comportamento de

indivíduos com risco de infecção de HIV. No entanto, nenhum estudo amplo estabeleceu que a TSS é um marcador valioso de infecção de HIV entre doadores de sangue (HERRERA *et. al.*, 1997 apud REICHE, 2003). Mas que, em seu estudo, concluiu que o TSS é um marcador pobre para prevenção de infecções de HIV (REICHE, 2003).

Rahlenbeck (1997) estudando doadores de sangue observou que os doadores HIV-positivos tiveram um risco aumentado para ser positivos para anticorpos da sífilis, mas não para HbsAg. Esses dados indicaram que a transmissão e a epidemia de HbsAg em Etiópia são diferentes daquela do HIV e da Sífilis.

Barcellos *et. al.* (2003) identificaram uma alta prevalência de infecção HIV entre indivíduos atendidos em centros de testagem anônima em Porto Alegre, no Rio Grande do Sul. E, segundo eles, alguns fatores de risco foram identificados para infecção por HIV, todos similares aos descritos em outros estudos, tais como: gênero masculino, idade superior a 20 anos, menos anos de frequência em escola e presença de sífilis.

3. MATERIAIS E MÉTODOS

3.1. MATERIAIS

3.1.1. Área de estudo

O estudo foi realizado no Hemocentro Recife da Fundação HEMOPE, Instituição vinculada à Secretaria Estadual de Saúde voltada para ações de assistência na área de hematologia e hemoterapia, com atendimento de 90% da demanda transfusional do Estado, desenvolvendo, também, programa de ensino e pesquisa científica no Estado de Pernambuco.

3.1.2. População de estudo e período de referência

A população estudada foi composta de candidatos à doação de sangue de primeira vez, aptos na triagem clínica e que participaram da triagem sorológica, no período de janeiro de 1998 a novembro de 2003.

3.1.3. Tamanho amostral

Realizou-se pré-estudo referente ao ano de 2002, onde se observou que de 18.603 doadores de 1ª doação do período apenas 51 (0,27%) foram diagnosticados como "HIV_positivo". Deste modo, os seguintes parâmetros foram utilizados para o cálculo do Tamanho da amostra (n):

- ✓ Tamanho da população: infinita,
- ✓ e = Erro máximo admitido =: 0,000311
- ✓ p = Proporção estimada de doadores de 1ª vez, aptos na TC, soropositivo para HIV = 0,0027
- ✓ $q = 1 - p = 0,9973$
- ✓ Intervalo de confiança: 95% $P\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}} \leq P \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}}\right) = 1 - \alpha$, e o

Cálculo do tamanho amostral: $n = \frac{z^2 \cdot (p \cdot q)}{e^2}$, onde

$$n = \frac{1,96^2 \times (0,0027 \times 0,9973)}{0,000311^2} \cong 106.950$$

$n = 106.950$ doadores consecutivos de 1ª VEZ aptos na TC.

Estudou-se 111.717 doadores de sangue de 1ª Vez, que foram aptos na triagem clínica, durante o período do estudo. Desses, verificou-se que 1.376 não participaram da triagem sorológica e foram retirados da pesquisa. Como também foram retirados da pesquisa 758 doadores que permaneceram inconclusivos para os testes sorológicos confirmatórios da infecção HIV, sendo ainda 2.743 registros retirados da amostra, devido a resultados “inconclusivos” para, pelo menos um dos marcadores de triagem das outras DST.

Analisando cada variável independente quanto ao registro “em branco” ou inconsistentes, à categorização do serviço e erro de digitação, retirou-se mais 692 doadores, que tinham essas características, pelo menos em uma dessas variáveis.

Assim, um total de 5.514 doadores foi excluído, o que corresponde a 4,94% da população inicial. Portanto a amostra final consistiu em 106.203 doadores de sangue de 1ªVez apto na Triagem Clínica, sendo que 217 doadores foram soropositivo para HIV, satisfazendo ao critério de (CONCATO, 1993), considerando que nesse estudo observou-se 11 variáveis independentes.

3.1.4 Desenho do estudo

Trata-se de estudo epidemiológico de corte transversal, no qual fator e efeito são observados num mesmo momento histórico, evidenciando as características e correlações naquele momento (ROUQUAYROL, 2003). Salvo exceções muito específicas, as únicas conclusões legítimas derivadas da análise de estudos de prevalência restringem-se a relação de associação, e não de causalidade. Em termos estatísticos, pode-se, no máximo, estabelecer que a causa suspeita e o efeito se encontram associados dentro de um certo nível de significância. Caso vinculados à ocorrência de doenças nessa etapa de análise, os fatores de exposição suspeitos passam a ser reconhecidos como fatores de risco em potencial (ROUQUAYROL, 2003), de acordo com o esquema da figura 1 (PEREIRA, 2000).

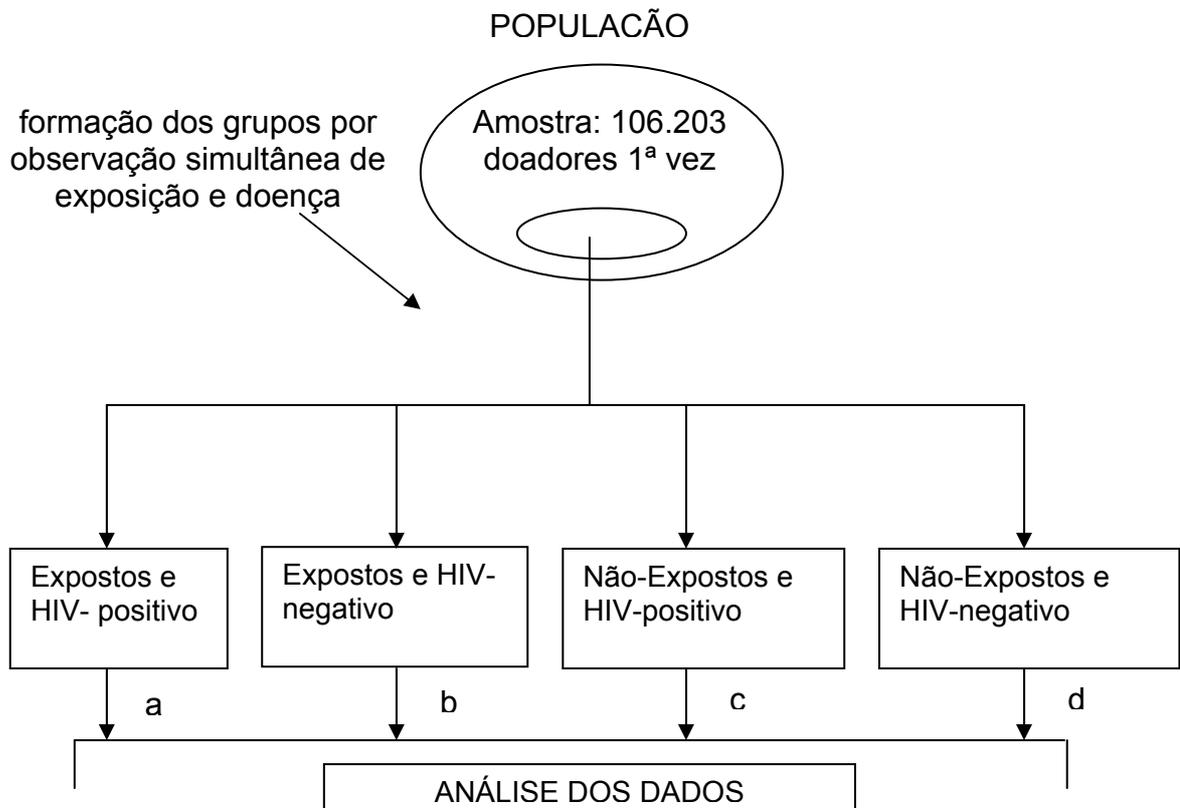


Fig. 1. Estrutura de um estudo transversal.
a, b, c, d – os quatro possíveis resultados.

3.1.5. Elenco das variáveis

As variáveis estudadas foram classificadas como: dependente (a infecção por HIV) e independentes (as de natureza sócio-demográfica: idade, gênero, situação conjugal, escolaridade e região de residência; as de natureza sorológica: status sorológicos Anti-HBc, HbsAg, Anti-HCV, Anti-HTLV e Sífilis; e o tipo de doação), categorizadas conforme quadro 1.

DEFINIÇÃO DA VARIÁVEL		TIPO	CATEGORIZAÇÃO/ CODIFICAÇÃO
Y	<i>Infecção por HIV</i>	<i>Dependente</i>	<i>1 - Positivo; 0 - Negativo.</i>
X1	<i>Gênero</i>	<i>Independente</i>	<i>1 - Masculino; 0 - Feminino.</i>
X2	<i>Idade</i>	<i>Independente</i>	<i>1 - 18 a 28; 2 - 29 a 39; 3 - 40 a 49; 4 - ≥ 50.</i>
X3	<i>Escolaridade</i>	<i>Independente</i>	<i>1 - Analfabeto; 2 - Fundamental; 3 - Médio; 4 - Superior.</i>
X4	<i>Situação conjugal</i>	<i>Independente</i>	<i>1 – Solteiro/ Divorciado/ Outros; 0 – Casado/ Viúvo.</i>
X5	<i>Região de Residência</i>	<i>Independente</i>	<i>1 - RMR; 0 - Interior/ outros estados.</i>
X6	<i>Tipo de doação</i>	<i>Independente</i>	<i>1 - Autóloga/ específica; 2 - Voluntária 3 – Reposição.</i>
X7	<i>Status sorológico do Anti-Hbc</i>	<i>Independente</i>	<i>1 - Positivo; 0 - Negativo;</i>
X8	<i>Status sorológico do HBsAg</i>	<i>Independente</i>	<i>1 - Positivo; 0 - Negativo;</i>
X9	<i>Status sorológico do Anti-HCV</i>	<i>Independente</i>	<i>1 - Positivo; 0 - Negativo;</i>
X10	<i>Status sorológico do Ant-HTLV</i>	<i>Independente</i>	<i>1 - Positivo; 0 - Negativo;</i>
X11	<i>Status sorológico do Sífilis</i>	<i>Independente</i>	<i>1 - Positivo; 0 - Negativo;</i>

Quadro 1. Variáveis relevantes associadas à infecção HIV, que entraram no estudo.

3.1.5.1. Variável dependente

- Infecção por HIV

Embora todo indivíduo sexualmente ativo corra risco de contaminação, estes riscos não são igualmente distribuídos. Eles variam em função de certas práticas

que constituem comportamentos de risco, propiciadores de maior contato com os agentes da doença. São exemplos: as relações homossexuais com vários parceiros, o uso de drogas ilícitas injetáveis e a prostituição (PEREIRA, 2000). Essas variáveis fazem parte da rotina de perguntas investigadas na triagem clínica, que classificam os doadores de sangue em aptos ou inaptos a doação.

Como se trabalhou com uma população de doador sanguíneo que foi aprovada na triagem clínica, considerada de bom comportamento, procurou-se estudar outras variáveis que fossem indiretamente relacionadas ao comportamento de risco, quanto ao aspecto epidemiológico da infecção HIV, e que tinham maior probabilidade de serem respondidas corretamente. Isso para definir fatores de riscos em potenciais, de modo a influenciar a direção das medidas de prevenção e controle, com vista a proteger os receptores.

Incluiu-se a análise de marcador sorológico para os anticorpos da hepatite B e C, HTLV e sífilis, por alguns serem considerados marcadores associados à infecção por HIV, em outros estudos. E estarem relacionados também à doenças sexualmente transmissíveis.

A infecção do HIV, neste estudo foi categorizada como: “1” em caso de doador soropositivo (indivíduo portador do vírus HIV) e “0” doador soronegativo (indivíduo não portador do vírus HIV).

3.1.5.2 Variáveis independentes

- Gênero

A razão homem/mulher expressa a distribuição dos casos de AIDS notificados segundo sexo. Ao longo do tempo, tem-se verificado que o excesso de casos do sexo masculino vem se reduzindo. De uma razão de 16H:1M, em 1984, constata-se que em 1990 alcançou 6H:1M, tendo atingido 2H:1M, em 1997/98. A diminuição da magnitude da razão de sexo tem levado a um diagnóstico atual de feminização da epidemia no País (ROUQUAYROL, 1999)

Foi tarefa levantar estas diferenças, apresenta-las objetivamente em relação à infecção HIV em doadores de sangue.

As categorias foram agrupadas em masculino e feminino, conforme definido pelo serviço.

- Idade

Dentre todas as variáveis relacionadas ao indivíduo, a idade é aquela que soma maior quantidade de relatos em estudos de epidemiologia (ROUQUAYROL, 2003). A relação entre idade e incidência da AIDS é muito evidente, e tendências no Brasil, incluindo Pernambuco, apontam para indivíduos mais jovens.

Foi computada em termos de anos decorridos até a data da doação, e distribuída em faixas etárias de 18 a 28; 29 a 39; 40 a 49 e maior ou igual a 50 anos. Correspondentes à exigência de idade para se doar sangue, que é de 18 a 65. Foi considerada na regressão logística como três variáveis dummy, utilizando-se a categoria maior ou igual a 50 anos(4) como o grupo de referência.

- Escolaridade

Visou-se avaliar neste estudo a possível relação entre a escolaridade e a infecção por HIV, em doadores de sangue, uma vez que a escolaridade estar relacionada estreitamente com o nível de saúde das pessoas. Há muitas hipóteses para explicar esta relação. Uma das razões aventadas é a de que os conhecimentos adquiridos permitem eleger modos de vida mais saudáveis, em acordo com o que já se sabe sobre as relações entre comportamento e saúde (PEREIRA, 2000).

As categorias foram agrupadas/codificadas em (1) analfabeto, (2) ensino Fundamental (1º Grau Completo ou Incompleto), (3) ensino Médio (2º Grau Completo ou Incompleto) e (4) ensino Superior (3º Grau Incompleto ou Completo, com ou sem Pós-graduação), conforme o critério do serviço. E tratada na regressão logística como três variáveis dummy, utilizando-se a categoria superior(4) como o grupo de referência.

- Situação conjugal

A situação das pessoas está associada ao estado civil. Casados têm diferentes modos de vida; esta diferença é associada, de maneira causal, a eventos favoráveis e desfavoráveis. A família confere proteção e segurança, que se refletem no nível de saúde. O casamento, ou, mais precisamente, a vida em comum, costuma trazer mudanças no comportamento da pessoa até então solteira. A mudança de estado civil também altera os padrões de comportamento sexual; logo, também se alteram os riscos venéreos – em geral, o casado está mais “protegido”. Os aspectos

positivos da união, em geral, pesam mais, devido à estabilidade de vida, à proteção mútua e à maior possibilidade de lidar com o estresse da vida moderna (PEREIRA, 2000).

Como a infecção HIV estar relacionada diretamente com aspectos comportamentais, precisamente sexuais, procurou-se estudar a relação entre a situação conjugal e essa infecção em doadores de sangue.

A situação conjugal foi coletada pelo serviço, de acordo com a condição declarada pelo doador: casado, solteiro, divorciado/desquitado, viúvo, e outros (na maior parte, indivíduos que se definiram como “amasiados”). Mas devido a esta variável não poder representar as formas de união, a categorização foi feita de acordo com o estudo de Andrade Neto *et. al.*(2002): (1) solteiro/ desquitado/ divorciado ou outros e (0) casado ou viúvo.

- Região de residência

A análise das regiões metropolitanas é uma possibilidade de investigar a magnitude da urbanização. È conglomerados que contêm a população de mais de um município, integrada social e economicamente (PEREIRA, 2000).

Estudou-se a região metropolitana brasileira do Recife, assim denominada em decreto do Governo Federal, de 1973, composta pela capital e arredores, num total de 14 municípios, tais como: Abreu e Lima, Araçoiaba, Cabo de Santo Agostinho, Camaragibe, Igarassu, Ipojuca, Itamaracá, Itapissuma, Jaboatão dos Guararapes, Moreno, Olinda, Paulista, Recife, e São Lourenço da Mata, quanto à associação com a Infecção por HIV, em doadores de sangue.

A categorização/codificação utilizada para Região foi que o doador reside na (1) RMR e (0) reside no Interior ou em Outros estados.

- Tipo de doação

Conforme o critério do serviço: Doação Autóloga (autodoação) - é aquela em que o candidato será o receptor; Doação Específica - aquela em que o candidato possui o mesmo tipo de sangue do paciente (para transfusão ou cirurgia) e faz parte do círculo familiar do mesmo; Doação Voluntária é quando o candidato se dirige ao Hemocentro independente de convocação e sim pelo espírito altruísta e de solidariedade. Seu sangue, desde que compatível, poderá ser liberado para

qualquer receptor; e Doação de Reposição - é aquela em que o candidato realiza doação para um determinado paciente que deverá receber ou já recebeu transfusão. Neste caso, o doador não precisará pertencer ao grupo sanguíneo do paciente e no ato da doação deverá informar o nome do paciente para o qual o sangue está sendo repostado ou do Hospital que encaminhou.

O objetivo foi verificar se existe associação do tipo de doação de sangue com a infecção HIV. Segundo a natureza da doação optou-se categorizar por (1) autóloga ou específica; (2) voluntária e (3) reposição. Para melhor interpretar as OR dessa variável, utilizou-se variável dummy, cujo grupo de referência foi variável autóloga ou específica(1).

- Status sorológico do Anti-Hbc

São anticorpos contra a partícula do core da hepatite B. Aparecem na circulação imediatamente após HbsAg, tem um aumento rápido no título e podem persistir por muitos anos após o desaparecimento do HbsAg. Durante o período da janela imunológica, pode ser o único sinal sorológico da exposição ao HBV (FUNDAÇÃO DE HEMATOLOGIA E HEMOTERAPIA DO ESTADO DE PERNAMBUCO, 2001).

Em alguns bancos de sangue em regiões com alta prevalência de infecção de HIV, o anti-HBC foi implementado como um marcador indireto para a infecção por HIV (REICHE *et al.*, 2003). Nesse estudo, procurou-se verificar essa evidência no Hemocentro Recife.

Categorizou-se como “1” o resultado reagente a esse marcador e “0” o resultado não reagente.

- Status sorológico do HbsAg

É o antígeno de superfície do Vírus da Hepatite B, antigamente conhecido como “Antígeno Austrália”. Aparece nas infecções agudas e crônicas, nos “portadores assintomáticos” e em muitos carcinomas hepatocelulares. A persistência do HbsAg por mais de quatro meses após o início da infecção aguda pelo vírus B sugere cronificação (MENDES, 1993).

A infecção pelo VHB tem distribuição epidemiológica universal, observando-se taxas de prevalência do HbsAg no soro, mais elevadas nos trópicos e em países

do chamado “Terceiro Mundo”, o que enfatiza a estreita correlação entre a disseminação da infecção e as características sócio-econômicas. E por ser uma infecção sexualmente transmissível, procurou-se verificar a possível associação do marcador HbsAg com a infecção por HIV.

Categorizou-se como “1” o resultado reagente a esse marcador e “0” o resultado não reagente.

- Status sorológico do Anti-HCV

O vírus da Hepatite C é conhecido hoje como vírus NÃO-A, NÃO-B, associado à transfusão, foi descoberto em 1988 (MENDES,1993) e a subsequente introdução de testes específicos de triagem para anti-HCV reduziram a taxa de 90% de infecção de hepatite após transfusão. Os anticorpos anti-HCV são detectados de 12 a 20 semanas depois do princípio de sinais e sintomas clínicos o que chamamos de janela imunológica (SÁEZ-ALQUÉZAR; SABINO, 2001 apud REICHE *et. al.*, 2003).

A infecção pelo vírus da hepatite C, geralmente é moderada e assintomática, apresenta distribuição global, aparecendo cerca de 150.000 novos casos por ano. A infecção ocorre principalmente por via sangüínea, com maior freqüência em pessoas que fazem uso de seringas compartilhadas. Dentre as formas não-parenterais de transmissão torna-se importante ressaltar a possibilidade da transmissão sexual (MENDES,1993).

Nesse estudo, procurou-se verificar a possível associação desse marcador com a infecção por HIV, no hemocentro Recife.

Categorizou-se como “1” o resultado reagente a esse marcador e “0” o resultado não reagente.

- Status sorológico do Anti-HTLV I / II (vírus linfotrópico da célula T humana tipo I e II)

O vírus HTLV-I foi descrito em 1980 como o primeiro retrovírus humano, o qual foi isolado de um paciente com leucemia de células T. É o agente etiológico de um tipo de leucemia e paralisia, sua transmissão se dar por via parenteral, vertical e sexual (MURRAY *et. al.*, 1992). O HTLV II foi identificado em 1982, numa linhagem de células T estabelecida de um paciente com leucemia de células T cabeluda

(FUNDAÇÃO DE HEMATOLOGIA E HEMOTERAPIA DO ESTADO DE PERNAMBUCO, 2001).

