



**UNIVERSIDADE FEDERAL DA PARAÍBA
CCEN - CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELOS DE DECISÃO E SAÚDE**

RAQUEL DE NEGREIROS MOREIRA

**QUALIFICAÇÃO E IMPUTAÇÃO DE DADOS SOBRE SATISFAÇÃO DE
HIPERTENSOS CADASTRADOS NA ESTRATÉGIA SAÚDE DA FAMÍLIA**

**João Pessoa, PB
2012**

RAQUEL DE NEGREIROS MOREIRA

**QUALIFICAÇÃO E IMPUTAÇÃO DE DADOS SOBRE SATISFAÇÃO DE
HIPERTENSOS CADASTRADOS NA ESTRATÉGIA SAÚDE DA FAMÍLIA**

Dissertação apresentada como requisito para a obtenção do grau de Mestre em Modelos de Decisão e Saúde da Universidade Federal da Paraíba.

Orientadores:

Prof. PhD. Neir Antunes Paes

Prof. Dr. Rodrigo Pinheiro de Toledo Vianna

João Pessoa, PB
2012

M838q *Moreira, Raquel de Negreiros.*

**Qualificação e imputação de dados sobre satisfação de hipertensos cadastrados na estratégia saúde da família/ Raquel de Negreiros
Moreira. - - João Pessoa: [s.n.], 2012.**

112f. : il.

*Orientadores: Neir Antunes Paes e Rodrigo Pinheiro e
Rodrigo Pinheiro de Toledo Vianna.*

Dissertação (Mestrado)-UFPB/CCEN.

*1. Saúde pública. 2. Hipertensão. 3. Hiperdia. 4.
Qualidade de dados. 5. Falta de dados. 6. Imputação.*

FOLHA DE APROVAÇÃO

Raquel de Negreiros Moreira

QUALIFICAÇÃO E IMPUTAÇÃO DE DADOS SOBRE SATISFAÇÃO DE HIPERTENSOS CADASTRADOS NA ESTRATÉGIA SAÚDE DA FAMÍLIA.

Dissertação apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde da Universidade Federal da Paraíba.

Aprovada em: 24/02/2012.

Prof. Dr. Neir Antunes Paes (orientador)
Instituição: UFPB

Prof. Dr. Rodrigo Pinheiro de Toledo Vianna (orientador)
Instituição: UFPB

Prof. Dr. César Cavalcanti Da Silva
Instituição: UFPB

Prof.^a Dra. Eduarda Ângela Pessoa Cesse
Instituição: Centro de Pesquisa Aggeu de Magalhães

Prof. Dr. Joab De Oliveira Lima
Instituição: UFPB

Prof. Dr. Mardone Cavalcante França
Instituição: UFRN

João Pessoa, PB
2012

Dedico a Deus e a Mãe Santíssima, pela força que sempre me deram nos momentos mais difíceis de minha vida. A meus pais pela paciência e apoio dado, e por toda a força e carinho, aos meus filhos que sentiram tanto a minha ausência e pelos sorrisos que me levantavam em meio às dificuldades.

AGRADECIMENTOS

Ao encerrar os trabalhos desse estudo, tenho o prazer de registrar alguns agradecimentos especiais a todos que, de forma direta ou indireta, acabaram por auxiliar, influenciar, opinar e corrigir o presente trabalho durante todo o transcorrer do curso. Meu agradecimento e carinho por cada um são enormes, mesmo não conseguindo me expressar tão bem com as palavras.

Ao Pai Eterno e a minha protetora Virgem Maria, que caminharam do meu lado, iluminando os caminhos a prosseguir, e que me deram toda a felicidade que tenho.

Aos meus familiares, pelo respeito, ajuda e compreensão nos dias difíceis. A minha avó D. Lourdes (in memorian) que tinha tanto orgulho de mim.

Aos meus filhos e ao meu esposo Neto, pelas tantas horas que deixei de me dedicar a eles na qualidade de mãe e esposa. Obrigada pela paciência, carinho e dedicação.

À mainha, que com tanta sabedoria me ensinou tudo que eu sei nessa vida, inclusive a ter sempre fé, porque Deus está acima de tudo...Sem você, mãezinha, eu não estaria aqui.