O HTLV-1 e o HIV-1 infectam os linfócitos T CD4, são transmitidos por formas semelhantes e possuem longos períodos de latência, mas causam doenças muito diferentes devido à natureza de sua interação com a célula do hospedeiro infectada (MURRAY *et. al.*, 1992). E por essa razão procurou-se verificar a possível associação do marcador Anti-HTLV com a infecção por HIV.

Categorizou-se como “1” o resultado reagente a esse marcador e “0” o resultado não reagente.

- Status sorológico do Sífilis

A sífilis é uma doença infecciosa crônica, cujo agente etiológico é o *Treponema Palidum*, caracterizada por lesões mucocutâneas recidivantes. A principal via de transmissão é a sexual; ocasionalmente pode transmitir-se através de transfusão sanguínea, ou por via vertical (ROUQUAYROL, 1993). É mais comum em indivíduos jovens sexualmente ativos, dos grandes centros urbanos.

E pelas mesmas razões já citadas em relação aos outros marcadores, procurou-se investigar a possível associação do marcador para sífilis com a infecção por HIV.

Categorizou-se como “1” o resultado reagente a esse marcador e “0” o resultado não reagente.

3.1.6. Coleta de Dados

Os critérios, tanto de pré-triagem como de triagem clínica e sorológica aplicados aos candidatos à doação, não foram criados, especificamente, para o presente estudo. Foram utilizados os mesmos critérios rotineiros empregados pela Fundação HEMOPE, para a seleção de candidatos e tais critérios obedeceram à legislação brasileira em vigor no período do estudo (BRASIL, 1993; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 1998; BRASIL, 2002; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2003).

A pré-triagem e a triagem clínica foram executadas por auxiliares de enfermagem/enfermeiras e por médicos, respectivamente, todos profissionais habilitados da equipe técnica da Fundação HEMOPE. Utilizou-se a Ficha de Doação

(anexo I), com dados da triagem e da doação, na qual constavam perguntas objetivas e técnicas e os médicos que as aplicavam estavam treinados a fazê-lo em linguagem acessível, e a explicar o significado de cada questão.

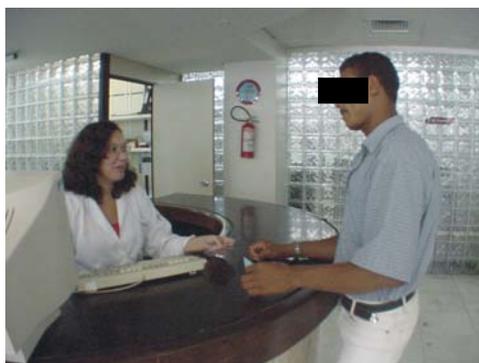
As informações obtidas na triagem clínica constavam das perguntas (anexo I) formuladas pelos médicos e, no momento em que eram questionados eram verificados itens relacionados à garantia da segurança do doador e do receptor. O procedimento classificava o candidato à doação em Apto e Inapto.

Considerou-se para esse estudo apenas os dados referentes aos doadores de “Primeira Vez” ou “de Primeira Doação” aquele que na data da triagem nunca tinha feito uma outra doação no Hemocentro. Caracterizando um total desconhecimento dos médicos sobre os aspectos comportamentais desses candidatos. Ressaltando que é nessa etapa em que a pesquisa deseja otimizar, indicando possíveis determinantes da infecção HIV, que deverão ser observados para a segurança do receptor.

Todos os candidatos que se apresentaram à doação no período do estudo, obrigatoriamente, passaram por essas etapas rotineiras do serviço descritas abaixo:

- Identificação

O candidato à doação que se dirigiu à recepção de doadores do hemocentro Recife (Figura 2) apresentou sua carteira de identidade ou profissional sendo preenchida sua ficha cadastral e observado se já tinha realizado alguma doação anteriormente, pois, caso contrário, era considerado doador de sangue de primeira vez, objeto do estudo. Nessa etapa se colhiam os dados sobre as variáveis: gênero, idade, região de residência, situação conjugal e tipo de doação.



*Figura 2. Recepção de doadores do Hemocentro Recife.
Fonte: Setor Comunicação/ Hemope.*

- Pré-triagem

Etapa em que se verificava a pressão arterial, temperatura, pulso, peso e altura, bem como a verificação do exame para detecção de anemia pela análise do resultado da dosagem da hemoglobina (Figura 3).



*Figura 3. Setor de Pré Triagem do Hemocentro Recife.
Fonte: Setor Comunicação/ Hemope.*

- Triagem Clínica

Fase do exame médico para verificação da situação clínica do candidato, apto ou inapto para a doação, através de entrevista individual, confidencial e sigilosa (Figura 4). Nesse exame, entre outras coisas, o candidato era investigado quanto à presença de doenças que já teve e aspectos de seu comportamento sexual. A questão mais importante, nessa etapa, era se o candidato tinha comportamento de risco para doenças transmissíveis pelo sangue.



*Figura 4. Setor de Triagem Clínica do Hemocentro Recife.
Fonte: Setor Comunicação/ Hemope.*

- Coleta de sangue

Sendo o candidato apto à doação, o mesmo era encaminhado à sala de coleta (Figura 5). De cada doação foi retirado cerca de 450ml de sangue do doador, em um circuito fechado de bolsa plástica, não permitindo qualquer contato com o

meio exterior. Ao término da doação de sangue, uma amostra de sangue do doador era coletada em dois tubos, a fim de realizar testes laboratoriais de sorologia e imunohematologia.



*Figura 5. Sala de Coleta do Hemocentro Recife.
Fonte: Setor Comunicação/ Hemope.*

- Triagem sorológica

Os candidatos à doação foram submetidos a testes de triagem sorológica para detecção de anticorpos para as seguintes viroses: vírus da Imunodeficiência Humana (anti-HIV), vírus da Hepatite B-antígeno do core (anti-Hbc), da Hepatite B-antígeno de superfície (HbsAg), vírus da Hepatite C (Anti_HCV), vírus linfotrópicos de células T-Humanas do tipo I e II (Anti_HTLV I/II) e Anticorpos para sífilis e chagas. Todos esses parâmetros são obrigatórios por determinação (BRASIL, 2002).

Neste estudo, só foram observados os testes sorológicos para detecção dos seguintes marcadores: anti_HIV, Anti-HBc, Hbs_Ag, Anti-HCV, Anti-HTLV I/II e para sífilis. Não foi incluído teste para Chagas, devido ao mesmo não ser relacionado a doenças sexualmente transmissíveis.

As informações obtidas do laboratório de sorologia (figura 6) constavam dos resultados positivos, negativos e inconclusivos dos testes para a exclusão das hepatites B e C, Sífilis, AIDS e HTLV-I/II. Dos quais, para esse estudo foram retirados os inconclusivos.

A metodologia utilizada pelo laboratório de sorologia do hemocentro Recife, no período da pesquisa foi “ELISA”, que é um teste imunoenzimático de detecção de anticorpos para HIV, HBC, HCV, HbsAg e HTLV. Para sífilis foi usado VDRL, que é um teste de floculação.

Para a infecção pelo HIV além dos testes de triagem foram feitos testes confirmatórios como a Imunofluorescência Indireta e Western Blot. Seguindo o fluxograma para a confirmação do HIV nas Unidades Hemoterápicas (Brasil, 1998).

Diagnosticando, todos os doadores observados nesta pesquisa, nas categorias: “positivo” e “negativo” para a infecção pelo HIV.



*Figura 6. Laboratório de Sorologia do Hemocentro Recife.
Fonte: Setor Comunicação/ Hemope.*

3.1.7. Processamento dos dados

Os procedimentos executados rotineiramente sobre uma doação são registrados no Sistema de Banco de Sangue – SBS, utilizado pela Fundação Hemope. Do qual foram obtidos os dados dos doadores de sangue estudados neste trabalho.

Foi solicitada a elaboração de um programa que gerasse um arquivo texto com todas os dados referentes as variáveis da pesquisa para a empresa SBS Sistemas Administração S/C Ltda, responsável pelo SBS.

A partir daí foram feitas às consistências e codificações dos dados através do Excel, sendo armazenados em arquivo com extensão DAT, e exportado para os pacotes de análise estatística: o SPSS versão 10 onde se processou a regressão logística clássica; e o WINBUGS 14 a regressão logística com enfoque bayesiano.

3.1.8. Aspectos éticos

O presente estudo foi submetido ao Comitê de Ética em Pesquisa (CEP) da Fundação HEMOPE e aprovado em 03 de dezembro de 2003.

3.2. MÉTODOS

3.2.1. Modelo de regressão logística binária

Um dos casos particulares dos modelos lineares generalizados são os modelos para variáveis que apresentam apenas duas categorias ou que foram de alguma forma dicotomizadas as variáveis que assumem valores “0” ou “1”. São chamadas variáveis “dummy”. Um dos mais importantes modelos é o de regressão logística, baseado na transformação logit para proporção.

Variáveis com duas categorias que podem ser classificadas como sucesso ou fracasso representando as possibilidades de respostas como, por exemplo 0 e 1, podem ser caracterizadas pela distribuição de Bernoulli. Comumente é chamado de “sucesso” o resultado mais importante de resposta ou aquele resultado que se pretende relacionar com outras variáveis de interesse. A distribuição de Bernoulli para a variável aleatória binária Y de parâmetro p_i especifica as probabilidades como:

$$P(Y_i = 1) = p_i \quad e \quad P(Y_i = 0) = 1 - p_i$$

Pois por definição,

$$E(Y_i) = 1(p_i) + 0(1 - p_i) = p_i$$

Que é a proporção de respostas em que $Y_i = 1$ e sendo,

$$Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = 1^2(p_i) + 0^2(1 - p_i) - (p_i)^2 = (p_i) - (p_i)^2 = p_i(1 - p_i)$$

Isto é,

$$Var(Y_i) = p_i(1 - p_i)$$

Então para esta variável aleatória a função de probabilidade é

$$f(Y_i; p_i) = p_i^{y_i} (1 - p_i)^{1 - y_i} \quad i = 1, \dots, n$$

No presente trabalho, $Y_i = 0$ designa “doador HIV-negativo” e $Y_i = 1$ designa “doador HIV-positivo”. Logo Y é uma v.a. Bernoulli.

- O modelo linear tradicional

Seja Y , uma variável dependente(resposta), e sejam $(p-1)$ variáveis independentes (fatores) X_1, X_2, \dots, X_{p-1} . O objetivo é a construção de um modelo que relacione a variável aleatória Y com as variáveis independentes X 's fixadas, dispondo-se para tanto de n observações, escritas na forma:

$$(Y_i, X_{i1}, X_{i2}, \dots, X_{i(p-1)}) \quad \text{onde} \quad i = 1, 2, \dots, n.$$

O modelo procurado pode ser escrito como

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} + \varepsilon = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ij} + \varepsilon \quad (1)$$

que, na notação matricial torna-se

$$\underline{Y} = \underline{X} \underline{\beta} + \underline{\varepsilon}$$

onde \underline{Y} é o vetor de dimensão $n \times 1$, dos valores observados da variável resposta.

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

\underline{X} é a matriz de ordem $n \times p$ do modelo, contém os n valores observados das variáveis explicativas.

$$\underline{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1(p-1)} \\ 1 & x_{21} & \dots & x_{2(p-1)} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{n(p-1)} \end{bmatrix}$$

$\underline{\beta}$ é o vetor dos parâmetros desconhecidos de dimensão p , e que devem ser estimados com base nos valores observados para \underline{Y} e os \underline{X} .

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_{p-1} \end{bmatrix}$$

e $\underline{\varepsilon}$ é o vetor de dimensão $n \times 1$, dos erros aleatórios, associado ao modelo.

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}$$

ε_i são assumidos variáveis aleatórias, independentes, com distribuição normal de média zero ($E(\varepsilon_i) = 0$) e variância constante ($Var(\varepsilon_i) = \sigma^2$) $\Rightarrow \varepsilon_i \sim N(0, \sigma^2 I_n)$

Sendo o valor esperado de Y , denotado por $E(\underline{Y}) = X \underline{\beta}$

Em qualquer problema de regressão, a quantidade chave é o valor médio da variável dependente, dado o valor da variável independente. Tal quantidade será representada por $E(Y | X)$, que se lê “valor esperado para Y , dado o valor de X ”. Na regressão linear parte-se da suposição que esta quantidade pode ser expressa por uma expressão da forma:

$$E(Y/X) = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_j$$

Esta relação torna admissível a possibilidade de que $E(Y/X)$ possa assumir qualquer valor para $X \in (-\infty, \infty)$.

No entanto existem alguns problemas para os quais o modelo linear tradicional não é apropriado para a situação aqui tratada, variável resposta dicotômica, por exemplo:

- A suposição de normalidade pode não ser adequada para modelar contagens ou proporções;
- Se a média dos dados é restrita naturalmente a uma amplitude de valores, o modelo linear tradicional pode não ser apropriado, desde que o preditor linear $E(Y) = X\beta$ pode assumir qualquer valor na reta real. Um exemplo seria a média de proporções que assume um valor no intervalo $[0, 1]$;
- Pode não ser real assumir homocedasticidade, ou seja, que a variância dos erros seja constante para todas as observações.
- O modelo logístico

Na regressão logística o que se tem é $0 \leq E(Y/X) = P(Y = 1/X) \leq 1$. Assim,

o modelo logístico é baseado na função logística. A relação é dada por

$$Y = \frac{e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j X_j}}$$

Para a variável dependente Y relacionada com uma única variável independente X , ou com várias variáveis X 's, a função é chamada *Sigmóide*, e o seu gráfico tem a forma da figura 7, a seguir. É fácil perceber que

$$\begin{cases} X \rightarrow -\infty \Rightarrow Y \rightarrow 0 \\ X \rightarrow +\infty \Rightarrow Y \rightarrow 1 \end{cases}$$

e, também, que

$$X = 0 \text{ e } \beta_0 = 0 \Rightarrow Y = \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{1}{2}$$

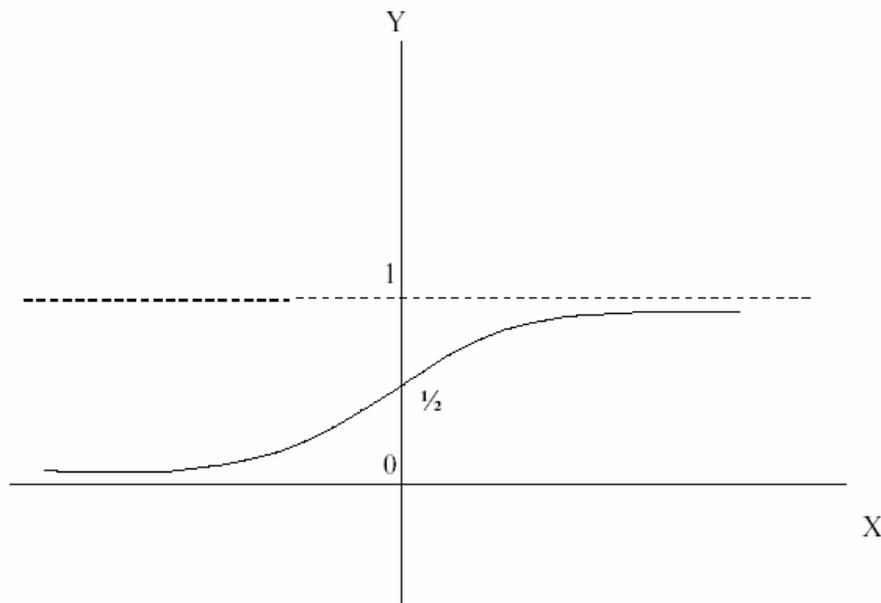


figura 7. Gráfico da função sigmóide

E essa é a principal razão da popularidade do modelo logístico, ele serve para descrever uma probabilidade, que no estudo corresponde ao risco individual de ser

um doador soropositivo para HIV.

A fim de simplificar a notação, a quantidade $E(Y/X)$, quando referente ao modelo logístico, será representada por $P(X)$, da seguinte forma:

$$P(X) = \frac{e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j X_j}} \quad (2)$$

Onde $Y = P(X) + \varepsilon$ e ε pode assumir um de dois possíveis valores:

- Se $Y=1$, então $\varepsilon = 1 - P(X)$, com probabilidade $P(X)$.
- Se $Y=0$, então $\varepsilon = -P(X)$, com probabilidade $[1 - P(X)]$.

Contudo, ε tem uma distribuição com média 0 e com a variância dada por $P(X)[1 - P(X)]$ isso dá uma distribuição de Bernoulli.

3.2.2. Transformação logit

Para evitar o problema restritivo de que os valores de probabilidade são números no intervalo de 0 e 1. A função logística pode ser linearizada pela transformação chamada logit.

A **transformação logit** é definida como

$$\text{logit } P(X) = \ln \left[\frac{P(X)}{1 - P(X)} \right]$$

$$\text{logit } P(X) = \ln \left[\frac{\frac{e^\mu}{1 + e^\mu}}{1 - \frac{e^\mu}{1 + e^\mu}} \right]$$

onde $\mu = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_j$

$$\text{logit } P(X) = \ln \left[\frac{\frac{e^\mu}{1 + e^\mu}}{1} \right]$$

$$\text{logit } P(X) = \ln[e^\mu]$$

$$\text{logit } P(X) = \mu$$

$$\text{logit } P(X) = \ln\left[\frac{P(X)}{1-P(X)}\right] = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_j \quad (3)$$

Para um conjunto de n observações independentes destas variáveis, escritas na forma $(Y_i, X_{1i}, X_{2i}, \dots, X_{(p-1)i})$, onde $i = 1, 2, \dots, n$.

Essa expressão (3) é Chamada função de regressão logística linear.

A importância dessa transformação é que o logit tem muito das desejáveis propriedades do modelo de regressão linear. O logit é linear em seus parâmetros, pode ser contínuo e pode alcançar de $-\infty$ a $+\infty$, dependendo do domínio de X .

3.2.3 Método Clássico de estimação dos Parâmetros

O método de estimação é a máxima verossimilhança, que consiste em maximizar a função de verossimilhança da amostra.

3.2.3.1 Modelo logístico simples

Seja uma amostra com n observações independentes de pares (X_i, Y_i) , $i = 1, 2, \dots, n$, onde os Y 's representam valores observados de uma variável *dicotômica*, isto é, Y assume seus valores no conjunto $\{0, 1\}$, e os X 's representam os valores observados de apenas uma variável independente., então a equação fica:

$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Para estimar o Modelo de Regressão Logística (MRL) é necessário estimar os valores para os parâmetros β 's, que são desconhecidos.

$P(X)$ dá a probabilidade condicional de que Y é igual a 1, dado o valor de X , isto será representado por $P(Y = 1/X)$. Segue que a probabilidade de que Y é igual a

0, dado o valor de X , fica

$$P(Y = 0/X) = 1 - P(X)$$

Desta forma

$$P(Y_i = 1/X_i) = P(X_i)$$

$$P(Y_i = 0/X_i) = 1 - P(X_i)$$

Seja a função de probabilidade de $Y \sim \text{Bernoulli}$

$$f(Y_i/X_i) = [P(X_i)]^{Y_i} [1 - P(X_i)]^{1-Y_i} \quad (4)$$

O método de estimação dos parâmetros na regressão logística clássica é a estimação por Máxima Verossimilhança (MV). Em linhas gerais, o Método da Máxima Verossimilhança fornece estimativas para os parâmetros que maximizam a probabilidade de obter o conjunto observado de dados. Para aplicar tal método deve-se, em primeiro lugar, construir a função chamada Função de Verossimilhança (FV). Os estimadores de máxima verossimilhança destes parâmetros são escolhidos entre aqueles que maximizam esta função. A Função de Verossimilhança é dada por

$$\lambda(\beta) = \prod_{i=1}^n f(Y_i/X_i) \quad (5)$$

A Máxima Verossimilhança implica que o estimador para β seja o valor que maximiza a expressão dada em (5). Contudo, é mais fácil, do ponto de vista matemático, trabalhar com o logaritmo da mesma. A expressão fica, então

$$L(\beta) = \ln[\lambda(\beta)]$$

$$L(\beta) = \ln \prod_{i=1}^n f(Y_i/X_i)$$

$$L(\beta) = \sum_{i=1}^n \{Y_i \ln[P(X_i)] + (1 - Y_i) \ln[1 - P(X_i)]\} \quad (6)$$

$$L(\beta) = \sum_{i=1}^n \left\{ Y_i \ln \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} + (1 - Y_i) \ln \left[1 - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right] \right\}$$

$$L(\beta) = \sum_{i=1}^n \left\{ Y_i (\beta_0 + \beta_1 X_i) - \ln(1 + e^{\beta_0 + \beta_1 X_i}) \right\}$$

Para obter $\underline{\beta}$ que maximiza $L(\beta)$ basta derivar a expressão em relação a β e igualar a zero as equações obtidas. As expressões resultantes são

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \quad (7)$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(X_i Y_i - \frac{X_i e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \quad (8)$$

ou, de outra forma,

$$\sum_{i=1}^n [Y_i - P(X_i)] = 0$$

$$\sum_{i=1}^n X_i [Y_i - P(X_i)] = 0$$

No Modelo Linear Geral as expressões (7) e (8) conduzem a um sistema de equações lineares, o que facilita em muito o cálculo dos estimadores para os parâmetros desconhecidos. No MRL as expressões são não lineares, requerendo, portanto, métodos especiais para a sua resolução. Tais métodos são iterativos, o que exige a utilização de *softwares* específicos, ou a construção de programas computacionais para implementação dos referidos métodos. O processo é iterativo, é dizer se dão aos coeficientes uns valores arbitrários, alguns pacotes estatísticos assumem zero como no SPSS. A partir destes valores iniciais se constrói uma matriz G com os valores previstos pelo modelo para as observações da variável dependente. A partir desta matriz e da matriz X de desenho (construída igual que em

regressão linear), calculam-se os novos estimadores, que se comprova se são já a solução, em cujo caso se pára o processo e em caso contrário se repete o processo. Na prática, e para evitar convergências assintóticas, também se os novos estimadores diferem dos anteriores em menos de uma certa quantidade, chamada limite de convergência.

3.2.3.2 Modelo logístico múltiplo

Sejam agora uma variável aleatória Y , que pode assumir qualquer um dos valores 0 ou 1, e $X' = (X_1, X_2, \dots, X_{p-1})$ um vetor de dimensão $(p - 1)$ de variáveis aleatórias independentes.

Sejam, também, n observações independentes destas variáveis, escritas na forma $(Y_i, X_{1i}, X_{2i}, \dots, X_{(p-1)i})$, onde $i = 1, 2, \dots, n$. O modelo de regressão logística pode também ser escrito na forma:

$$P(X) = \frac{e^{\beta_0 + \sum_{j=1}^{p-1} \beta_j X_j}}{1 + e^{\sum_{j=1}^{p-1} \beta_j X_j}}$$

O método usado para a estimação dos parâmetros é o mesmo adotado para o caso univariado, Máxima Verossimilhança. As equações de verossimilhança são obtidas pela derivação parcial em relação a cada um dos p parâmetros da função $L(\beta)$.

As expressões resultantes para as derivadas ficam

$$\sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (9)$$

$$\sum_{i=1}^n X_{ij} [Y_i - P(X_i)] = 0 \quad \text{para } j = 0, 1, 2, \dots, p-1. \quad (10)$$

Como no modelo univariado, as equações obtidas com a derivação da Função Verossimilhança não são lineares, aqui também são necessários métodos iterativos para a resolução do sistema de equações resultante. O vetor de soluções das equações será representado por $\hat{\underline{\beta}}$ e é chamado de estimador de máxima

verossimilhança de β . Desta forma, os valores estimados para o Modelo de Regressão Logística múltiplo são $\hat{P}(x_i)$, o valor da expressão (2) calculada usando $\hat{\beta}$ e X_i .

Neste caso a idéia que cerca a estimação dos parâmetros pode ser colocada alternativamente como

$$\min \sum_{i=1}^n \left[y_i - \hat{P}(X_i) \right]^2 = \min SQR \quad (11)$$

, onde *SQR* é a Soma dos Quadrados dos Resíduos.

Segundo *Walker & Duncan(1967) apud Guimarães (2000) e Martins(1988)*, a exata determinação dos parâmetros não é possível, em função da conhecida complexidade do problema resultante. Entretanto, considerando que o problema da estimação dos parâmetros pode ser entendido como um problema de minimização, pode-se considerar a utilização de algum método de minimização para a obtenção dos mesmos. Entre estes métodos, um dos mais conhecidos é o *Método de Newton-Raphson*, que apresenta a vantagem de convergir rapidamente para a solução.

Pelo teorema do limite central na estimação por máxima verossimilhança, diz-se que estes estimadores são assintoticamente normais e sua matriz de variância e covariância é obtida pela matriz de segunda derivada parcial da função de log verossimilhança (6). Essas derivadas parciais têm a seguinte forma geral:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n X_{ij}^2 P(X_i) [1 - P(X_i)] \quad (12)$$

e

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_u} = - \sum_{i=1}^n X_{ij} X_{iu} P(X_i) [1 - P(X_i)] \quad (13)$$

para $j, u=0, 1, 2, \dots, p-1$, e $i=1, \dots, n$.

Construir a matriz de dimensão p com os termos dados nas equações (12) e (13) e denotar como $I(\beta)$. Essa matriz é chamada a matriz de informação. A variância e covariância dos coeficientes estimados são obtidos pela inversa dessa

matriz, denota-se por $\Sigma(\beta) = I^{-1}(\beta)$. Conseqüentemente, usa-se a anotação $\sigma^2(\beta_j)$ para denotar o j-ésimo elemento da diagonal desta matriz, que é a variância de $\hat{\beta}_j$, e $\sigma(\beta_j, \beta_u)$ para denotar um elemento arbitrário fora-diagonal, que é a covariância de $\hat{\beta}_j$ e $\hat{\beta}_u$, onde $j \neq u$. Os estimadores das variâncias e covariâncias que são denotados por $\hat{\Sigma}(\hat{\beta})$, são obtidos calculando $\Sigma(\beta)$ utilizando $\hat{\beta}$. Usa-se $\hat{\sigma}^2(\hat{\beta}_j)$ e $\hat{\sigma}^2(\hat{\beta}_j, \hat{\beta}_u)$, $j, u = 0, 1, \dots, p-1$, para denotar os valores nesta matriz.

Logo, as estimativas dos erros padrões dos coeficientes estimados, são calculadas e denotadas como:

$$\hat{SE}(\hat{\beta}_j) = [\hat{\sigma}^2(\hat{\beta}_j)]^{1/2} \quad \text{para } j = 0, 1, 2, \dots, p-1 \quad (14)$$

Usa-se esta anotação no desenvolvimento de métodos para testar coeficiente e estimação de intervalo de confiança.

Uma formulação da matriz de informação é $\hat{I}(\hat{\beta}) = X'VX$ onde X é uma matriz $n \times p$, contendo o dado para cada sujeito, e V é uma matriz diagonal de dimensão $n \times n$ com elemento geral $\hat{P}(X_i)[1 - \hat{P}(X_i)]$. Isto é:

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1(p-1)} \\ 1 & x_{21} & \dots & x_{2(p-1)} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{n(p-1)} \end{bmatrix}$$

e

$$V = \begin{bmatrix} \hat{P}(x_1)[1 - \hat{P}(x_1)] & 0 & \dots & 0 \\ 0 & \hat{P}(x_2)[1 - \hat{P}(x_2)] & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{P}(x_n)[1 - \hat{P}(x_n)] \end{bmatrix}$$

A forma do método de Newton é:

$$\beta_{k+1} = \beta_k + \{I(\beta_k)\}^{-1} \{[S(\beta_k)]\}$$

onde

$$S(\beta_k) = \begin{bmatrix} \frac{\partial L(\beta)}{\partial \beta_0} \\ \dots \\ \frac{\partial L(\beta)}{\partial \beta_{p-1}} \end{bmatrix} \text{ função escore (1ª derivadas)}$$

Com aproximação inicial dada, neste trabalho por $\beta'_{k=0} = [0 \dots 0]$ no pacote estatístico SPSS versão 10.

- Estimação por intervalos

Tendo em conta o anterior, um intervalo de confiança em nível de $(1-\alpha)\%$ para o coeficiente β_j é:

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} \hat{SE}(\hat{\beta}_j) \quad (15)$$

3.2.3.3. Testes de hipóteses sobre os coeficientes

Depois de calcular os coeficientes, o passo seguinte consiste em determinar se as variáveis independentes são associadas significativamente à variável de resultado.

Os métodos de inferência nos modelos lineares generalizados baseiam-se, fundamentalmente, na teoria de máxima verossimilhança. Os testes a seguir mencionados são assintoticamente equivalentes, sendo a razão de verossimilhança o critério que define um teste uniformemente mais poderoso (CORDEIRO, 1986; DEMETRIO, 2002; PAULA, 2004). De acordo com essa teoria, existem duas estatística mais importantes, que foram utilizadas, neste estudo, para testar hipóteses relativas ao vetor de parâmetros $\underline{\beta}$, que são deduzidos de distribuições assintóticas.

A primeira denominada estatística de Wald (1943) é baseada na distribuição assintótica de $\underline{\hat{\beta}}$ e é geralmente mais usada no caso de hipóteses relativas a um único coeficiente β_j . É uma generalização da estatística t de Student. A estatística do teste:

$$\text{Como } \beta_j \sim N(0, \sigma^2(\beta_j)) \text{ e sob o teste de hipótese } \begin{cases} H_0 : \beta_i = 0 \\ H_a : \beta_i \neq 0 \end{cases} \text{ é}$$

$$W_j = \frac{\hat{\beta}_j}{\hat{SE}(\hat{\beta}_j)} \sim N(0,1) \quad (16)$$

Onde $\hat{SE}(\hat{\beta}_j)$ é o desvio padrão de $\hat{\beta}_j$, visto em (14)

A segunda estatística é a razão de verossimilhança e é geralmente preferida no caso de hipóteses relativas a vários parâmetros β 's.

Utilizar o teste da razão de verossimilhança para verificar a adequação do modelo como um todo, isto é:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0 \\ H_a : \text{O Modelo é adequado} \end{cases}$$

A estatística é:

$$G = -2 \ln \left[\frac{\lambda(\text{modelo sem variável})}{\lambda(\text{modelo com variável})} \right] \sim \chi_k^2, \text{ onde } k \text{ é o número de } \beta \text{'s}$$

("variáveis") no modelo e λ é a função de verossimilhança (5) para o modelo múltiplo. Ou seja:

$$G = 2[L(\hat{\beta}) - L(\beta^0)] \sim \chi_k^2 \quad (17).$$

onde β^0 é o vetor de parâmetro sob a hipótese nula e L é a função de log-verossimilhança (6) para o modelo múltiplo.

Rejeita-se a hipótese de nulidade se o p-valor associado com esse teste é $P[\chi^2(k) > G] < 0,05$, isto é rejeitar H_0 ao nível de significância de 5%.

3.2.3.4. Bondade do ajuste

Sempre que se constrói um modelo de regressão é fundamental, antes de passar a extrair conclusões, comprovar que o modelo calculado se ajusta efetivamente aos dados usados para estimá-lo.

No caso da regressão logística uma idéia bastante intuitiva é calcular a probabilidade de aparição do acontecimento, presença da infecção por HIV para todos os doadores da amostra. Se o ajuste é bom, é de se esperar que um valor alto de probabilidade se associe com presença real da infecção por HIV, e vice-versa, se o valor dessa probabilidade calculada é baixo, cabe esperar também ausência de infecção por HIV.

Esta idéia intuitiva se leva a cabo formalmente mediante o teste de Hosmer-Lemeshow (1989), que basicamente consiste em dividir o percurso da probabilidade em “g” grupos de decis de risco (isto é probabilidade de HIV-positivo $<0,1$; $<0,2$ e assim até <1) e calcular tanto a distribuição de HIV-positivo, como HIV-negativo prevista pela equação e os valores realmente observados. A estatística Hosmer-Lemeshow é obtida, calculando-se a estatística do qui-quadrado de Pearson da tabela $2 \times g$ das freqüências observadas e previstas, onde g é o número de grupos.

A estatística é escrita

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n_k \hat{P}(X)_k)^2}{n_k \hat{P}(X)_k (1 - \hat{P}(X)_k)} \quad (18)$$

Onde, n_k é o número total de doadores no k -ésimo grupo.

$$O_k = \sum_{j=1}^{c_k} Y_j \quad \text{é o número de respostas entre as } c_k \text{ fatores padrões (é a}$$

freqüência total de resultados do evento *no k-ésimo grupo*) e $\hat{P}(X)_k = \sum_{j=1}^{c_k} \frac{m_j \hat{P}(X)_j}{n'_k}$

é a probabilidade média estimada de um resultado do evento para o k -ésimo grupo e c_k denota o número de fatores padrões no k -ésimo decil e m_j é o número de doadores com $X=X_j$, num total de n amostra.

Ambas as distribuições, esperada e observada, contrastam-se mediante uma prova de χ^2 com $g - 2$, onde $g=10$ (número de grupos).

O número de grupos, g , pode ser menor de 10 se houver menos de 10 testes padrões de variáveis explanatórias. Deve haver ao menos três grupos para que a estatística de Hosmer-Lemeshow seja calculada.

Sob a hipótese de nulidade de que os dados não se ajustam ao modelo, se o valor p é menor que o ponto de corte aceito de 0,05, o teste mostra evidência de ajuste inadequado, pois não se pode rejeitar a hipótese de nula (EBRAHIM, 2001).

3.2.4. Método bayesiano de estimação dos parâmetros

Os problemas de inferência paramétrica bayesiana se concentram na obtenção de uma certa quantidade a posteriori de interesse, que será função dos parâmetros do problema. Assim, se o que se deseja é a inferência sobre $\underline{\beta} = (\beta_0, \beta_2, \dots, \beta_{p-1})$, ou sobre qualquer transformação do parâmetro, $g(\underline{\beta})$, do ponto

de vista bayesiano, a formulação tem de contemplar a informação amostral, dada pelos dados $Y = (Y_1, Y_2, \dots, Y_n)$ e $X = (X_1, X_2, \dots, X_{p-1})$ com função de verossimilhança $\lambda(Y \setminus X\beta)$, e a escolha da distribuição a priori sobre o parâmetro, $P(\beta)$, que possui as informações subjetivas de quem conhece o que está sendo estudado. A combinação de ambos ingredientes conduz a obtenção, através do Teorema de Bayes, da distribuição a posteriori (ANTELMAN, 1977),

$$P(\beta_0, \dots, \beta_{p-1} \setminus Y, X) = \frac{P(\beta_0, \dots, \beta_{p-1}) \lambda(y \setminus \beta_0 + \beta_1 X_1 + \dots + \beta_{(p-1)} X_{(p-1)})}{\int_{\beta} P(\beta_0, \dots, \beta_{p-1}) \lambda(y \setminus \beta_0 + \beta_1 X_1 + \dots + \beta_{(p-1)} X_{(p-1)}) d\beta}$$

Logo,

$$P(\beta_0, \dots, \beta_{p-1} \setminus Y, X) \propto P(\beta) * \lambda(Y \setminus X\beta) \quad (19)$$

Com a qual se pretende calcular a quantidade a posteriori de interesse, que se pode expressar por:

$$E[g(\beta) \setminus Y, X] = \int_{\beta} g(\beta) P(\beta \setminus Y, X) d\beta \quad (20).$$

O caso é que tanto a obtenção da densidade a posteriori, como o cálculo da quantidade a posteriori, contém, na maioria dos casos, integrais que só podem ser resolvidas por métodos de simulação de Monte Carlo Cadeias de Markov (MCMC), que permitem simular uma cadeia que converge para distribuição a posteriori, de maneira que quantidades amostrais obtidas a partir das realizações da citada cadeia convirjam para quantidade a posteriori de interesse (HASTINGS, 1970 e GILKS, 1996).

A idéia básica do método MCMC é construir uma cadeia de Markov que seja fácil de simular e que tenha como distribuição de equilíbrio a distribuição de interesse. Essas técnicas podem ser aplicadas com sucesso em problemas em que a dimensão do espaço paramétrico seja grande (LEANDRO, 2001).

Neste estudo, considerou-se as seguintes situações para estimar os parâmetros do modelo final encontrado na regressão logística clássica stepwise:

- Verossimilhança

O modelo observacional, ou o modelo probabilístico para os dados, como foi visto anteriormente, é o modelo de Bernoulli com função de probabilidade dada pela expressão (4):

$$f(Y_i/X_i) = [P(X_i)]^{Y_i} [1 - P(X_i)]^{1-Y_i}$$

E função de verossimilhança dada pela expressão (5):

$$\lambda(Y/X\beta) = \prod_{i=1}^n f(Y_i/X_i)$$

Onde $P(X)$ é dado pela expressão (2):

$$P(X) = \frac{e^{\beta_0 + \sum_{i=1}^{p-1} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{p-1} \beta_i X_i}}$$

Para assegurar que $0 < P(X) < 1$, usa-se a transformação logit assim, o modelo é:

$$\log it(p_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$$

Onde k é o número de variáveis do modelo final.

- Distribuições Prioris não informativa para “ β_j ”

Considerou-se que cada parâmetro do modelo se distribui segundo uma normal com média zero e variância 0,001.

$$\beta_j \sim N(0, .001) \forall j \quad (21)$$

- Distribuição Posteriori

Pela expressão de posteriori definida em (19) e substituindo nessa expressão a verossimilhança dos dados vista em (6) e priori considerada em (21) encontra-se a distribuição posterior em comum dos parâmetros, determinada pela expressão seguinte:

$$P(\beta_0, \dots, \beta_{p-1} / Y, X) \propto p(\beta_0, \dots, \beta_{p-1}) \prod_i p_i^Y (1 - p_i)^{1-Y}$$

$$\propto p(\beta_0, \dots, \beta_{p-1}) \prod_i \text{logit}^{-1}(\beta_0 + \beta_1 X_{1i} + \dots + \beta_{p-1} X_{(p-1)i})^Y * \\ \text{logit}^{-1}(1 - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_{p-1} X_{(p-1)i}))^{1-Y}$$

A obtenção da expressão dessas *posteriors* marginais de forma analítica é inviável nesse caso, devido à complexidade dos cálculos envolvidos. Uma alternativa é gerar as distribuições a *posteriori* de cada um dos parâmetros através da técnica denominada *Amostrador de Gibbs*. Em Gamerman(1997); Robert e Casella(1999) e Chen et al.(2000) contem uma revisão destes métodos com aplicações para modelos de regressão logística.

O programa (anexo 3) para realizar o *Amostrador de Gibbs* utilizado neste trabalho, foi executado no pacote estatístico WinBUGS 14 de livre acesso na *internet* <http://www.mrc-bsu.cam.ac.uk/bugs>, criado em 1999, pelos britânicos Spiegelhalter, Thomas e Best, que implementaram estas técnicas, que simulam realizações amostrais através de “burn in”, ou melhor, de “m” iterações.

Por hipótese, a formulação do modelo é um primeiro passo para o uso do programa. Deve-se deixar claramente estabelecidos todos os elementos que intervêm no modelo assim como qual é a quantidade a *posteriori* de interesse;

Com o modelo formulado, se criará o “Desenho do doodle” (ferramenta gráfica característica do WinBUGS para a visualização do modelo). Para isso abre-se uma janela “doodle” em WinBUGS onde se introduzirá todos os “nodes”(parâmetros) do problema. Uma vez criado o “doodle”, abre-se a ferramenta de especificação de modelos. Esta etapa é concluída com a verificação da correção sintática do modelo. Uma vez comprovado que o modelo está sintaticamente correto, pode-se carregar os dados e inicializar os valores.

A simulação se realiza com uma instrução que nos permitirá decidir o número de amostras a realizar, esta inclui a parte “burn in” da amostra. Uma vez fixados os “nodes”, se procederá à parte da simulação que será armazenada, e que intervirá no cálculo das quantidades a *posteriori*. A partir desse momento será possível analisar os resultados.

O WinBUGS oferece diferentes resultados estatísticos sobre os “nodes” de interesse (média, mediana, quartis, intervalos de credibilidade, variância) assim como a medida do erro cometido. Também realiza os gráficos da densidade a *posteriori* de cada “node” (um histograma amostral), o traço da série gerada, e assim como a representação dos coeficientes de autocorrelação. Estes gráficos junto com

a medida do erro são instrumentos que permitem decidir se o tamanho da amostra e do “burn in” têm sido adequados.