Ao meu pai, que além de me aconselhar como pai amoroso, me fez crescer com pessoa e como profissional, me ajudando a refletir e ser uma “pesquisadora”. Espelhamos-nos em você painho!

Aos meus irmãos Raoni e Itaquê “Tezinho”, tantos momentos passamos juntos, bons e ruins, sempre com frases engraçadas, gratidão, alegria e muito trabalho. O sucesso é nosso!

A todos os meus colegas de mestrado cujos desafios enfrentamos juntos, resultando no sucesso dessa caminhada.

Ao amigo e orientador (um segundo pai), Prof. Dr. Neir, pela paciência, persistência, assistência e estímulos contínuos no desenrolar desse estudo, cuja sabedoria e inteligência são gigantes. Obrigada por ter acreditado em mim sempre mostrando que cada um tem seu potencial!

Ao Prof. Dr. Joab, cuja competência, disponibilidade, companheirismo e dedicação foram únicos. Obrigada pela ajuda imprescindível, sem esta jamais teria concluído tal trabalho.

Aos professores do Departamento de Estatística e do Programa de Pós-Graduação: Ulisses, Eufrásio, César, Rodrigo, Marcelo, Jozemar, pela enorme ajuda nesses anos de estudo. Jamais esquecerei tais ensinamentos...

Aos membros da Banca Examinadora, competentes profissionais da ciência, que se propuseram a contribuir no aprimoramento deste estudo.

*“Pesquisar é ver o que todos vêem
e pensar o que ninguém mais
pensou ”*
Albert Szent-Gyorgyi

RESUMO

MOREIRA, Raquel de Negreiros. **Qualificação e imputação de dados sobre satisfação de hipertensos cadastrados na Estratégia Saúde da Família.** João Pessoa, 2012. Dissertação (Mestrado em Modelos de Decisão em Saúde), Universidade Federal da Paraíba - UFPB. João Pessoa.

A qualidade das informações tem sido objeto de interesse particularmente na área da saúde. Sabe-se que a incompletude de informações é um problema muito comum nos sistemas de informação e em estudos epidemiológicos. Desta forma, tem-se como solução a imputação de dados, onde são criados conjunto de dados artificialmente completos passíveis de análise estatística. Esse estudo objetivou analisar a qualidade dos dados do HIPERDIA e dos itens sobre satisfação de usuários hipertensos da Estratégia Saúde da Família no município de João Pessoa/PB sobre o serviço e o uso de métodos de imputação para dados faltantes. Os dados secundários foram obtidos da segunda via do HIPERDIA, dos hipertensos cadastrados entre 2006/2007 em 36 equipes de Saúde da Família, resultando numa amostra representativa de 343 usuários do município de João Pessoa/PB. Como fonte primária foi construído um instrumento composto por 8 dimensões essenciais da atenção básica, mensurados na Escala de Likert variando de "0" a "5". As técnicas foram aplicadas para o método de Imputação Única: Substituição por um Valor de Tendência Central (TC), *Hot Deck*, Estimativa de Máxima Verossimilhança (MV) e Regressão Logística Multinomial (RLM), sendo comparados através do percentual de acerto, erro médio quadrado (RMSE) e erro percentual médio absoluto (MAPE). Foi construída a simulação de dois cenários amostrais distintos com diferentes proporções de dados faltantes (5%, 10%, 15%, 30% e 40%). Na comparação dos métodos de imputação, para cenário com variável apresentando um tipo de resposta sobrepondo às outras, o método de TC foi o que obteve melhor performance, seguido do método de RLM. Para o cenário com homogeneidade de frequência de respostas, o melhor método foi o de RLM. O estudo permitiu demonstrar que ainda existem falhas no preenchimento do HIPERDIA e que a imputação permitiu resgatar as características da representação dos dados originais, verificando que os métodos de imputação adotados trouxeram fidedignidade e diminuição de vieses na amostra para proporções de até 40% de dados faltantes.

PALAVRAS CHAVES: hipertensão, HIPERDIA, qualidade dos dados, falta de dados, imputação.

ABSTRACT

MOREIRA, Raquel de Negreiros. **Eligibility and imputation of data on satisfaction of hypertensive registered in the Family Health Strategy**. João Pessoa, 2012. 115f. Dissertation (MSc in Decision Models in Health), Universidade Federal da Paraíba - UFPB. João Pessoa.

The quality of information has been of particular interest in health. It is known that the incompleteness of information is a very common problem in information systems and epidemiological studies. Thus, it has been imputation as a solution of the missing data, which are created artificially complete set of data subject to statistical analysis. This study aimed to analyze the data quality HIPERDIA items on satisfaction and hypertensive patients of the Family Health Strategy in the city of João Pessoa / PB on the service and the use of imputation methods for missing data. Secondary data were obtained from duplicate HIPERDIA, the hypertensive patients enrolled between 2006/2007 in 36 family health teams, resulting in a representative sample of 343 users in the city of João Pessoa / PB. As a primary source was constructed an instrument consisting of eight core dimensions of primary care, measured on Likert scale ranging from "0" to "5". The techniques were applied to the method of Single Imputation: Replacement for Central Value Trend (TC), Hot Deck, Estimated Maximum Likelihood (MV) and Multinomial Logistic Regression (RLM), were compared using the percentage of correct answers, average error square (RMSE) and mean absolute percentage error (MAPE). Was built to simulate two different scenarios sample with different proportions of missing data (5%, 10%, 15%, 30% and 40%). The comparison of the allocation methods, for variable setting with a type having overlapping response to the other, the method was that TC gave better performance, followed by the method of RLM. For the scenario with homogeneous frequency response, the best method was to RLM. The study has demonstrated that there are still errors in the completion of HIPERDIA and that allowed us to recover the imputation characteristics of the representation of the original data, verifying that the imputation methods adopted brought reliability and reduction of bias in the sample proportions of up to 40% of missing data.

KEYWORDS: hypertension, HIPERDIA, data quality, missing data, imputation.

LISTA DE FIGURAS

FIGURA 3.1	Esquema do Processo de Simulação para o Cenário 1.....	60
FIGURA 3.2	Esquema do Processo de Simulação para o Cenário 2.....	62
FIGURA 4.1	Avaliação dos métodos de imputação com relação à eficiência e à estabilidade.....	76
FIGURA 4.2	Distribuição amostral do RMSE por método e cenários, segundo as proporções de valores faltantes	82
FIGURA 4.3	Distribuição amostral do MAPE por método e cenários, segundo as proporções de valores faltantes.....	83
FIGURA 4.4	Distribuição amostral do Percentual de Acerto por método e cenários, segundo as proporções de valores faltantes	84
FIGURA 5.1	Decisão para escolha de métodos de imputação em amostras ou variáveis com presença de frequência de tipo de resposta superior às demais.....	91
FIGURA 5.2	Decisão para escolha de métodos de imputação em amostras ou variáveis com presença de frequência com tipo de respostas homogêneas....	92

LISTA DE QUADROS

QUADRO 2.1	Classificação do risco estratificado segundo o grau de hipertensão e presença de fatores de risco ou doença.....	26
-------------------	--	----

LISTA DE TABELAS

TABELA 3.1	Análise descritiva do item F4.....	59
TABELA 3.2	Análise descritiva do item simulado para construção do Cenário 2.....	61
TABELA 4.1	Compleitude no preenchimento das variáveis do cadastro HIPERDIA em João Pessoa/PB, 2006/2007.....	68
TABELA 4.2	Compleitude no preenchimento das variáveis no prontuário dos usuários em João Pessoa/PB, 2009.....	69
TABELA 4.3	Informações descritivas dos usuários com pelo menos um dado ausente no cadastro HIPERDIA em João Pessoa/PB, 2006/2007.....	70
TABELA 4.4	Compleitude de dimensões do estudo no município de João Pessoa/PB, 2009.....	72/73
TABELA 4.5	Médias e desvios padrão para os erros de imputação (RMSE e MAPE) e para o percentual de acerto segundo o método e proporção de valores faltantes para o Cenário 1.....	77
TABELA 4.6	Médias e desvios padrão para os erros de imputação (RMSE e MAPE) e para o percentual de acerto por método e proporção segundo os valores faltantes para o Cenário 2.....	80
TABELA 4.7	Porcentagem das frequências das respostas aos itens das dimensões características ao Cenário 1	85
TABELA 4.8	Porcentagem das frequências das respostas aos itens das dimensões característicos ao Cenário 2	86