Pode-se, assim, verificar a convergência através da visualização minuciosa dos gráficos do traço (LEANDRO, 2001). E finalmente, calculou-se os resumos desejados utilizando estatística descritiva, isto é, cálculos de certas características da distribuição posteriori tais como: média, desvio padrão e intervalos de credibilidades. Visto que, a distribuição posteriori contém toda a informação sobre o parâmetro β (LEANDRO, 2001), conseqüentemente calculou-se as OR e seus intervalos de credibilidades.

3.2.5. Medida de associação

Há que ter em conta que os estimadores habituais da associação não são os coeficientes β_j senão os OR, portanto os intervalos de confiança de interesse para calcular são os OR, já mencionados.

Para um conjunto de n observações, $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, isto é, com apenas uma variável independente, X , a expressão (3) fica

$$\log it \quad P(X) = \ln \left[\frac{P(X)}{1 - P(X)} \right] = \beta_0 + \beta_1 X_1$$

Onde β_0 e β_1 são parâmetros desconhecidos.

Esse modelo foi aplicado para analisar a associação entre a infecção por HIV e a ocorrência ou não de um fator particular. A chance de ser um doador HIV-positivo, com presença do fator ($x=1$) fica dada por:

$$\frac{P(x=1)}{1 - P(x=1)} = e^{\beta_0 + \beta_1}$$

Enquanto que a chance de ser um doador HIV-positivo, com ausência do fator é simplesmente.

$$\frac{P(x=0)}{1 - P(x=0)} = e^{\beta_0}$$

Logo, o “Odds Ratio” (razão de chances) fica dada por:

$$OR = \left[\frac{\frac{P(x=1)}{1-P(x=1)}}{\frac{P(x=0)}{1-P(x=0)}} \right] = \frac{P(x=1)[1-P(x=0)]}{P(x=0)[1-P(x=1)]} = e^{\beta_1} \quad (22)$$

Essa é uma grande vantagem da regressão logística, a possibilidade de interpretação direta dos coeficientes como medidas de associação.

Um intervalo assintótico de confiança para OR com coeficiente $(1-\alpha)$, terá os limites:

$$(\hat{OR}_i, \hat{OR}_s) = e^{\left\{ \hat{\beta} \pm Z_{1-\frac{\alpha}{2}} \hat{SE}(\hat{\beta}) \right\}} \quad (23)$$

As estimativas dos parâmetros do modelo são feitas sob a suposição de que as variáveis têm uma escala de medida.

Segundo Martins(1988); Hosmer e Lemeshow (1989), um problema que surge frequentemente com as variáveis explicativas em um modelo de regressão logística é a forma de interpretação dos parâmetros. Variáveis discretas, tais como, sexo, tipo de doação e marcador sorológico, não têm escala ordenada de medida. Desta forma, se codificarmos de forma arbitrária, a variável “tipo de doação” como: 1, 2 e 3 respectivamente as suas categorias, o parâmetro estimado não tem interpretação, pois dizer que o $\log [P_i / (1-P_i)]$ é acrescido de uma unidade se passarmos de um doador do tipo reposição para um do tipo voluntária, não faz nenhum sentido.

Variáveis indicadoras (apêndice), tomando os valores um ou zero para designar a presença ou ausência do atributo, devem ser usadas para representar corretamente os efeitos de tais variáveis em um modelo de regressão logística. Desta forma, a variável “tipo de doação”, tendo três categorias, necessita usar dois variáveis indicadoras. Primeiramente, designa-se uma das categorias como referência e para cada uma das demais, introduz-se uma variável R_i , que é codificada como um (presença) ou zero (ausência). No estudo tomou-se o tipo de doação “autóloga ou específica”(categoria 1, vê quadro 1) como grupo de referência, introduziu-se duas variáveis indicadoras para o tipo de doação: R_1 , codificada como

um (voluntária) ou zero (não voluntária); R_2 , codificada como um (reposição) ou zero (não reposição). A variável X_6 apresentada no quadro 1, é, então, trocada por duas variáveis indicadoras R_1 e R_2 e a saída dos dados para análise da regressão deve ser recodificada para refletir esta mudança. Em geral, se uma variável discreta tem K categorias, será necessário usar k-1 variáveis indicadoras e o coeficiente β_i associado com cada uma das variáveis indicadoras representa a mudança no log [$P_i / (1-P_i)$] para esta categoria relativa à categoria de referência.

Convém destacar que estas variáveis indicadoras não têm nenhum sentido por si só, e, portanto, devem figurar nos modelos e se deve contrastar sua inclusão sempre em bloco.

Para as variáveis Idade (X_2) e Escolaridade (X_3), procedeu-se da mesma forma, sendo criado três variáveis indicadoras (R_1, R_2 e R_3) para cada uma dessas duas variáveis. Onde o grupo de referência para X_2 foi idade “maior ou igual a 50 anos” (categoria 4, vê quadro 1) e para X_3 foi escolaridade “superior” (categoria 4, vê quadro 1).

3.2.6. Análise estatística

Inicialmente foi feito o teste de independência χ^2 - Pearson (com correção de Yates para tabela 2x2) para cada variável explicativa em relação a variável dependente, a infecção por HIV, com a finalidade de verificar a existência de associação em nível inferior a 20% ($p < 0,20$), no SPSS.

Uma vez definido, pela etapa inicial, o conjunto de covariáveis a ser incluído no modelo logístico múltiplo, restou saber qual a melhor maneira de encontrar um modelo reduzido que incluísse apenas as covariáveis mais importantes para explicar a probabilidade de ser doador HIV-positivo $P(Y = 1/\underline{X})$.

A investigação dos fatores associados à infecção por HIV foi feita utilizando-se a regressão logística múltipla clássica, no SPSS. Baseada num dos métodos mais aplicados em regressão logística, o método stepwise (HOSMER e LEMESHOW 1989), que baseia-se num algoritmo misto de inclusão e eliminação de covariáveis segundo a importância das mesmas de acordo com algum critério estatístico. Esse grau de importância foi avaliado pelo nível de significância do teste da razão de verossimilhança entre os modelos que incluem ou excluem as covariáveis em

questão. Quanto menor fosse esse nível de significância tanto mais importante foi considerada a covariável (foram consideradas as variáveis associadas à infecção por HIV que apresentaram o valor de “p” para o teste da razão de verossimilhança igual ou inferior a 0,05).

Avaliou-se também a significância de cada um dos coeficientes das covariáveis selecionadas no processo anterior, através do teste de Wald (1943), considerando o nível de significância de 5%, no SPSS. Se alguma covariável ou fator não fosse significativo deveria ser excluído do modelo.

Uma vez selecionadas as covariáveis “mais importantes”, ou os efeitos principais, e testado-se interações, deu-se entrada na etapa da análise dos resíduos desse modelo final. A análise dos resíduos foi realizada através do teste de Hosmer-Lemeshow, para um nível de significância de 5%, no SPSS.

Com a finalidade de maximizar o número de doadores de sangue (produto) e ao mesmo tempo minimizar o risco de infecção (insumo) por HIV, foi realizada a análise do ponto ótimo da função eficiência, definida como a razão entre o número de doadores e o risco. O estudo foi feito com o Logit distribuído em intervalos de Classes, calculando-se, para cada classe i :

$$Eficiência = \frac{Pr\ oduto}{Insumo}$$

Onde,

$$insumo = \frac{N^{\circ} \text{acumulado de HIV Positivo na Classe } i}{N^{\circ} \text{acumulado de Doadores na Classe } i}$$

e

$$Pr\ oduto = N^{\circ} \text{acumulado de Doadores na Classe } i$$

Depois de verificado o bom ajuste do modelo final e feita à análise da função eficiência, passou-se para a etapa de estimar os parâmetros desse modelo ajustado pelo método bayesiano, utilizando o programa (anexo 3) no Winbugs 14.

E com esses resultados executar a comparação entre os dois métodos: bayesiano e clássico. Utilizando-se como medida de avaliação os “Odds Ratio estimadas e seus respectivos intervalos de 95% de confiança.

4. RESULTADOS E DISCUSSÃO

Foram estudados 106.203 doadores de sangue de 1ª vez, aptos na triagem clínica. Dessa amostra, conforme a tabela 1, houve uma predominância acentuada de doadores do gênero masculino, 84.073 (79%); com idade entre 18 a 39 anos, 81.368 (77%); com ensino fundamental, 80.238(76%); residente na RMR 92.425 (87%); de tipo de doação reposição, 85.146 (80%), com status sorológico para anti-Hbc negativo, 91.796(86%); com HbsAg negativo, 104.989 (99%); com anti-HCV negativo, 105.316 (99%); com anti-HTLV negativo, 105.431 (99%); com Sífilis negativo, 102.733 (97%). Por outro lado quanto a variável situação conjugal houve uma pequena diferença (4%), ou seja, 55.508 (52%) solteiro/divorciado/outros e 50.695 (48%) casado/viúvos.

Dos 106.203 doadores de 1ª vez aptos na triagem clínica, observados neste estudo, 217 (0,204%) foram diagnosticados como HIV-positivo, que corresponde a 2,04 por mil doações, verificando-se uma pequena queda da proporção, quando comparada com outros dados, nos quais na região nordeste essa proporção chegou em 1997 a 2,650 por mil doações (BRASIL, 2003; KUPEK, 2000). Tais resultados não eram de se esperar, uma vez que a doação é voluntária e entre doadores de sangue, supunha-se não existirem indivíduos doentes, considerando-se o fato de que a doação deve ser sempre um ato altruísta e, além disso, os candidatos serem submetidos à rigorosa anamnese (anexo 1). Segundo Rouquayrol (1999) candidatos à doação de sangue ainda procuram centros de doação de sangue para realizarem testes sorológicos na tentativa de obterem diagnóstico para DST, pondo o sistema, conseqüentemente, em risco.

Além disso, um outro risco a ser considerado é o da janela imunológica (COVAS, 1998; CANUTTI JÚNIOR, 1998), isto é, período no qual os teste não detectam a soropositividade, em virtude da pequena quantidade de anticorpos no sangue examinado, mas com possibilidade dos HIV-negativos, nessa condição de transmitirem o vírus. Para eliminar os que estejam no período de janela imunológica utilizam-se os critérios rigorosos da triagem clínica, valorizando-se os fatores de risco identificados na presente pesquisa, uma vez que ainda não são adotados na rotina testes sorológicos mais sensíveis.

A tabela 1 apresenta a distribuição quanto à associação entre a infecção por HIV e cada variável independente. Os resultados referentes ao teste qui-quadrado de independência de Pearson, indicaram que as variáveis Gênero ($p=0,261$) e os marcadores sorológicos: Anti-HCV ($p=0,608$) e Anti-HTLV ($p=1,0$) não tiveram contribuição significativa estatisticamente ($p<0,20$) com a infecção por HIV, e não entraram na regressão logística múltipla. Esses resultados estão de acordo com os estudos citados na revisão de literatura.

Tabela 1. Análise univariada da infecção por HIV segundo gênero, idade, escolaridade, situação conjugal, região de residência, tipo de doação, status sorológico para: anti-Hbc, HbsAg, anti-HCV, anti-HTLV e sífilis.

Variável independente		Infecção por HIV			Análise estatística		
nome	categoria	Positiva	%	Negativa %		Total	
Gênero	Feminino	38	0,17	22.092	99,83	$\chi_1^2=1,263$ p=0,261	
	Masculino	179	0,21	83.894	99,79		84.073
Idade	18 a 28	89	0,20	44.720	99,80	$\chi_3^2=7,446^a$ $p=0,059$	
	29 a 39	90	0,25	36.469	99,75		36.559
	40 a 49	31	0,17	18.141	99,83		18.172
	Maior ou igual a 50	7	0,11	6.656	99,89		6.663
Escolaridade	Analfabeto	5	0,41	1.222	99,59	$\chi_3^2=9,451$ $p=0,024$	
	Ensino fundamental	170	0,21	80.068	99,79		80.238
	Ensino médio	38	0,21	18.083	99,79		18.121
	Ensino superior	4	0,06	6.613	99,94		6.617
Situação conjugal	Casado/ viúvo	87	0,17	50.608	99,83	$\chi_1^2=4,788$ $p=0,029$	
	Solteiro/ divorciado/outros	130	0,23	55.378	99,77		55.508
Região de residência	RMR	204	0,22	92.221	99,78	$\chi_1^2=8,781$ $p=0,003$	
	Interior/ outros estados	13	0,09	13.765	99,91		13.778
Tipo de doação	Autóloga/ específica	1	0,03	3.314	99,97	$\chi_2^2=69,031$ $p=0,000$	
	Voluntária	81	0,46	17.661	99,54		17.742
	Reposição	135	0,16	85.011	99,84		85.146
Anti – Hbc	Negativo	164	0,18	91.632	99,82	$\chi_1^2=20,947$ $p=0,000$	
	Positivo	53	0,37	14.354	99,63		14.407
HbsAg	Negativo	210	0,20	104.779	99,80	$\chi_1^2=6,602$ $p=0,010$	
	Positivo	7	0,58	1.207	99,42		1.214
Anti – HCV	Negativo	214	0,20	105.102	99,80	$\chi_1^2=0,264$ p=0,608	
	Positivo	3	0,34	884	99,66		887
Anti – HTLV	Negativo	215	0,20	105.216	99,80	$\chi_1^2=0,000$ p=1,0	
	Positivo	2	0,26	770	99,74		772
Sífilis	Negativo	193	0,19	102.540	99,81	$\chi_1^2=39,344$ $p=0,000$	
	Positivo	24	0,69	3.446	99,31		3.470
Total		217	0,20	105.986	99,80	106.203	

4.1. MÉTODO CLÁSSICO

A seleção das variáveis independentes mais importantes foi analisada através do SPSS versão 10, via dois métodos stepwise Backward e Forward (razão de

verossimilhança), pelos quais se obtiveram os mesmos resultados, utilizando-se apenas as oito variáveis explicativas que foram associadas com a infecção HIV pelo teste χ^2 , tais como idade ($p=0,059$), escolaridade ($p=0,024$), situação conjugal ($p=0,029$), região de residência ($p= 0,003$), tipo de doação ($p=0,000$), anti-HBC ($p=0,000$), HbsAg ($p=0,010$) e sífilis ($p=0,000$) (Tabela 1).

As variáveis idade, escolaridade e tipo de doação foram analisadas a partir da construção de variáveis dummy (vê explicação no apêndice). Segundo a tabela 2, observa-se que o grupo de referência para a variável independente escolaridade é “ensino superior”, para idade é “maior ou igual a 50 anos” e para tipo de doação “autóloga/específica”.

Tabela 2. Codificação Dummy das variáveis categóricas: escolaridade, idade e tipo de doação, do modelo final no SPSS.

Variável Independente		frequência	Dummy		
Nome	Categorização		D(1)	D(2)	D(3)
Escolaridade	analfabeto	1.227	1	0	0
	ensino fundamental	80.238	0	1	0
	ensino médio	18.121	0	0	1
	ensino superior	6.617	0	0	0
Idade	18 a 28	44.809	1	0	0
	29 a 39	36.559	0	1	0
	40 a 49	18.172	0	0	1
	maior ou igual a 50	6.663	0	0	0
Tipo de doação	Autóloga/Específica	3.315	0	0	
	Voluntária	17.742	1	0	
	Reposição	85.146	0	1	

Utilizando-se o método backward com as oito variáveis relacionadas significativamente com a infecção por HIV, através do teste qui-quadrado, observou-se que o mesmo eliminou no segundo passo a variável HbsAg ($p=0,103$) para o teste da razão de verossimilhança e em seguida a variável situação conjugal pois a mesma não foi estatisticamente significativa, pelo teste de Wald ($W=2,752$; $p=0,097$) ao nível de significância de 5%.

Por outro lado foi também testado o modelo, considerando-se o método backward para o modelo completo sem realização do teste qui-quadrado, observou-se que o mesmo eliminou no segundo passo a variável Anti-HTLV ($p=0,905$); no terceiro eliminou a variável Anti-HCV ($p=0,477$); no quarto eliminou a variável gênero ($p=0,173$) e no quinto eliminou a variável HbsAg ($p=0,103$), de acordo com o teste

da razão de verossimilhança, ao nível de significância de 5%. Como também a variável situação conjugal não foi significativa para o teste de Wald ($W=2,752$; $p=0,097$). Concluindo-se o mesmo resultado do procedimento anteriormente descrito.

Logo, a seleção de variáveis independentes relacionadas à infecção por HIV e de seus respectivos coeficientes pelo modelo de regressão logística, permitiu a elaboração de uma fórmula matemática cujo resultado indica o logit da probabilidade de ser doador soropositivo para HIV, tal como:

$$\begin{aligned} \text{logit}(HIV_i) = & \beta_0 + \beta_1(\text{idade } 18 \text{ a } 28) + \beta_2(\text{idade } 29 \text{ a } 39) + \beta_3(\text{idade } 40 \text{ a } 49) + \beta_4(\text{analfabeto}) + \\ & + \beta_5(\text{Ens. fundamental}) + \beta_6(\text{Ens. médio}) + \beta_7(\text{Anti - Hbc}) + \beta_8(\text{Sífilis}) + \\ & + \beta_9(\text{Re g. residência}) + \beta_{10}(\text{Doação voluntária}) + \beta_{11}(\text{Doação reposição}) \end{aligned} \quad (24)$$

Substituindo-se os parâmetros pelas estimativas encontradas pelo método clássico de verossimilhança, temos:

$$\begin{aligned} \text{logit}(HIV_i) = & -10,719 + 0,872(\text{idade } 18 \text{ a } 28) + 1,009(\text{idade } 29 \text{ a } 39) + 0,561(\text{idade } 40 \text{ a } 49) + 2,071(\text{analfabeto}) + \\ & + 1,083(\text{Ens. fundamental}) + 1,089(\text{Ens. médio}) + 0,677(\text{Anti - Hbc}) + 1,173(\text{Sífilis}) + \\ & + 0,804(\text{Re g. residência}) + 2,468(\text{Doação voluntária}) + 1,455(\text{Doação reposição}) \end{aligned}$$

Tabela 3. Tabela de contingência para o teste de Hosmer-Lemeshow no SPSS.

Decil	Infecção por HIV = negativa		Infecção por HIV = positiva		Total
	observada	esperada	observada	esperada	
1	10.050	10.049,942	4	4,058	10.054
2	10.579	10.577,688	6	7,312	10.585
3	10.782	10.782,001	12	11,999	10.794
4	19.720	19.723,605	33	29,395	19.753
5	7.388	7.383,850	7	11,150	7.395
6	16.281	16.282,078	29	27,922	16.310
7	10.464	10.461,307	20	22,693	10.484
8	9.991	10.000,654	46	36,346	10.037
9	10.731	10.724,871	60	66,129	10.791

No nível de significância de 5%, existe evidência suficiente para rejeitar a afirmação de que nenhum dos fatores de risco do modelo estão relacionados com a

variável resposta, segundo o resultado do teste da razão de verossimilhança ($\chi^2_{11} = 126,579$, $p = 0,000$), para o modelo completo. Podemos dizer que existe evidência estatística de que no mínimo um dos fatores de risco está associado à ocorrência da infecção por HIV, em doadores de sangue.

O valor da estatística do teste de Hosmer-Lemeshow calculada pelas frequências da tabela 3 é $\hat{C} = 5,733$, e o p-valor correspondente calculado pela distribuição qui-quadrado com 7 graus de liberdade é 0,571. Isso indica que o modelo parece ajustar bastante bem. Uma comparação das frequências observadas e esperadas, em cada das 18 células da tabela 3, mostrou o modelo ajustado dentro de cada decil de risco. Pode-se aceitar que o ajuste do modelo foi satisfatório.

Após verificado o bom ajuste do modelo resultante do processo stepwise backward, procedeu-se a análise de associações das variáveis independentes com a infecção por HIV, observando-se os resultados da tabela 4.

Para a importância de cada variável no modelo foi verificada a estatística do teste de Wald (Tabela 4), onde se observou que apenas as variáveis “idade 40 a 49 anos” (Wald=1,792; $p = 0,181$) e tipo de doação “reposição” (Wald=2,106; $p = 0,147$) não foram associadas à infecção por HIV para o teste de Wald no nível de significância de 5%. Portanto, verificou-se que todas as outras variáveis do modelo final, tais como: idade 18 a 28 anos (Wald=4,824; $p = 0,028$), idade 29 a 38 anos (Wald=6,538; $p = 0,011$), escolaridade analfabeto (Wald=9,325, $p = 0,002$), escolaridade ensino fundamental (Wald=4,551; $p = 0,033$), escolaridade ensino médio (Wald=4,281; $p = 0,039$), anti-Hbc (Wald=17,313; $p = 0,000$), sífilis (Wald=28,501, $p = 0,000$), região de residência (Wald=7,739, $p = 0,005$) e tipo de doação voluntária (Wald=6,032; $p = 0,014$), tiveram valores de “p” menores que 0,05, indicando que existe evidência suficiente de que os coeficientes de $\hat{\beta}_i$ são diferentes de zero para um nível de significância de 5%, baseados no teste de Wald aplicado individualmente. Tais variáveis foram estatisticamente significantes como fatores de risco em potenciais à infecção por HIV.

Procedendo-se agora com a análise da força da associação de cada potencial fator de risco, observou-se (Tabela 4), com relação a variável idade, mantendo-se todas as outras variáveis independentes, constantes, e tomando-se como grupo de referência a categoria idade maior ou igual a 50 anos, que a chance de ser HIV-positivo para os doadores com idade entre 18 a 28 anos foi de 2,392 vezes maior

que a idade de referência com (IC95%: 1,098 - 5,209). Para os doadores com idade entre 29 a 39 anos foi de 2,742 vezes maior que a referência, com (IC95%: 1,266 - 5,940). E para os com idade entre 40 a 49 anos, a chance decresceu para 1,753 vezes maior que a idade de referência, com (IC95%: 0,771 - 3,988). Observando-se ainda que a associação com a infecção por HIV foi estatisticamente significativa apenas nos dois primeiros grupos etários acima para o teste de Wald, $p=0,028$ e $p=0,011$, respectivamente, e para o último grupo a relação não foi estatisticamente significativa ($p=0,181$), indicando uma tendência da faixa etária acima de 40 anos de apresentar menor risco para o receptor. Conclui-se que fatores de risco em potencial para a infecção HIV, analisando-se a variável idade e mantendo-se todas as outras variáveis explicativas constantes, foram relacionados apenas com os grupos etários de 18 a 28 e 29 a 39 anos com maior risco entre os de 29 a 39 (OR=2,742). Esses resultados estão de acordo com os estudos de Sampaio (2001); Zachariah *et. al.* (2002); Rouquayrol (1999) e Boletim Informativo DST/AIDS (2002b).

No que se refere a variável escolaridade, verificou-se (Tabela 4) que a chance de ser doador HIV-positivo cresceu à medida que a escolaridade foi diminuindo, mantendo-se constantes todas as outras variáveis do modelo. Para doadores com tipo de escolaridade analfabeto a chance de ser doador HIV-positivo foi de 7,935 vezes maior que a de doadores com ensino superior (grupo de referência), com (IC95%: 2,100 - 29,987). Para doadores com ensino fundamental a chance foi de 2,955 vezes maior que o grupo de referência, com (IC95%: 1,092 - 7,996). E para doadores com ensino médio a chance foi de 2,971 vezes maior, com (IC95%: 1,059 - 8,335). Todas as categorias foram associadas com a infecção por HIV para o teste de Wald, analfabeto ($p=0,002$), ensino fundamental ($p=0,033$) e ensino médio ($p=0,039$). Considerando-se a categoria analfabeta a de maior risco de ser doador HIV-positivo. Isso confirma os resultados de Souza (1998), e aqueles apresentados pela Secretaria de Saúde de Pernambuco sendo que em vista disso a Secretaria aponta a tendência atual da epidemia para a “pauperização” da AIDS (BOLETIM INFORMATIVO DST/AIDS, 2003).

Conforme pode ser observado na tabela 4, a análise do parâmetro positivo $\hat{\beta}_9 = 0,804$ para a variável “Região de residência” representa a estimada alteração na função logit, quando doadores que moram na RMR são comparados com os que não moram, mantendo as outras variáveis constantes. Teve-se um $\hat{OR} = 2,234$ com

(IC95%: 1,268 - 3,937) para a variável “região de residência”, indicando que doadores da RMR apresentaram uma chance 2,234 vezes maior de ser HIV-positivo, do que os que não moram na RMR.

Nenhuma associação foi observada (Tabela 4) entre a doação do tipo reposição e a infecção por HIV (teste de Wald $p=0,147$), sendo sua chance de 4,283 vezes maior que a de doadores com doação autóloga/específica, com (IC95%: 0,601 - 30,540). Já para os doadores com doação do tipo voluntária, a chance de ser doador HIV-positivo foi 11,799 vezes maior que a de doadores com doação autóloga/específica, com (IC95%: 1,646 - 84,568), e observou-se associação com a infecção por HIV estatisticamente significativa (Wald=6,032; $p=0,014$). Embora o doador voluntário seja considerado o doador ideal, porque na teoria o mesmo estaria praticando um ato de altruísmo, e que por isso seria mais isento de soropositividade, não foi o que se encontrou na presente pesquisa. Isso pode indicar, entre outros aspectos, que os voluntários HIV-positivos da amostra, optaram por doar sangue em busca de diagnóstico. Isso devido ao fato de considerarem, por desconhecimento, que os resultados de testes obtidos em hemocentros e banco de sangue sejam mais seguros em comparação aos exames realizados em laboratórios de análises clínicas, os quais são especificamente realizados visando o diagnóstico da doença.

A análise do OR igual a 1,968 com (IC95%: 1,430 - 2,706) para a variável status sorológico ANTI – Hbc (Tabela 4) indica que doadores com ANTI-Hbc – positivo, com mesmo valor para as outras variáveis do modelo, apresentaram uma chance de ser HIV – positivo 1,968 vezes maior que os doadores com resultado anti- Hbc – negativo. Isso demonstrou que o resultado anti-Hbc-positivo representa um fator de risco em potencial para a infecção por HIV. Tais resultados estão de acordo com aqueles encontrados por REICHE *et. al.* (2003) em doadores de sangue. Portanto, a associação entre as duas infecções, HIV e hepatite B, pode ocorrer devido ao fato dessas viroses apresentarem a mesma forma de transmissão, ou pode ser devido ao alto percentual de reações falso-positivas do anti-Hbc.

A análise do OR igual a 3,232 com (IC95%: 2,101 - 4,972) para a variável “status sorológico para sífilis” (Tabela 4) indica que doadores com sífilis-positivo, com mesmo valor para as outras variáveis do modelo, apresentaram uma chance de ser HIV-positivo 3,232 vezes maior que aqueles com sífilis-negativo. Esse resultado diverge daquele encontrado por Reiche *et. al.* (2003) cuja conclusão foi de que o teste para sífilis é considerado de pouca importância como marcador indireto para

identificar e excluir doações de pessoas com alto risco de infecção por HIV. Tal resultado foi encontrado também por Herrera et. al.(1997) apud REICHE (2003), nos E.U.A.. Enquanto Sáez-Alquézar et. al. (1997) apud Reiche, 2003, encontrou que o teste de sífilis na triagem pode indiretamente prevenir o risco residual de transmissão de HIV. No Rio Grande do Sul, Barcellos et. al.(2003) também mostraram relação de sífilis com HIV-positivo em um centro de testagem anônimo. Na Etiópia, Rahlenbeck (1997), estudando doadores de sangue, igualmente, relatou que os HIV-positivos tiveram um risco aumentado para sífilis-positivo. Ele relacionou seus resultados ao tipo de transmissão, DST, das duas infecções.

Tabela 4. Estimativas dos parâmetros β , seu desvio padrão(SE), teste de Wald,, OR e Intervalo de confiança(IC) para OR, referentes ao modelo (24), no SPSS.

Variável Independente	β	SE(β)	Teste de Wald			OR=Exp(β)	IC95% para OR	
			Valor	gl	p-valor		inferior	superior
Idade								
Idade 18 a 28	$\beta_1 = 0,872$	0,397	4,824	1	0,028	2,392	1,098	5,209
Idade 29 a 38	$\beta_2 = 1,009$	0,394	6,538	1	0,011	2,742	1,266	5,940
Idade 39 a 49	$\beta_3 = 0,561$	0,419	1,792	1	0,181	1,753	0,771	3,988
≥ 50						1		
Escolaridade								
analfabeto	$\beta_4 = 2,071$	0,678	9,325	1	0,002	7,935	2,100	29,987
Ens. fundamental	$\beta_5 = 1,083$	0,508	4,551	1	0,033	2,955	1,092	7,996
Ens. médio	$\beta_6 = 1,089$	0,526	4,281	1	0,039	2,971	1,059	8,335
Ens. superior						1		
Reg. residência								
RMR	$\beta_9 = 0,804$	0,289	7,739	1	0,005	2,234	1,268	3,937
Interior/ out. Est.						1		
Tipo de doação								
Autóloga/específica						1		
voluntária	$\beta_{10} = 2,468$	1,005	6,032	1	0,014	11,799	1,646	84,568
reposição	$\beta_{11} = 1,455$	1,002	2,106	1	0,147	4,283	0,601	30,540
Ant-Hbc								
positivo	$\beta_7 = 0,677$	0,163	17,313	1	0,000	1,968	1,430	2,706
negativo						1		
Sífilis								
positivo	$\beta_8 = 1,173$	0,220	28,501	1	0,000	3,232	2,101	4,972
negativo						1		
Constante	$\beta_0 = -10,719$	1,200	79,849	1	0,000	0,000		

IC-95% = intervalo de 95% de confiança do OR

Não se observou interação estatisticamente significativa ou biologicamente importante.

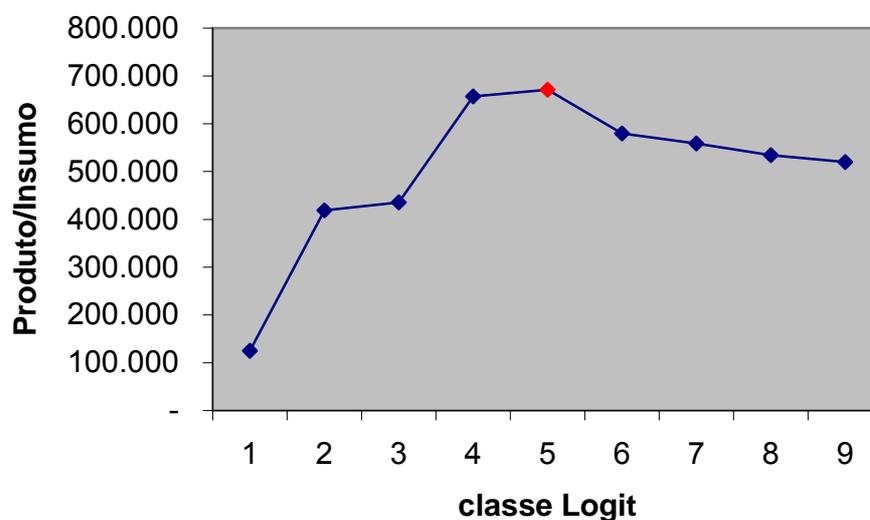
Observou-se que o ponto de máximo da função eficiência, definida como razão entre o número de pessoas e o risco, pertenceu a classe 5 (Figura 8). Esse

resultado indicou que se existisse uma pré-triagem antes da triagem sorológica, baseada num algoritmo que calcule o logit para cada doador, rejeitando aquele que estivesse entre as classes 6 a 9, observar-se-ia, para a amostra estudada, que se rejeitaria um total de 23.455 (22,09% do total pesquisado) doadores, dos quais 115 (53,00% do total infectado) eram doadores positivos para a infecção HIV(Tabela 5).

Tabela 5. Distribuição da Função Eficiência por classe do Logit.

Classe i	Logit HIV	Infectados	Doadores	Infectados acumulado	Insumo	Produto	Eficiência
1	-10,719 a -7,776	2	5.000	2	0,04000	5.000	125.000,00
2	-7,715 a -7,219	5	12.124	7	0,04088	17.124	418.901,97
3	-7,193 a -6,700	15	13.827	22	0,07108	30.951	435.438,36
4	-6,694 a -6,362	73	48.069	95	0,12022	79.020	657.280,04
5	-6,358 a -5,898	7	3.728	102	0,12327	82.748	671.297,21
6	-5,828 a -5,482	62	14.763	164	0,16819	97.511	579.780,19
7	-5,476 a -5,123	28	6.056	192	0,18539	103.567	558.652,27
8	-5,120 a -4,494	17	2.086	209	0,19782	105.653	534.093,61
9	-4,471 a -3,194	8	550	217	0,20433	106.203	519.773,14
Total		217	106.203	-	-	-	-

Figura 8. Ponto que Maximiza o nº de doador e Minimiza o risco de Infecção por HIV.

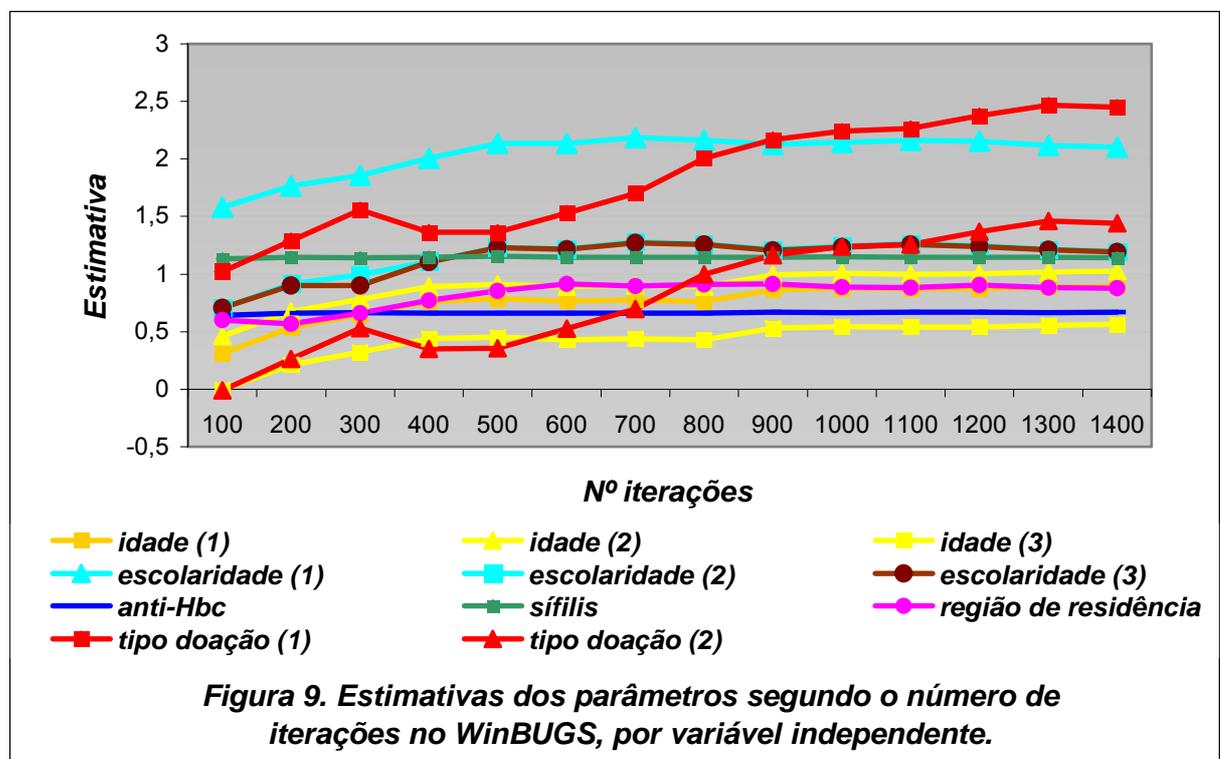


4.2. MÉTODO BAYESIANO

Foi utilizado na análise bayesiana o modelo final resultante do stepwise backward da regressão logística clássica:

$$\log it(HIV_i) = \beta_0 + \beta_1(idade18a28) + \beta_2(idade29a39) + \beta_3(idade40a49) + \beta_4(analfabeto) + \\ + \beta_5(Ens.fundamental) + \beta_6(Ens.médio) + \beta_7(Anti-Hbc) + \beta_8(Sífilis) + \\ + \beta_9(Re g.residência) + \beta_{10}(Doação voluntária) + \beta_{11}(Doação reposição)$$

Nesse método a dificuldade foi definir o número de iterações para equilíbrio das estimativas dos parâmetros, principalmente, para os relacionados com as variáveis doação voluntária(β_{10}) e doação reposição (β_{11}), que exigiram um número maior de iterações para a distribuição posteriori desses parâmetros (Figura 9). Depois de minucioso estudo de 100 a 12.000 iterações, foi verificado que 1.400 seria o número mínimo de iterações necessárias para realizar o estudo descritivo da distribuição posteriori dos referidos parâmetros de interesse.



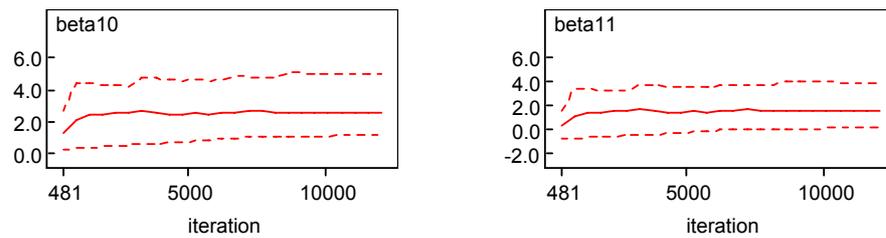


Figura 10. gráficos dos quantis da cadeia para os parâmetros β_{10} e β_{11} com 12.000 iterações.

Além da figura 9, pode-se verificar nos gráficos dos quantis da distribuição posteriori (Figura 10), para os parâmetros das variáveis doação voluntária (β_{10}) e doação reposição (β_{11}), para 12.000 iterações, que depois de 481 e antes de 2500 iterações já se observa o equilíbrio da cadeia. Pelo fato dos outros parâmetros convergirem mais rápido, o que é observado na figura 9, não foi necessário um estudo minucioso para os mesmos. Detendo-se toda a análise nesse dois parâmetros da figura 10. Ressaltando-se que, para cada 100 iterações no winbugs, o tempo de processamento foi de 10 minutos, num computador de processador AMD Athlon 1.1GHz com 128MB de memória.

Tabela 6. Resultados das estimativas dos parâmetros, seu desvio padrão e IC 95%, para o modelo final múltiplo, realizando-se 1.400 iterações no Winbugs.

Variável independente	β	SE(β)	IC95% para β	
			inferior	superior
Idade				
Idade 18 a 28	$\beta_1 = 0,896$	0,420	0,096	1,925
Idade 29 a 39	$\beta_2 = 1,027$	0,418	0,207	2,004
Idade 40 a 49	$\beta_3 = 0,565$	0,441	-0,280	1,577
≥50 anos				
Escolaridade				
analfabeto	$\beta_4 = 2,101$	0,642	0,813	3,323
Fundamental	$\beta_5 = 1,198$	0,448	0,373	2,056
Médio	$\beta_6 = 1,191$	0,465	0,279	2,099
Superior				
Anti-Hbc				
positivo	$\beta_7 = 0,670$	0,160	0,354	0,969
negativo				
Sífilis				
positivo	$\beta_8 = 1,143$	0,226	0,675	1,575
negativo				
Reg. residência				
RMR	$\beta_9 = 0,879$	0,349	0,199	1,603
Interior\ outros estado				
Tipo de doação				
Vonluntária	$\beta_{10} = 2,480$	1,089	0,406	4,411
Reposição	$\beta_{11} = 1,440$	1,085	-0,628	3,374
Autóloga\ específica				
Constante	$\beta_0 = -10,920$	1,236	-12,790	-7,502

Com a realização de 1.400 iterações foram efetuadas as estimativas das médias, desvio padrão e IC95% para os parâmetros do modelo (Tabela 6), esses resultados foram obtidos diretamente do programa no winbugs (anexo3) e em seguida calculados as OR, seus respectivos IC95% e amplitude dos IC (Tabela 7).

Observou-se pelo gráfico do traço da série (Figura 11), em todos os parâmetros do modelo, que a partir de 1350 iterações a série se estabilizou, confirmando-se o número de 1400 iterações. Como é visto também pelo gráfico dos quantis posteriori (Figura 13). Observou-se ainda, pelos gráficos da figura 12, referentes à densidade posteriori dos parâmetros, que as distribuições se aproximam na sua totalidade da distribuição normal.

Tabela 7. Comparação das estimativas das OR, seus IC 95% e amplitude do IC, para o modelo final múltiplo, entre os dois métodos de estimação.