LISTA DE SIGLAS

ACS – Agente Comunitário de Saúde
AVC – Acidente Cardiovascular
CONASS – Conselho Nacional dos Secretariados de Saúde
CNES - Cadastro Nacional de Estabelecimentos de Saúde
CNS - Cartão Nacional de Saúde
CTIIS - Câmara Técnica de Informação e Informática
DATASUS - Departamento de Informática do SUS
DM – Diabetes Mellitus
EM - Expectation-Maximization
ESF – Estratégia de Saúde da Família
HAS – Hipertensão Arterial Sistêmica
HD – Hot Deck
IM – Imputação Múltipla
OMS- Organização Mundial da Saúde
PACS – Programa de Agentes Comunitários de Saúde
PAD – Pressão Arterial Diastólica
PAS – Pressão Arterial Sistólica
PCAT - *Primary Care Assessment Tool*
PNE- CHE – Plano Nacional de Educação e Controle da Hipertensão Arterial
PRAHADM – Plano de Reorganização da Atenção à Hipertensão Arterial e ao Diabetes Mellitus
PSF – Programa de Saúde da Família
MAPE – Erro Percentual Médio Absoluto
MAR - (Missing at Random
MCAR- *Missing Complety at Random*
MS – Ministério da Saúde
MV – Máxima Verossimilhança
NOAS - Norma Operacional Básica de Assistência à Saúde
QME – Quadrado Médio do Erro
RMSE - Rootmean-Square Deviation
RLM – Regressão Logística Multinomial
SIA - Sistema de Informação Ambulatorial

SIAB – Sistema de Informação da Atenção Básica

SIA-SUS - Sistema de Informações Ambulatoriais

SIH - Sistema de Informações Hospitalares

SI - Sistema de informação

SIM - Sistema de Informação de Mortalidade

SISPACS - Sistema de Informação do Programa de Agentes Comunitários de Saúde

SIS – Sistema de Informação em Saúde

SINAM - Sistema de Agravos de Notificação

SINASC - Sistema de Informação de Nascidos Vivos

SisHiperdia - Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos

SUS – Sistema Único de Saúde

TC – Tendência Central

UBS – Unidade Básica de Saúde

USF – Unidade de Saúde da Família

SUMÁRIO

APRESENTAÇÃO.....	I
CAPÍTULO 1: INTRODUÇÃO.....	16
1.1 PROBLEMATIZAÇÃO.....	20
1.2 JUSTIFICATIVA.....	22
1.3 OBJETIVOS.....	24
1.3.1 Geral.....	24
1.3.2 Específicos.....	24
CAPÍTULO 2: REFERENCIAL TEÓRICO.....	25
2.1 HIPERTENSÃO ARTERIAL SISTÊMICA.....	26
2.2 SAÚDE PÚBLICA E SISTEMAS DE INFORMAÇÃO DIRECIONADOS AOS USUÁRIOS HIPERTENSOS.....	28
2.2.1 Sistema Único de Saúde e Estratégia Saúde da Família.....	28
2.2.2 Sistema de Informação em Saúde.....	31
2.2.3 Sistema de informação HIPERDIA: origem e Contemporaneidade.....	33
2.3 INCOMPLETUDE DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE E EM PESQUISAS EPIDEMIOLÓGICAS.....	37
2.4 SOLUÇÃO PARA VARIÁVEIS COM DADOS FALTANTES: IMPUTAÇÃO DE DADOS	38
2.4.1 Mecanismos de Informações Perdidas.....	39
2.4.2 Imputação Única.....	41
2.4.3 Imputação Múltipla.....	43
2.5 PROJETO SOBRE A HIPERTENSÃO ARTERIAL SISTÊMICA (PAES, 2008;2009).....	45
CAPÍTULO 3: MATERIAIS E MÉTODOS.....	47
3.1 FONTES E BASE DOS DADOS.....	48
3.2 PERFIL DESCRITIVO DOS USUÁRIOS COM DADOS FALTANTES NO HIPERDIA.....	51
3.3 MÉTODOS DE IMPUTAÇÃO.....	52
3.3.1 Método 1: Substituição por um Valor de Tendência Central	52

3.3.2 Método 2: <i>Hot deck</i>	52
3.3.3 Método 3: Estimativa de Máxima Verossimilhança.....	53
3.3.4 Método 4: Regressão Logística Multinomial	55
3.4 ESTUDO DE SIMULAÇÃO.....	56
3.4.1 Simulação para o Cenário 1.....	58
3.4.2 Simulação para o Cenário 2.....	61
3.5 ACURÁCIA DOS MÉTODOS.....	62
3.6 CRITÉRIOS DE DECISÃO.....	64
3.7 ORGANIZAÇÃO DO BANCO DE DADOS.....	64
3.8 CONSIDERAÇÕES ÉTICAS DA PESQUISA.....	65
CAPÍTULO 4: RESULTADOS	66
5.1 CENÁRIO DOS DADOS PERDIDOS NO HIPERDIA.....	67
5.2 CENÁRIO DOS DADOS PERDIDOS NAS DIMENSÕES DO QUESTIONÁRIO.....	71
5.3 SIMULAÇÃO E ACURÁCIA DOS MÉTODOS.....	74
CAPÍTULO 5: DISCUSSÃO	87
CAPÍTULO 6: CONSIDERAÇÕES FINAIS	93
REFERÊNCIAS	96
ANEXOS	105
ANEXO 1: FICHA DO CADASTRO DO HIPERDIA.....	106
ANEXO 2: QUESTIONÁRIO.....	107
ANEXO 3: APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA DO CCS.....	110
ANEXO 4: ENCAMINHAMENTO DA PREFEITURA MUNICIPAL DE JOÃO PESSOA.....	111

APRESENTAÇÃO

A qualidade dos dados dos sistemas de informação em saúde vem sendo amplamente discutida no Brasil. A falta de dados nos cadastros que compõe esses sistemas é um problema muito comum. A ausência de dados em pesquisas também vem sendo discutida. A imputação de dados surgiu com o propósito de tentar diminuir tais danos. Nesse sentido, esse estudo tenta abordar, como tema central, o problema de dados incompletos, seja nos sistemas de informação, quanto em banco de dados.

Tal estudo é uma ramificação de um projeto de maior abrangência (PAES, 2008/2009) que objetivou avaliar a efetividade do controle pressórico dos usuários cadastrados no HIPERDIA no período de 2006 a 2007 e com acompanhamento entre 2009 a 2010.

Diante do problema de falta de dados nos cadastros do HIPERDIA e no banco de dados da pesquisa é que surgiu a ideia de abordar tais paradigmas e suas possíveis soluções.

Têm-se assim o intuito de verificar a incompletude dos dados do HIPERDIA e a utilização de técnicas de imputação para construção de bancos de dados completos.

Espera-se que este estudo possa subsidiar os debates necessários para que, em breve, a qualidade das informações oriundas dos nossos sistemas e das pesquisas esteja garantida, possibilitando que sejam amplamente utilizadas para tomada de decisão.

Capítulo 1

1. INTRODUÇÃO

Ao se valorizar o emprego da epidemiologia para a avaliação dos sistemas e serviços de saúde, reconhece-se o papel desta na produção de conhecimentos para o processo de tomada de decisão, no que se refere à formulação de políticas de saúde, à organização do sistema e às intervenções necessárias para o enfrentamento de problemas específicos nessa área de produção do conhecimento (TIGRE et al, 1990).

Nos países em desenvolvimento, as condições para o adoecimento por patologias crônicas estão relacionadas ao nível de atenção primária e devem ser tratadas principalmente nesse âmbito. No entanto, grande parte da atenção primária está voltada para problemas agudos e para as necessidades mais urgentes dos pacientes (OMS, 2003).

Este é um dos desafios da Atenção Básica, notadamente da Estratégia Saúde da Família, espaço prioritário e privilegiado de atenção à saúde que atua com equipe multiprofissional e cujo processo de trabalho pressupõe vínculo com a comunidade e a clientela adscrita, levando em conta a diversidade racial, cultural, religiosa e os fatores sociais envolvidos (GIACOMOZZI; LACERDA, 2006).