Variáveis Independentes	IC95,0% para Exp(OR)							
	Bayesiana				Clássica			
	OR	Inferior	Superior	Amplitude	OR	Inferior	Superior	Amplitude
<i>Idade 18 a 28</i>	2,450	1,101	6,855	5,754	2,392	1,098	5,209	4,111
<i>Idade 29 a 39</i>	2,793	1,229	7,419	6,190	2,742	1,266	5,940	4,674
<i>Idade 40 a 49</i>	1,759	0,756	4,840	4,084	1,753	0,771	3,988	3,217
<i>Analfabeto</i>	8,174	2,254	27,743	25,489	7,935	2,100	29,987	27,887
<i>Ens. Fundamental</i>	3,313	1,452	7,815	6,363	2,955	1,092	7,996	6,904
<i>Ens. Médio</i>	3,290	1,321	8,158	6,837	2,971	1,059	8,335	7,276
<i>Anti-Hbc</i>	1,954	1,424	2,635	1,211	1,968	1,430	2,706	1,276
<i>Sífilis</i>	3,136	1,964	4,831	2,867	3,232	2,101	4,972	2,871
<i>Reg. Residência</i>	2,408	1,220	4,968	3,748	2,234	1,268	3,937	2,669
<i>Voluntária</i>	11,941	1,502	82,352	80,850	11,799	1,646	84,568	82,922
<i>Reposição</i>	4,221	0,533	29,195	28,662	4,283	0,601	30,540	29,939

Quanto à análise da força da associação pelo método bayesiano, verificou-se que a chance de ser doador HIV-positivo para os doadores com idade entre 18 a 28 anos foi 2,450 vezes maior do que o grupo de referência, com (IC95%: 1,101 -

6,855). Para os doadores com idade entre 29 a 39 anos foi de 2,793 vezes maior que a referência, com (IC95%: 1,229 - 7,419). E para os com idade entre 40 a 49 anos decresceu para 1,759 com (IC95%: 0,756 - 4,840), observando-se também que esta última categoria não representa fator de risco, pois o valor “1” pertence ao IC95%.

No que se refere a variável escolaridade, verificou-se (Tabela 7) que a chance de ser doador HIV-positivo cresceu à medida que a escolaridade foi diminuindo, mantendo-se constantes todas as outras variáveis do modelo. Para doadores com tipo de escolaridade analfabeto a chance de ser doador HIV-positivo foi de 8,174 vezes maior que a de doadores com ensino superior, com (IC95%: 2,254 - 27,743). Para doadores com ensino fundamental a chance foi de 3,313 vezes maior que o grupo de referência, com (IC95%: 1,452 - 7,815). E para doadores com ensino médio a chance foi de 3,290 vezes maior, com (IC95%: 1,321 - 8,158).

Teve-se uma $\hat{OR} = 2,408$ com (IC95%: 1,220 - 4,968) para a variável “região de residência”, indicando que doadores da RMR apresentaram uma chance 2,408 vezes maior de ser HIV-positivo, do que os que não moram na RMR.

Nenhuma associação foi observada (Tabela 7) entre o tipo de doação reposição e a infecção por HIV pelo fato do IC95% conter o “1”, sendo sua chance de 4,221 vezes maior que a de doadores com doação autóloga/específica, com (IC95%: 0,533 - 29,195). Já para os doadores com tipo de doação voluntária, a chance de ser doador HIV-positivo foi 11,941 vezes maior que a de doadores com doação autóloga/específica, com (IC95%: 1,502 - 82,352).

A análise da OR igual a 1,954 com (IC95%: 1,424 - 2,635) para a variável status sorológico ANTI – Hbc (Tabela 7) indica que doadores com ANTI-Hbc – positivo, com mesmo valor para as outras variáveis do modelo, apresentaram uma chance de ser HIV – positivo 1,954 vezes maior que os doadores com resultado anti- Hbc – negativo.

A análise da OR igual a 3,136 com (IC95%: 1,964 - 4,831.) para a variável “status sorológico para sífilis” (Tabela 7) indica que doadores com sífilis-positivo, com mesmo valor para as outras variáveis do modelo, apresentaram uma chance de ser HIV-positivo 3,136 vezes maior que aqueles com sífilis-negativo.

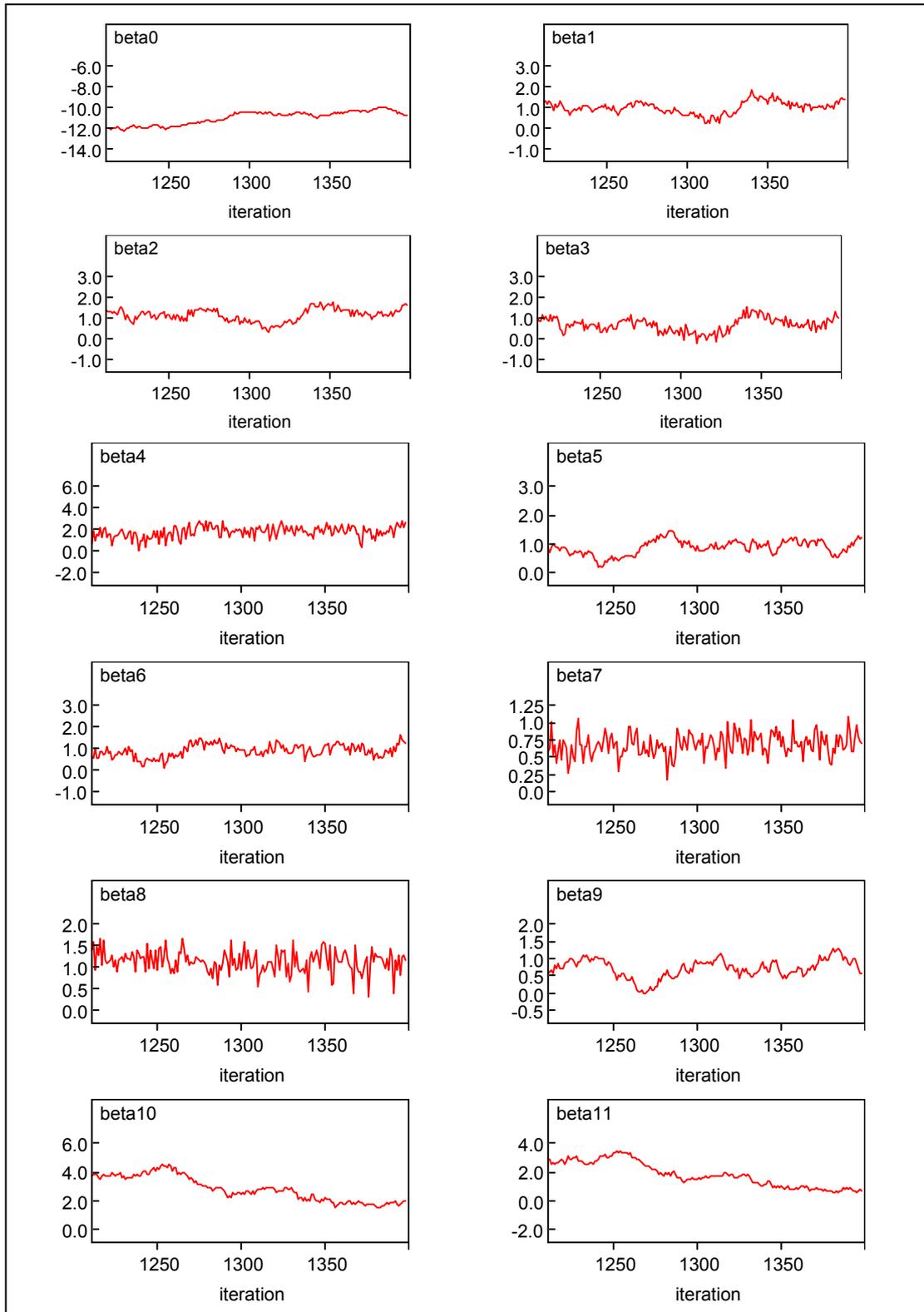


Figura 11. Gráficos do traço para cada parâmetro do modelo, realizando-se 1400 iterações, no WinBUGS 14.

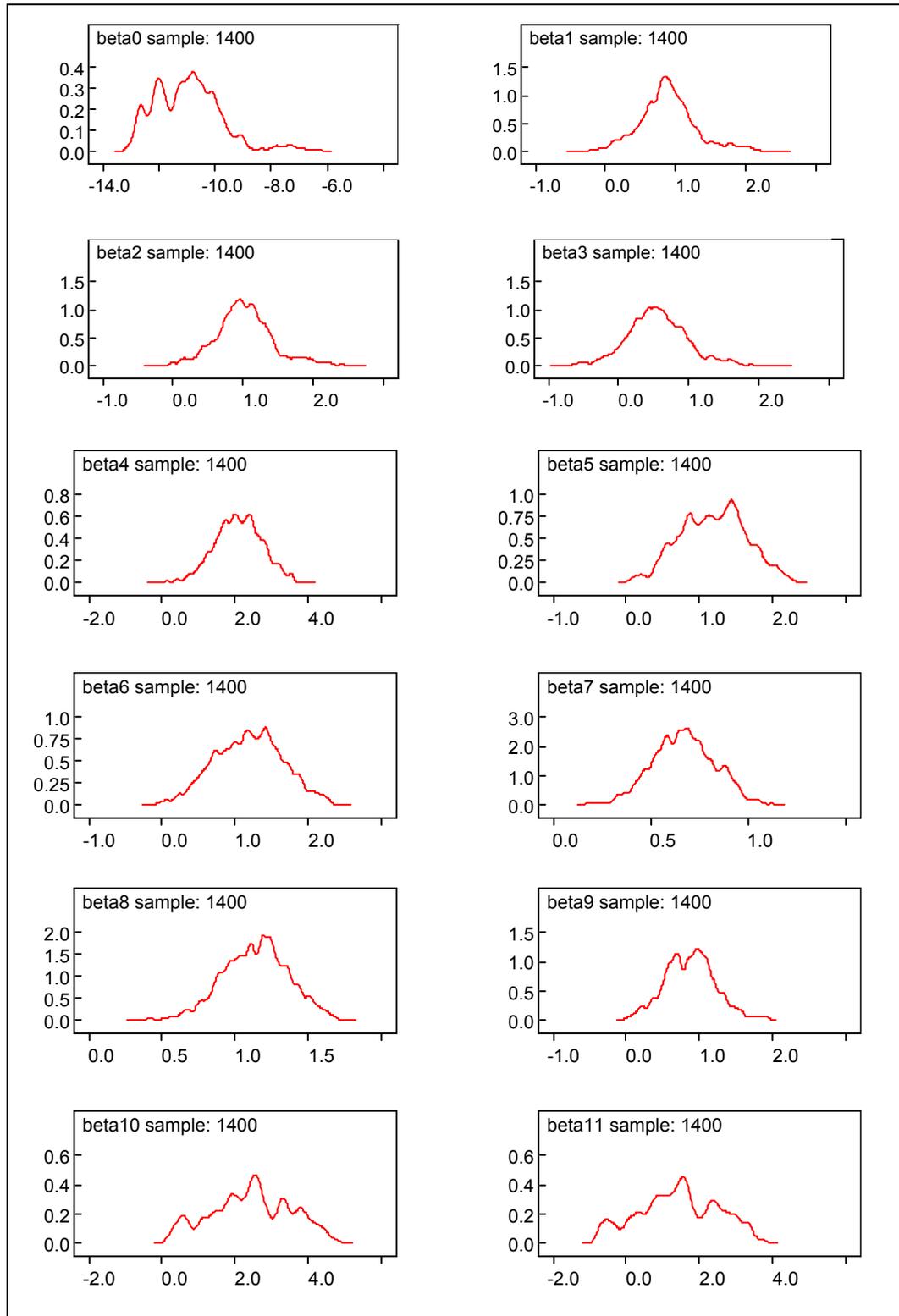


Figura 12. Gráficos da densidade posteriori para cada parâmetro do modelo, realizando-se 1.400 iterações, no WinBUGS 14.

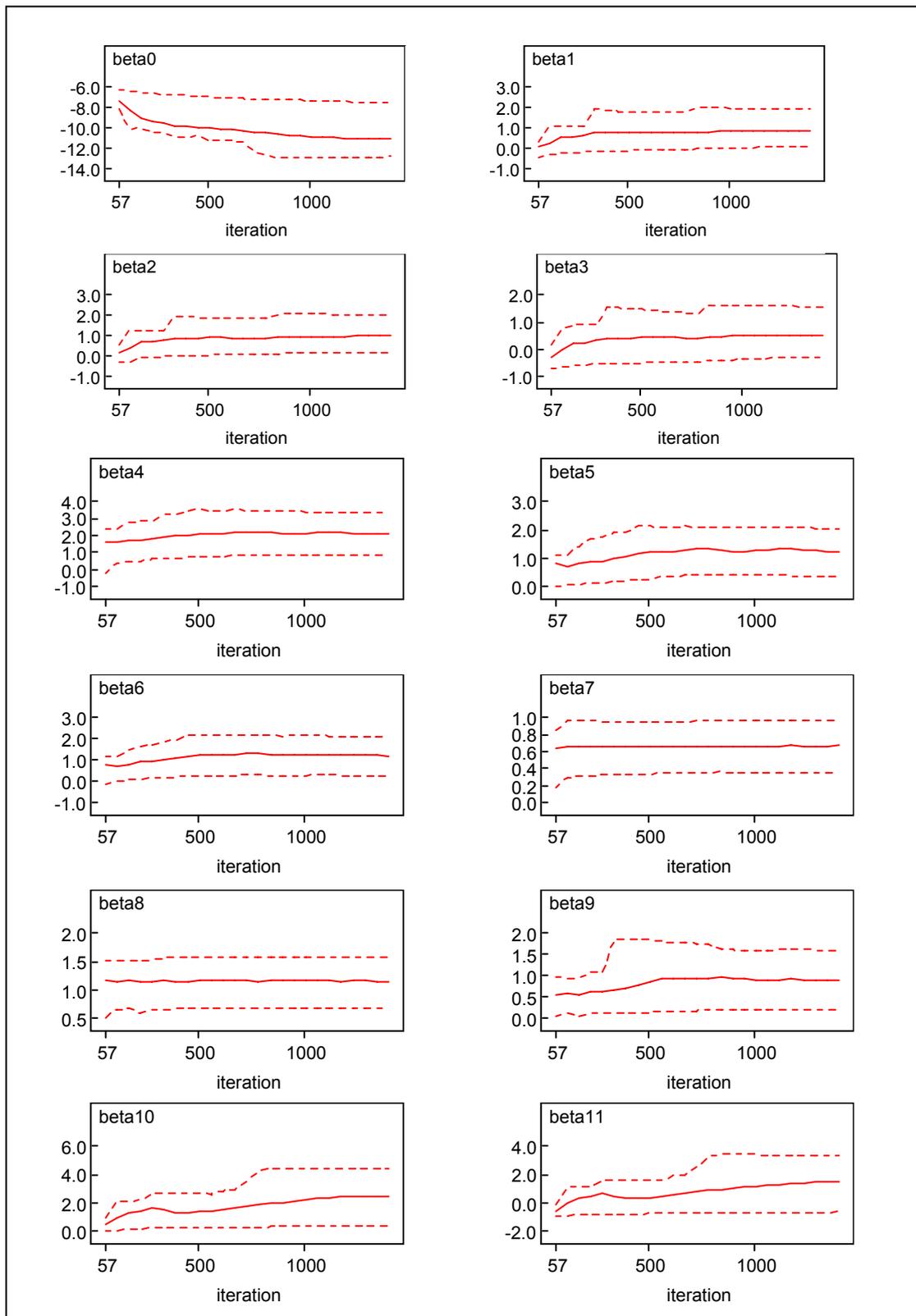


Figura 13. Gráficos dos quantis posteriori para cada parâmetro do modelo, realizando-se 1.400 iterações, no WinBUGS 14.

Observando-se a tabela 7, temos que a estimativa bayesiana no que refere aos IC 95% foi de amplitude menor em 7 variáveis, tais como: analfabeto cuja diferença foi de $D=2,398$, ensino fundamental ($D=0,541$), ensino médio ($D=0,439$), anti-Hbc ($D=0,065$), sífilis ($D=0,004$), voluntária ($D=2,072$) e reposição($D=1,277$). E pelo método clássico a amplitude foi menor em 4 variáveis: idade entre 18 a 28 anos ($D=1,643$), idade entre 29 a 39 ($D=1,516$), idade entre 40 a 49 ($d=0,867$) e região de residência ($D=1,079$). Mas que com relação às OR estimadas pelos dois métodos, a diferença chegou no máximo a $(0,355)$, observado na variável ensino fundamental.

5. CONCLUSÕES

Conclui-se que, pelo método clássico os fatores de risco em potencial, relacionado à infecção HIV comparados com o grupo HIV-negativo foram: idade 18 a 28 anos (OR=2,392; IC95%: 1,098 - 5,209), 29 a 39 anos (OR=2,742; IC95%: 1,266 - 5,940); escolaridade analfabeto (OR=7,935; IC95%: 2,100 - 29,987), ensino fundamental (OR=2,955; IC95%: 1,092 - 7,996), ensino médio (OR=2,971; IC95%: 1,059 - 8,335); status sorológico positivo para anti-Hbc (OR=1,968; IC95%: 1,430 - 2,706); status sorológico positivo para sífilis (OR=3,232; IC95%: 2,101 - 4,972); residir na RMR (OR=2,234; IC95%: 1,268 - 3,937) e tipo de doação voluntária (OR=11,799; IC95%: 1,646 - 84,568).

.Pelo método bayesiano, concluiu-se que a força da associação entre a infecção HIV e os fatores de risco em potencial, entre doadores de sangue, foram estimadas como segue: idade 18 a 28 anos 2,450 (IC95%: 1,101 - 6,855), idade 29 a 39 anos 2,793 (IC95%: 1,229 - 7,419), escolaridade: analfabeto 8,174 (IC95%: 2,254 - 27,743), ensino fundamental 3,313 (IC95%: 1,452 - 7,815), ensino médio 3,290 (IC95%: 1,321 - 8,158), anti-Hbc positivo 1,954 (IC95%: 1,424 - 2,635), sífilis positivo 3,136 (IC95%: 1,964 - 4,831), residir na RMR 2,408 (IC95%: 1,220 - 4,968), tipo de doação voluntária 11,941 (IC95%: 1,502 - 82,352).

Assim, quanto à aplicação dos dois métodos: clássico e bayesiano, concluiu-se que as duas metodologias foram aproximadamente equivalentes nas estimativas dos OR e de seus intervalos de confiança.

Finalmente, os critérios considerados de fragilidade quanto à associação a infecção por HIV em doadores de sangue, são indivíduos com idade mais próxima ao limite mínimo para doação, com menor nível de escolaridade, que forem positivos para o anti-Hbc e Sífilis, residentes na RMR e cujo tipo de doação for à voluntária. Portanto, esses critérios devem ser considerados na rotina da triagem clínica de doadores de sangue, como também, nas campanhas de captação de doadores, visando a melhoria da qualidade do sangue a ser transfundido.

6. RECOMENDAÇÕES

- Sugere-se a necessidade de medidas conjuntas, que envolvam o indivíduo, a família, os profissionais de saúde e ações governamentais, no controle tanto da infecção por HIV, quanto de outras patologias. A prevenção da infecção por HIV, no que se refere à doação de sangue, necessita de aprofundamento científico através de novas pesquisas que auxiliem na tomada de decisão quanto a essas ações e de testes mais sensíveis, para garantir a segurança transfusional.
- Sugere-se a necessidade de aprofundamento científico através de novas pesquisas que auxiliem na explicação quanto ao tipo de doação voluntária ter sido considerado risco, pois a literatura especializada tem definido doadores voluntários como aqueles de baixo risco de transmissão de doenças infecciosas e principalmente aqueles cujo risco de transmitir doenças em período de janela imunológica é próximo de zero. Portanto sugere-se ainda a realização de campanhas de conscientização quanto à doação voluntária de sangue.
- Recomenda-se a realização de campanhas de captação de doadores de sangue que focalize os indivíduos de maior idade e os que têm maior nível de escolaridade.
- Propõe-se, em continuidade do presente estudo, a implementação de modelos mistos com efeitos aleatórios para o mesmo conjunto de dados.
- Propõe-se a implementação de priori informativas, na abordagem bayesiana, com base no resultado dessa dissertação.
- Sugere-se, ainda, implantar rotina de pré-triagem antes da triagem sorológica, baseada no resultado de que um doador com valor de Logit maior ou igual a (-5,828) seja rejeitado para doação de sangue.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). Portaria nº 488, de 1 de Junho de 1998. Estabelece procedimentos seqüenciados para detecção de anticorpos anti-HIV, que deverão ser seguidos pelas unidades hemoterápicas, públicas ou privadas, visando a redução de resultados falso-positivos ou falso-negativos. **Diário Oficial [da] República Federativa do Brasil**, 1998. Disponível em: <http://www.anvisa.gov.br/legis/portarias/488_98htm>. Acesso em: 29 dez. 2003.

_____. Resolução RDC nº 343, de 13 de dezembro de 2002. Aprovar o regulamento técnico para obtenção, testagem, processamento e controle de qualidade de sangue e hemocomponentes para uso humano, que consta como Anexo I. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 17 jan. 2003. Seção I, p.40-50.

ALBERT, J. H. AND CHIB, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. **Journal of the American Statistical Association** 88, 669-679.

AMORIM, M. M. R.; FAÚNDES, A. ; SANTOS, L. C. Fatores de risco para síndrome do desconforto respiratório do recém-nascido na pré-eclâmpsia grave. **Revista Brasileira de Saúde Materno Infantil**, Recife, v.1, n.1, p. 65-69, jan./abr. 2001.

ANDRADE NETO, J.L.; PINTARELLI, V.L.; FELCHNER, P.C.; MORAIS, R.L.; NISHIMOTO, F.L. HIV prevalence among blood donors in a blood bank in Curitiba (Brazil). **Braz J Infect Dis**, Curitiba, v.6, n.1, p.15-21, feb. 2002.

ANDRIEU, N.; PREVOST, T.; ROHAN, T.E.; LUPORSI, LÊ M.G.; GERBER, M.; ZARIDZE, D.G.; LIFANOVA, Y.; RENAUD, R.; LEE, H.P.; DUFFY, S.W. Variation in the interaction between familial and reproductive factors on the risk of breast cancer according to age, menopausal status, and degree of familiarity. **Internation journal of epidemiology**. v. 29, n.2, p.214-223. 2000.

ANTELMAN, GORDON. **Elementary Bayesian Statistics**, Edward Elgar Publishing Limited & Lansdown Place, Cheltenham Glas,UK, 1977.

BARCELLOS N.T, FUCHS S.C, FUCHS F.D. Prevalence of and risk factors for HIV infection in individuals testing for HIV at counseling centers in Brazil. **Sex Transm Dis**. Rio Grande do Sul, v.30, n.2, p.166-73, fev.2003.

BEDRICK, E. J., CRISTENSEN, R.; JOHNSON, W. Bayesian binomial regression. **The American Statistician**. V.51, n.3, p.211-218.1997.

BERGERUD, W. A.; REED, W. J. **Bayesian statistical methods**. Chapter 7. In V. Sit, and B.. Taylor. (editors). 1998. Statistical methods for adaptive management studies. Res. Br., B. C. Min. For., Res. Br., Victoria, BC, Land

Manage. Handb. No.

BOLETIM INFORMATIVO DST/AIDS. Sífilis congênita. Recife: Secretaria de Saúde de Pernambuco, ano 2, n.1, jan./abr.2002.

_____. Transmissão materno-infantil do HIV. Recife: Secretaria de Saúde de Pernambuco, ano 2, n.2, maio/ago. 2002.

_____. A polêmica posição da igreja católica em relação ao uso do preservativo como prevenção do HIV. Recife: Secretaria de Saúde de Pernambuco, ano III, n.4, maio./dez. 2003.

BOX, G.E.P.; TIÃO, G.C. **Bayesian inference in statistical analysis**. New York: Addison Wesley, 1973. 588p.

BRASIL, G.H. **Estatística Bayesiana**: alguns aspectos Teóricos e Práticos, In: Reunião Regional de Estatística, 22., 1991, Recife. Mini curso ...Recife, UFPE,1991. Apostila.

BRASIL. Ministério da Saúde. Gabinete do Ministro. Portaria nº1376 de 19 de novembro de 1993. Aprova alterações na Portaria nº721/GM de 09.08.89, que aprova Normas Técnicas para coleta, processamento e transfusão de sangue, componentes e derivado, e dá outras providências. **Diário Oficial da União**, Brasília, DF, 2 dez.1993. Seção I, p.18.405.

BRASIL. Ministério da Saúde. Portaria nº 59, de 28 de janeiro de 2003. Definir e normatizar a sub-rede de laboratórios do Programa Nacional de DST e Aids, que realizam testes sorológicos para a detecção de anticorpos anti-HIV, integrante da Rede Nacional de Laboratórios clínicos, em conformidade com a Portaria Nº 15, de 03 de janeiro de 2002. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 30 jan. 2003. seção I, p.87-89.

BRASIL.Ministério da Saúde.Secretaria de Políticas de Saúde.Coordenação Nacional de DST e AIDS. **Simpósio satélite**: a epidemia de Aids no Brasil: situação e tendência. Brasília (DF), 1997. 311p.

_____. **Prevalência do HIV nos bancos de sangue**. Brasília.. Disponível em: <<http://www.aids.gov.br/uvad/bancosangue.htm>>. Acesso em: 11/12/2003.

CANUTTI Júnior. Risco transfusional: metodologia e estudo. **Série de Monografia Escola Brasileira de Hematologia**, Ribeirão Preto, v.5, p.90-99, 1998.

CARY,N.C. **Statistical analysisn system**: guide for personal computer-version 6. North Carolina: SAS Institut IMC, 1985. 378p.

CHEN, M. H. , SHAO, A. & IBRAHIM, J. **Monte Carlo Methos in Bayesian Computation**. Springer-Verlag, NY, 2000.

CONCATO, J. ; FEINSTEIN, A. R.; HOLFORD, T. R. The risk of determining

risk with multivariable models. **Annals of Internal Medicine**, v.118, n.3, p.201-210, fev.1993.

CORDEIRO, G.M. **Modelos lineares generalizados**. Campinas, VII Simpósio Nacional de Probabilidade e Estatística, 1986.286p.

COVAS, D. T. Risco de transmissão do HIV-1 pelas transfusões de sangue. **Série de Monografias Escola Brasileira de Hematologia**, Ribeirão Preto, v.5, p.100-106, 1998.

COX, D. R. **The analysis of binary data**. London: Methuen, 1969. 142p.

DEMÉTRIO, C.G.B. **Modelos lineares generalizados em experimentação agrônômica**. Piracicaba, SP. ESALQ/USP, 2002.

EBRAHIM, G. J. **Análise multivariada**. Tradução: Antônio Carlos Figueira et. al. Recife: IMIP, 2001. 160p.

FREIRE, J.C.M. **Métodos para estimar “odds ratio” através de abordagens clássicas e bayesianas, em estudos epidemiológicos**. 2001. 62 f. (Dissertação de Mestrado) – Biometria, Universidade Federal Rural de Pernambuco, Recife.

FUNDAÇÃO DE HEMATOLOGIA E HEMOTERAPIA DO ESTADO DE PERNAMBUCO. **Curso de sorologia**. 2001. 57f. Apostila.

GAMERMAN, D. **Markov Chain Monte Carlo: Stochastic simulation for bayesian inference**. Chapman & Hall, 1997.

GILKS, W. R.; RICHARDSON, S., SPIEGELHALTER, D. J. **Markov Chain Monte Carlo in Practice**. London: Chapman & Hall, 1996.

GOMES, M. M. **Medicina baseada em evidências: princípios e práticas**. Rio de Janeiro; Reichmann & Affonso, 2001.150p.

GUIMARÃES, I. A. **Construção e avaliação de uma regra de reconhecimento e classificação de clientes de uma instituição financeira com base na análise multivariada**. 2000. 87 f. (Dissertação de Mestrado) - Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2000.

HASTINGS, W. K. **Monte Carlo sampling methods using Markov chains and their applications**. *Biometrika*, 57, 97 a 109, 1970.

HOSMER, D.W.; LEMESHOW, S. **Applied Logistic Regression**. New York: John Wiley, 1989. 307p.

HULLEY, S. B.; CUMMING, S. R.; BROWNER, W. S.; GRADY, D.; HEARST, N.; NEWMAN, T. B. **Delineamento a pesquisa clínica: uma abordagem epidemiológica**, 2. ed. Porto Alegre: Artmed, 2003. 374p.

JEFFREYS, H. **Theory of probability**. 3 ed., Oxford: Clarendon Press, 1961. 578p.

KAHN, S.R.; JOSEPH, L.; ABENHAIM, L.; LECIERC, J.R. Clinical Prediction of Deep Vein Thrombosis in Patients with Leg Symptoms. **Thromb Haemost.** v .81, n.3, p.353-357,1999.

KUPEK E.J. HIV seroprevalence among blood donors in southern Brazil in the decade of 1990. **Braz J Infect Dis.** Santa Catarina, v.4, n.5, p.217-225, oct. 2000.

KUPEK E.J. The Reduction of HIV transfusion risk in southern Brazil in the 1990s. **Transfus Med.** Santa Catarina, v.11, n.2, p.75-78, apr. 2001.

LEANDRO R. A. **Introdução à Estatística Bayesiana**. Departamento de Ciências Exatas, ESALQ/USP Piracicaba, SP - 7 a 13/07/2001.

LEE SY, SONG XY. Bayesian analysis of structural equation models with dichotomous variables. Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong. **Stat Med.** V.22, n.19, p.3073-30, oct. 2003.

MARSHALL G., SHROYER A.L.W., GROVER F.L. & HAMMERMEISTER K.E. Bayesian-logit model for risk assessment in coronary artery bypass grafting. **Ann Thorac Surg.** v.57, n.6, p.1492-1499, 1994.

MARTINS, C. A. C. **Análise de regressão logística**. 1998. 53 f. (Dissertação de Mestrado) – Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife,1988.

MCFARLAND, W.; Mvere, D.; Shamu, R.; Katzenstein, D. Risk factors for HIV seropositivity among first-time blood donors in Zimbabwe. **Transfusion,** Arlington, v.38, p. 279-284, mar. 1998.

MENDES, T. F.; PITTELLA, A. M. **Recentes avanços em hepatites**. São Paulo: Fundo Editorial BYK, 1993.191p.

MING-HUI, CHEN; QI-MAN, SHAO. Propriety of posterior distribution for dichotomous quantal response models. **Proceedings of the American Mathematical Society.** v.129, n.1, p.293-302, aug.2000.

MURRAY, P.R.; DREW, W.L.; KOBAYASHI, G.S.; THOMPSON Jr, J.H. **Microbiologia médica**. Rio de Janeiro. Guanabara Koogan, 1992. 513p.

PAULA, G.A. **Modelos de regressão com apoio computacional**. Instituto de Matemática e Estatística, Universidade de São Paulo, 2004. 294p. Disponível em <<http://www.ime.usp.br/~giapaula>> acesso em 06/03/2004.

PEREIRA, M. G. **Epidemiologia: teoria e prática**. Rio de Janeiro: Guanabara

Koogan, 2000. 596p.

PILOTE, L.; JOSEPH, L.; BÉLISLE P.; ROBINSON, K.; LENTE, F. V.; TAGER, I. B. Iron stores and coronary artery disease: a clinical application of a method to incorporate measurement error of the exposure in a logistic regression model. **Journal of Clinical Epidemiology**. v. 53, p.809-816.

POLLARD, William E., **Baysian satistics for evaluation research: an introduction**, Beverly Hills: sage publications, 1986.

RAHLENBECK SI, YOHANNES G, MOLLA K, REIFEN R, ASSEFA A. Infection with HIV, syphilis and hepatitis B in Ethiopia: a survey in blood donors. Department of Microbiology, Gondar College of Medical Sciences, Ethiopia. **Int. J. STD AIDS**. V.8, n.4, p.261-264, Apr.1997.

REICHE E.M.V.; VOGLER I.H.; MORIMOTOH. K.; BORTOLEIRO A.L.; MATSUO T.; YUAHASI K.K.; CANCIAN S.Jr., KOGUEICHI R.S. Evaluation of surrogate markers for human immunodeficiency virus infection among blood donors at the blood bank of “hospital universitário regional norte do Paraná”, Londrina, PR, Brazil. **Rev. Inst. Med. Trop. S.Paulo**, v.45, n.1, p.23-27, jan./feb.2003.

ROBERT, C. P.; CASELLA, G. **Monte Carlo Statistical Methods**. Springer Texts in Statistics, 1999.

ROUQUAYROL, M. Z.; ALMEIDA FILHO, N. **Epidemiologia & Saúde**. 4.ed. Rio de Janeiro: MEDSI, 2003. 728p.

_____. _____. 5.ed. Rio de Janeiro: MEDSI, 1999. 570p.

SAMPAIO, D. A. **Caracterização descritiva de candidatos à doação de sangue do grupo etário de 50 a 60 anos, de ex-doadores excluídos por limite de idade, de não-doadores acima de 50 anos e análise comparativa com candidato de 18 a 28 anos da cidade do Recife**. 2001. 210 f. Tese (Doutorado) - Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, 2001.

SANTOS, J.N.M. **Abordagem bayesiana do equilíbrio de Hardy-weinberg**. 2001. 52f. Dissertação (Mestrado) – Biometria, Universidade Federal de Pernambuco, Recife.

SOUZA, S. M. B. Perfil epidemiológico da clientela dos centros de testagem e aconselhamento para o HIV em Goiás – Brasil. **Revista Eletrônica de Enfermagem**. vol. 4, nº2, p. 49. 2002. Disponível em < <http://www.fen.ufg.br> > acesso em 02/02/2004.

SCHIAFFINO, A.; RODRIGUEZ, M.; PASARIN, M. I.; REGIDOR, E.; BORRELL, C.; FERNANDEZ, E. Odds ratio or prevalence ratio? Their use in cross-sectional studies, **Gac Sanit**, Spanish, v.17, n.1, p.70-74, jan./feb.2003.

SCHMID C. H.; ROSNER B. A bayesian approach to logistic regression models having measurement error following a mixture distribution. **Stat Med.** v.12, n.12, p.1141-1153, 1993.

SANTANA F. A.; POLO F. J. V.; PÉREZ J. C. R. Herramientas para la investigación biomédica: la perspectiva bayesiana. **Med Clin.** Barcelona, v.119, n.7, p. 265-268, 2002.

SILVA, A. L. C.; SUÁREZ G. P. Qué es la inferência bayesiana? **JANO EMC.** Viernes, v.58. n.1338, p.65-66, mar.2000.

STROMBERG U. Prevalence ODDS ratio v prevalence ratio. **Occup Environ Med.** v. 51, p.143-144, 1994.

VIALLEFONT, V.; RAFTERY, A.E.; RICHARDSON, S. Variable selection Bayesian model averaging in case-control studies. **Statistics in Medicine.** v.20, n.21, p.3215-3230, 2001.

WALD, A. (1943) " Tests of statistical hypotheses concerning several parameters when the number of observations is large". **Trans. Amer. Math. Soc.**, 54, 426-482.

ZACHARIAH, R.; HARRIES, A. D.; NKHOMA, W., ARENDT, V.; SPIELMANN, M. P.; BUHENDWA, L.; CHINGI, C.; MOSSONG, J. HIV prevalence and demographic risk factors in blood donors. **East Med J.** v.79, n.2, p.88-91, feb.2002.

ZELLNER, A.; ROSSI, P. Bayesian analysis of dichotomous quantal response models. **Journal of Econometrics.** 25, 365- 393.22, 1984.

APÊNDICE

Medidas de risco

Muitas medidas de risco são utilizadas, em epidemiologia, para representar quantitativamente a relação entre os eventos. Essas medidas estão relacionadas ao tipo do delineamento do estudo (Hulley, 2003). A partir da tabela 8 e utilizando os dados desse estudo calculou-se as principais medidas de risco:

Tabela 8. Tabela-padrão para a apresentação dos resultados em quatro tipos de estudo: ensaio clínico randomizado e estudos de coorte, de caso-controle e transversal.

Exposição ao fator	Doença		Total
	Sim	Não	
Sim	a	b	a + b
Não	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

Significado das letras:

a = número de indivíduos expostos e doentes

b = número de indivíduos expostos e sadios

c = número de indivíduos não-expostos e doentes

d = número de indivíduos não-expostos e sadios

N = a + b + c + d = número total de pessoas

Neste trabalho, por exemplo, a associação entre o marcador sorológico da sífilis e a infecção por HIV é vista na tabela 9 de contingência, abaixo.

Tabela 9. Marcador sorológico da sífilis versus infecção por HIV. jan98-nov/03, Recife-PE.

Variável Explicativa(X) Marcador sorológico da sífilis	Variável resposta(Y): ser doador HIV-positivo		
	sim	não	Total
positivo	24(a)	3.446(b)	3.470(a+b)
negativo	193(c)	102.540 (d)	102.733(c+d)
Total	217(a+c)	105.986(b+d)	106.203(a+b+c+d)

As medidas de risco calculadas pelos dados observados neste estudo, observando a tabela 9, são:

Risco Relativo(RR):

È a razão entre dois coeficientes de incidência, ou melhor, entre o coeficiente de infecção por HIV nos expostos e nos não-expostos.

$$Risco\ Relativo(RR) = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{0,692}{0,188} = 3,682$$

Odds Ratio(OR):

$$Odds\ Ratio(OR) = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c} = \frac{24 * 102.540}{3446 * 193} = 3,700$$

Mostrando que, quando a doença é rara, o OR fornece uma boa estimativa do risco relativo:

OR \cong Risco Relativo.

De outra forma pode-se calcular esse Odds Ratio usando a regressão Logística Simples, utilizando-se apenas a variável independente “Status sorológico da sífilis”, o modelo ajustado, para cada método (razão de verossimilhança e Bayesiano), foi:

Método clássico:

Utilizando o SPSS obtiveram-se os seguintes dados na tabela 10 abaixo:

Tabela 10. Estimativas dos β e OR do modelo apenas com a variável sífilis, no SPSS.

Variável	β	SE	Wald	gl	p_wald	OR=Exp(β)	IC 95%	
							inferior	superior
sífilis	1,308	0,217	36,308	1	0,000	3,700	2,418	5,663
constante	-6,275	0,072	7.586,128	1	0,000	0,002		

Assim,

$$\hat{L}ogit(P(y = positivo \setminus x = Sífilis)) = \hat{\beta}_0 + \hat{\beta}_1 Sífilis = -6,275 + 1,308 Sífilis$$

$$\text{Odds Ratio estimado}(OR_{Sifilis=0}^{\wedge}) = e^{\hat{\beta}_0} = e^{-6,275} = 0,002$$

Logo,

$$\text{Odds Ratio estimado}(OR_{Sifilis=1}^{\wedge}) = e^{\hat{\beta}_1} = e^{1,308} = 3,699$$

Método Bayesiano:

Utilizando-se 800 iterações, verificou-se, na figura 14 abaixo, uma boa convergência do processo de estimação, no Winbugs 14.

No primeiro gráfico verifica-se a boa convergência da cadeia de markov para estimar os parâmetros da equação logística abaixo. No segundo e terceiro gráficos, observa-se à suposição de distribuição posterior normal. Na tabela 11, verifica-se as estimativas dos parâmetros β , seu erro padrão(SE) e intervalos com 95% de credibilidade, respectivamente. Dos quais pode-se escrever o modelo:

$$\hat{\text{Logit}}(P(y = \text{positivo} \mid x = \text{Sifilis})) = \hat{\beta}_0 + \hat{\beta}_1 \text{Sifilis} = -6,279 + 1,292\text{Sifilis}$$

E calcular as OR,

$$\text{Odds Ratio estimado}(OR_{Sifilis=0}^{\wedge}) = e^{\hat{\beta}_0} = e^{-6,279} = 0,002$$

$$\text{Odds Ratio estimado}(OR_{Sifilis=1}^{\wedge}) = e^{\hat{\beta}_1} = e^{1,292} = 3,640$$

Tabela 11. Estimativas dos parâmetros e OR do modelo apenas com a variável sífilis, no Winbugs, com 800 iterações.

Variável	Média (β)	SE(β)	IC 95% p/ β		OR=Exp(β)	IC 95% para OR	
			inferior	superior		inferior	superior
Constante	-6,279	0,068	-6,411	-6,144	0,002	0,0016	0,0021
Sifilis	1,292	0,217	0,852	1,695	3,640	2,3439	5,4466

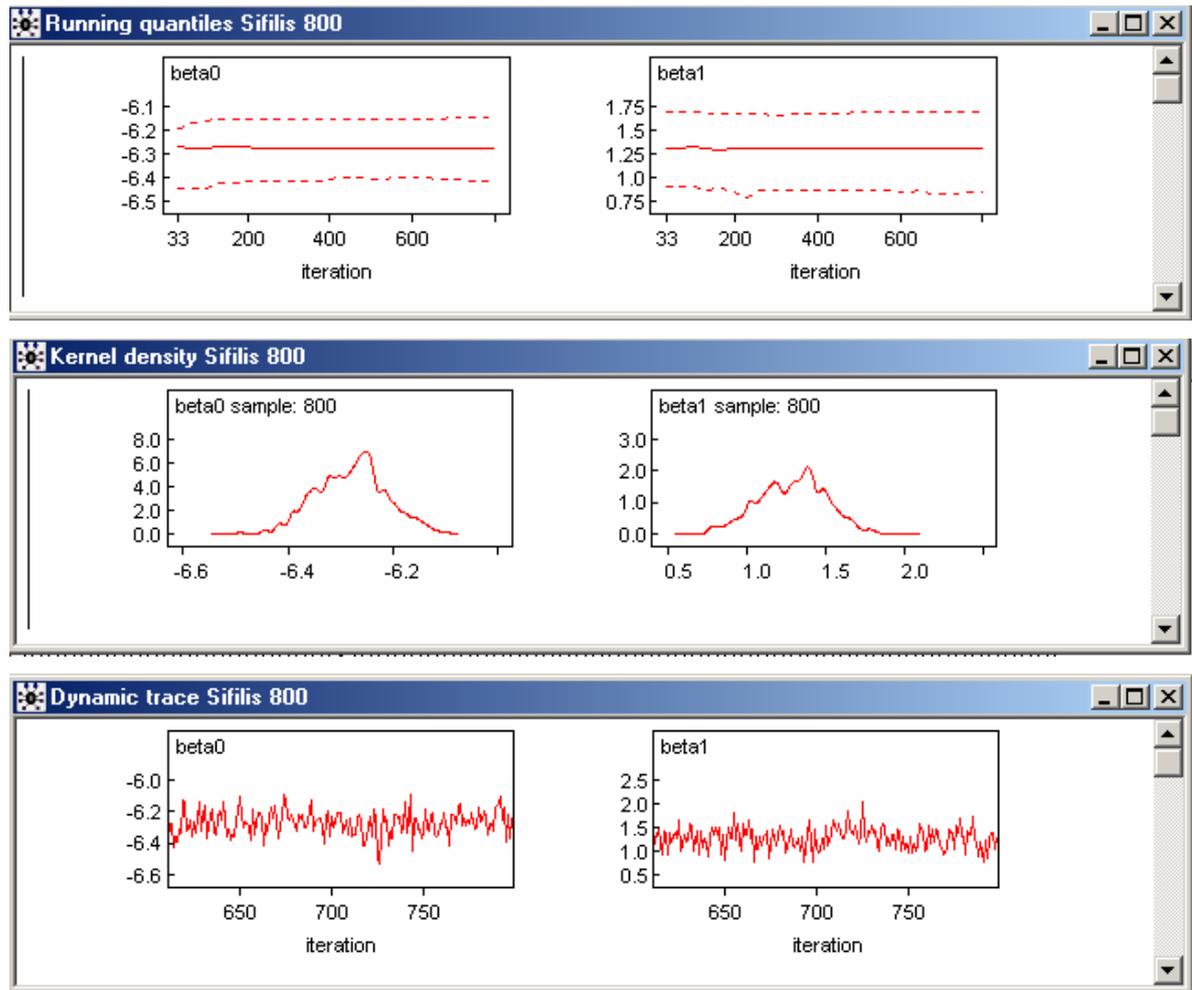


Figura 14. Gráficos de convergência (do traço, da densidade posteriori, dos quantis posteriori) das estimativas dos parâmetros, do modelo da infecção HIV, apenas com a variável sífilis, no Winbugs 14, com 800 iterações.

Nota-se que, a estimação do OR para o status sorológico da Sífilis coincide com a obtida pela tabela 9, como também é bem aproximada da calculada pelo método clássico.

Para a variável independente politômica como “tipo de doação” com três categorias, como mostrada na tabela 12, a análise foi através da criação de duas variáveis “dummy”.

Baseando-se na tabela 12, os Odds Ratio foram calculados para cada tipo de doação, usando a variável “autóloga ou específica” como o grupo de referência, logo os cálculos das estimativas dos odds ratio observados são:

$$OR_{Voluntária} = \frac{(81 \times 3.314)}{(1 \times 17.661)} = 15,199$$

$$OR_{Re\ posição} = \frac{(135 \times 3.314)}{(1 \times 85.011)} = 5,263$$

Tabela 12. Classificação cruzada entre infecção HIV e o tipo de doação. Recife-PE. jan1998-nov2003.

Tipo de doação	Infecção por HIV (Y)		Total	Odds Ratio (OR)
	positiva	negativa		
<i>Autóloga</i>	1	3.314	3.315	1
<i>Voluntária</i>	81	17.661	17.742	15,199
<i>Reposição</i>	135	85.011	85.146	5,263
<i>Total</i>	217	105.986	106.203	-

Essa saída da tabela acima é encontrada na literatura quando existe um grupo referência observado, do qual os outros grupos são comparados.

Essas mesmas estimativas dos OR podem ser observadas pela regressão logística com uma escolha apropriada de variáveis dummy.

O método para especificar as variáveis dummy envolve estabelecer todas elas iguais a zero para o grupo de referência, e então estabelecer uma única variável dummy igual a “1” para cada um dos outros grupos de categoria. Isso é ilustrado na tabela 13.

Tabela 13. Especificação de variáveis dummy para tipo de doação usando autóloga/ específica como grupo de referência.

Tipo de doação	Variável Dummy	
	D1	D2
<i>Autóloga/ específica (1)</i>	0	0
<i>Voluntária (2)</i>	1	0
<i>Reposição (3)</i>	0	1

Para o método clássico, no programa SPSS é só especificar que o grupo de referência é o primeiro. Obteve-se para a regressão logística com infecção por HIV (variável dependente) e tipo de doação (variável independente) as seguintes estimativas para os parâmetros da equação é vista na tabela 14.

Tabela 14. Estimativas dos parâmetros e OR do modelo apenas com a variável tipo de doação transformada em dummy, com grupo de referência “autóloga ou específica”, no SPSS.

Tipo de doação	β	SE	Wald	gl	p_wald	OR=Exp(β)	IC 95%	
							inferior	superior
<i>Autóloga/específica</i>						1		
<i>voluntária</i>	2,634	0,964	7,467	1	0,006	13,927	2,106	92,109
<i>reposição</i>	1,573	0,961	2,679	1	0,102	4,823	0,733	31,734
<i>constante</i>	-8,019	0,957	70,144	1	0,000	0,000		

$$\hat{\text{Logit}}(P(y = \text{positivo} \mid x = \text{Voluntária, Re posição})) = \hat{\beta}_0 + \hat{\beta}_1 * \text{Voluntária} + \hat{\beta}_2 * \text{Re posição}$$

$$\hat{\text{Logit}}(P(y = \text{positivo} \mid x = \text{Voluntária, Re posição})) = -8,019 + 2,634 \text{Voluntária} + 1,573 \text{Re posição}$$

Logo,

$$\text{Odds Ratio estimado}(OR_{(\text{Voluntária, autóloga / específica})}) = e^{\hat{\beta}_1} = e^{2,634} = 13,93$$

$$\text{Odds Ratio estimado}(\hat{OR}_{(\text{Re posição, Autóloga / específica})}) = e^{\hat{\beta}_2} = e^{1,573} = 4,821$$

E para o método bayesiano, através do winbugs temos, para 1400 iterações, as estimações (tabela 15) e boa convergência observada na figura 15, assim temos:

$$\hat{\text{Logit}}(P(y = \text{positivo} \mid x = \text{Voluntária, Re posição})) = \hat{\beta}_0 + \hat{\beta}_1 * \text{Voluntária} + \hat{\beta}_2 * \text{Re posição}$$

$$\hat{\text{Logit}}(P(y = \text{positivo} \mid x = \text{Voluntária, Re posição})) = -8,113 + 2,721 \text{Voluntária} + 1,665 \text{Re posição}$$

Logo,

$$\text{Odds Ratio estimado}(OR_{(\text{Voluntária, autóloga / específica})}) = e^{\hat{\beta}_1} = e^{2,721} = 15,196$$

$$\text{Odds Ratio estimado}(\hat{OR}_{(\text{Re posição, Autóloga / específica})}) = e^{\hat{\beta}_2} = e^{1,665} = 5,286$$

Tabela 15. Estimativas dos parâmetros e OR do modelo apenas com a variável tipo de doação, no Winbugs, com 1.400 iterações.

Tipo de doação	Média (β)	SE(β)	IC 95% p/β		OR=Exp(β)	IC 95% para OR	
			Inferior	superior		inferior	superior
constante	-8,113	0,984	-9,775	-6,527	0,0003	0,0001	0,0015
voluntária	2,721	0,993	1,088	4,420	15,1955	2,968	83,096
reposição	1,665	0,990	0,077	3,370	5,2857	1,0800	29,078

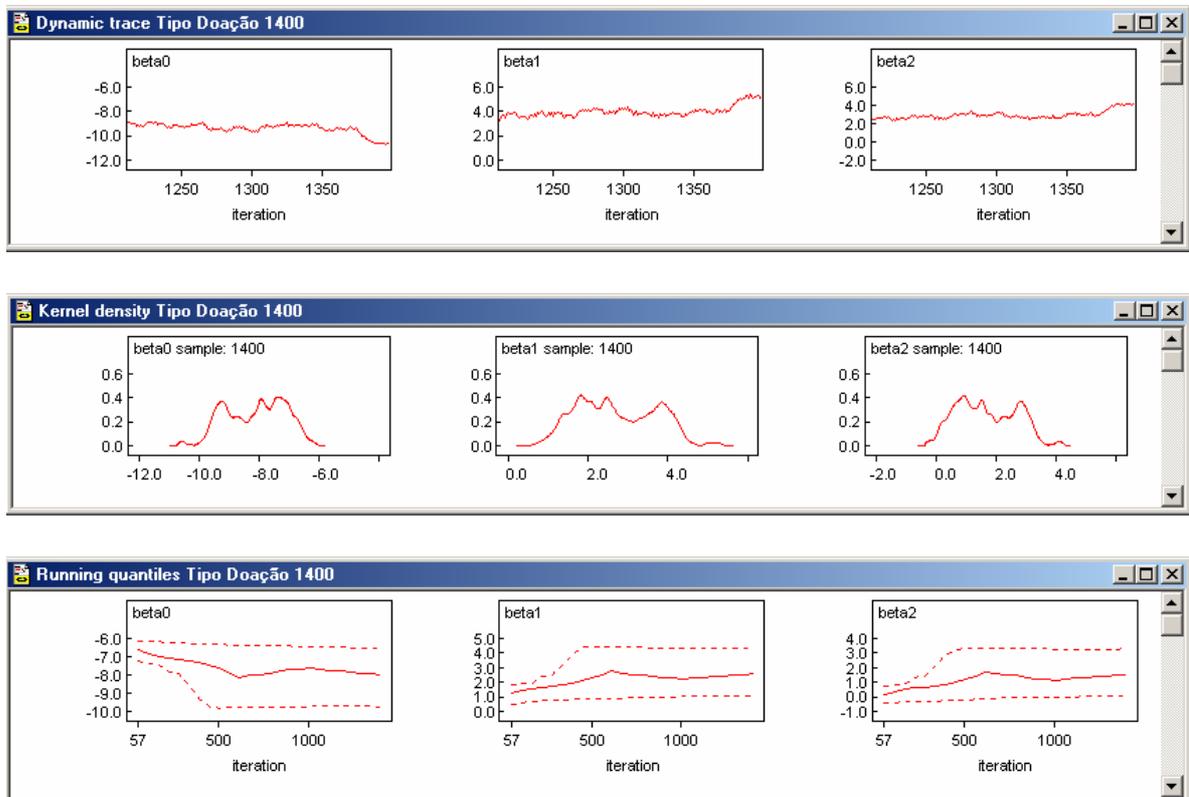


Figura 15. Gráficos de convergência (do traço, da densidade posteriori, dos quantis posteriori) das estimativas dos parâmetros, do modelo da infecção HIV, apenas com a variável tipo de doação, no Winbugs 14, com 1.400 iterações.

ANEXO 1. FICHA DE DOAÇÃO

Verso	
TRAGEM CLÍNICA	
01- Qual a data da última doação? 02- Exames Alérgicos/Problemas na doação? 03- Jejum ou vigília? 04- Alimentação copiosa há menos de 4 horas? 05- Gravidez ou amamentando? 06- Perda acentuada de peso recentemente? 07- Aborto nos últimos 3 meses? 08- Vacina no último ano? Qual? 09- Resir. / Dox. Garg. / Febre / Vômito / Diarréia? 10- Ictericia ou Hepatite? Com que idade? 11- Contato íntimo com quem já teve hepatite? 12- Doença pele/autoimune/cancer/cirroses/gonítes? 13- Doença de Chagas/Placido pelo barbeiro? 14- Malaria/Profilaxia Malaria/Área endêmica? 15- Sífilis ou outra doença venérea? 16- Portador do vírus HTLV? 17- Doença grave / Cirurgia / Intern. Hospitalar? 18- Transf. sangue/Hemodér. Recipientes? 19- Doença de sangue ou Hemorragia? 20- Candidato / Respirat. / Alérgico/ Distúrbio Renal? 21- Bebida alcoólica diariamente/ 24 horas?	S N 22- Desmaio/ Epilepsia/Convulsão? 23- Uso hormônio do crescimento? 24- Medicação nos últimos 15 dias? 25- E ou já foi violado em drogas? 26- Tratagem / Acupuntura últimos 12 meses? 27- Apresentação ínguas no corpo? 28- Diarréia ou febre persistente? 29- Sudorese noturna? 30- Marchas estrano, na boca/lacrimas/pele? 31- Voto apurais p/ realizar o teste p/ o HIV? 32- Teve alguma relação homossexual? 33- Felic. Sexual c/ Desc. / Prost. / Garot. Prog? 34- Deu / recebeu drogas / dinheiro por sexo? 35- Relação sexual c/ suspeito de AIDS? 36- Relacionamento sexual c/ + de 1 parceiro? 37- Estiver preso / sexo c/ ac-deternto
Resp. Triagem	Código
Assinatura	Assinatura
Avaliação	Desc. Motivo de Recusa
Mot. Rec.	1- () Doação s/ Transfusão 2- () Coleta de Amostra 3- () Doação p/ Transfusão
TERMO DE RESPONSABILIDADE Declaro que compreendi as questões a mim formuladas e que as informações que eu prestei são verdadeiras, e desde que atendam aos requisitos para doação de sangue, permito que sejam retirados _____ ml de meu sangue para uso a critério do HEMOPE. Estou ciente de que sejam feitos testes de triagem sorológica em meu sangue e se algum resultado se apresentar alterado, serei convocado para receber informações e se necessário repetir os exames. Estou ainda ciente de que posso vir a sofrer alguma reação à doação e que fui orientado sobre ela. Recife, _____ de _____ de 20____	
Assinatura do doador	Impressão Digital
TERMO DE CONHECIMENTO DA INAPTIÇÃO Declaro estar ciente de não estar apto(a) à doação de sangue segundo os critérios e normas vigentes do Ministério da Saúde e da Fundação HEMOPE por um período de _____	
Data	Assinatura do Doador
	Impressão Digital

Frente	
FICHA DE DOAÇÃO DE SANGUE DIVISÃO DE COLETA - DCPOS	
Código Triagem Local de Coleta <input type="radio"/> Interna <input type="radio"/> Unidade Móvel <input type="radio"/> Específica <input type="radio"/> <input type="radio"/> Posto Fixo <input type="radio"/> Empresa (E-A) <input type="radio"/> Voluntário <input type="radio"/> Sexo () M () F	Tipo de Doação <input checked="" type="radio"/> Resposta <input type="radio"/> Específica <input type="radio"/> <input checked="" type="radio"/> Ausotega <input type="radio"/> Voluntário <input type="radio"/> Tipo Documento Órgão Expedidor
Local Nascimento Nº Documento Tipo Documento Órgão Expedidor	Sexo () M () F
Nome do Pai Nome da Mãe	Cor da Pele () A () B () M () N Branco
Estado Civil J () C () D () S () V () Endereço Residencial	CEP Telefone
Município Endereço Comercial (Empresa, Telefone e Ramal)	CEP Telefone
Ocupação Escolaridade () 1º Grau incompleto () 2º Grau incompleto () 3º Grau incompleto () 1º Grau completo () 2º Grau completo () 3º Grau completo	Registro Internação Cód. do Hospital
Hospital Estado Civil - Casado, D - Divorçado, S - Solteiro, V - Viúvo Sr da Mãe - A - Amarela, B - Branca, M - Mulata, N - Negra	
Hemoglobina Hematócrito Temperatura Pressão Arterial Peso Altura Pulso	Assinatura Assinatura Assinatura Assinatura
Opção de braço () D () E Hora início da coleta Hora término da coleta Respostas adversas	Assinatura Assinatura Assinatura
Código Bolete Código de Amostra	
OBSERVAÇÕES:	

ANEXO 2. DECLARAÇÃO DO COMITÊ DE ÉTICA



FUNDAÇÃO HEMOPE – ASSESSORIA DE PESQUISA
Rua Joaquim Nabuco, 171
CEP: 52011-000 – Graças – Recife -PE
Fones: (81) 3416-4660
Fone/Fax: (81) 3421-6946
E-mail: pesquisa@hemope.pe.gov.br

COMITÊ DE ÉTICA EM PESQUISA

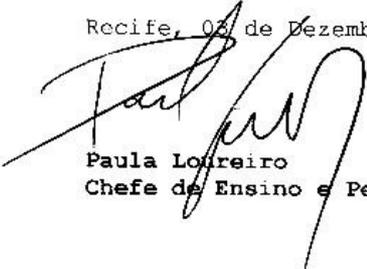
Projeto de Pesquisa: *"Regressão Logística uma abordagem bayesiana na determinação de fatores de risco para HIV, em doadores de sangue.*

Pesquisadora/Orientadora: *Eufrázio de Souza Santos*

Instituição: Fundação HEMOPE

O Comitê de Ética em Pesquisa da Fundação HEMOPE, de acordo com o previsto na resolução CSN nº 196/96, aprova o projeto acima citado, desenvolvido pela Bolsista **Niédja Maristone de Oliveira Barreto Queiroz**.

Recife, 03 de Dezembro de 2003


Paula Loureiro
Chefe de Ensino e Pesquisa

ANEXO 3. PROGRAMA REGRESSÃO LOGÍSTICA MÚLTIPLA NO WINBUGS

The screenshot displays the WinBUGS interface with the following components:

- Main Window:** Contains the model script for a multi-class logistic regression. The script defines parameters for HIV status based on age, education, and other variables.
- Update Tool:** A floating window with controls for 'updates' (100), 'refresh' (100), 'update' (1), 'iteration' (1400), 'thin' (1), 'over relax' (checkbox), and 'adapting' (checkbox).
- Sample Monitor Tool:** A floating window for monitoring the MCMC process. It includes fields for 'node', 'beg', 'end', 'chains', 'thin', and 'percentiles'. Buttons for 'clear', 'set', 'trace', 'history', 'density', 'stats', 'coda', 'quantiles', 'bgr diag', and 'auto cor' are also present.
- Specification Tool:** A floating window for model specification. It includes buttons for 'check model', 'load data', 'compile', 'load inits', 'for chain', and 'gen inits'. A 'num of chains' field is set to 1.

```

model
{
  for (i in 1:N) {
    HIV[i] ~ dbern(p[i])
    logit(p[i]) <- beta0 + beta1*idade101[i] + beta2*idade102[i] + beta3*idade103[i] +
    beta4*Escolar1[i] + beta5*Escolar2[i] +
    beta6*Escolar3[i]+beta7*Ant_Hbc[i]+beta8*Sifilis[i]+beta9*Reg_D
    icol[i]+beta10*tipdoa1[i]+beta11*tipdoa2[i]
  }
  beta0 ~ dnorm(0.0,0.001)
  beta1 ~ dnorm(0.0,0.001)
  beta2 ~ dnorm(0.0,0.001)
  beta3 ~ dnorm(0.0,0.001)
  beta4 ~ dnorm(0.0,0.001)
  beta5 ~ dnorm(0.0,0.001)
  beta6 ~ dnorm(0.0,0.001)
  beta7 ~ dnorm(0.0,0.001)
  beta8 ~ dnorm(0.0,0.001)
  beta9 ~ dnorm(0.0,0.001)
  beta10 ~ dnorm(0.0,0.001)
  beta11 ~ dnorm(0.0,0.001)
}

```

Data → click on one of the arrows to open the data ←

Inits ← list(beta0=0, beta1=0, beta2=0, beta3=0, beta4=0, beta5=0, beta6=0, beta7=0, beta8=0, beta9=0, beta10=0, beta11=0)←