Uma das patologias crônicas que exigem acompanhamento de maior regularidade, com avaliações rotineiras presentes na Estratégia Saúde da Família é a Hipertensão Arterial Sistêmica (HAS).

Pelo fato de a hipertensão arterial não acarretar, na maioria das vezes, qualquer sintoma aos pacientes e por envolver orientação para vários objetivos, o sucesso no seu tratamento depende de uma ação integrada multiprofissional, fortalecendo um vínculo dos pacientes à unidade e à organização de rede de atendimento (BRASIL, 2001).

A hipertensão é uma doença que deve ser prevenida e tratada nas unidades de saúde. Estas são responsáveis por uma determinada área, onde se acompanha e geram informações para a aquisição e distribuição de medicamentos de forma regular a todos os pacientes cadastrados pelo instrumento elaborado para tal (CHAZAN; PEREZ, 2008).

O reconhecimento de que a modificação dos hábitos de vida, como a prevenção dos fatores de risco e o tratamento adequado de desvios da normalidade

quando estabelecidos modificam a história evolutiva desses agravos e torna ainda mais estratégico o conhecimento de sua prevalência para sua prevenção e/ou tratamento (CONTIERO et al, 2009).

Na Atenção Básica, há um sistema de informação que também cadastra, acompanha e monitora os usuários com hipertensão arterial e *Diabetes Mellitus* conforme as diretrizes do Plano de Reorganização da Atenção a Hipertensão Arterial e *Diabetes Mellitus* (BRASIL, 2001). Lançado no ano de 2000, este plano teve o intuito de minimizar os problemas crescentes de doenças crônicas não transmissíveis, principalmente a hipertensão e diabetes. Diversas ações foram implementadas nos estados e municípios, como capacitações profissionais na atenção básica, pactuação de normas e metas entre as três esferas da gestão de saúde, atenção à assistência farmacêutica com dispensação de medicamentos de uso contínuo, e promoção de atividades educativas (OLIVEIRA, 2005).

O Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos (SisHiperdia) foi desenvolvido em 2002 em consequência desse Plano de Reorganização, permitindo cadastrar e acompanhar os portadores de Hipertensão Arterial Sistêmica (HAS) e/ou *Diabetes Mellitus* (DM), em todas as Unidades Básicas do SUS; possuindo também o propósito de gerar informações para os trabalhadores e gestores de saúde, para aquisição, dispensação e distribuição de medicamentos a todos os usuários cadastrados (CONTIERO, et al, 2009).

É notório que a informação consistir em uma importante ferramenta para gestores e planejadores de saúde na formulação e avaliação de políticas públicas, assim, por meio desse sistema informatizado, tornou-se possível a descrição do perfil epidemiológico dos pacientes cadastrados. Tal potencialidade, no entanto, pode ser mais explorada no país e agregada ainda aos indicadores de controle de hipertensão presentes no Pacto de Indicadores da Atenção Básica (BOING, BOING, 2007).

O HIPERDIA é disponibilizado pelo DATASUS, sendo de uso obrigatório para os municípios que aderiram à Estratégia Saúde da Família. Esse programa é “alimentado” pelos dados de consulta médica e de enfermagem, com atuação do agente comunitário de saúde com um acompanhamento sistemático do usuário hipertenso pela equipe de Saúde da Família (ZILMER; SCHWARTZ; MUNIZ, 2010).

Avaliar a qualidade das informações em saúde, nos Sistemas de Informação de Saúde (SIS), é imprescindível, na medida em que eles são instrumentos

importantes para o diagnóstico da situação de saúde, pois caracterizam populações em risco e possibilitam planejar estratégias terapêuticas de acordo com as necessidades e especificidades de cada grupo populacional (ZILMER; SCHWARTZ; MUNIZ, 2010). A qualidade da informação depende da coleta adequada, da correta estruturação e da correta interpretação (fidedignidade) dos dados (TEIXEIRA, et al, 2005).

Zilmer, Schwartz e Muniz, (2010), abalizam a incompletude nas variáveis do HIPERDIA, mostrando limitações na qualidade dos dados do cadastro, podendo interferir na análise final dos dados da população com Hipertensão Arterial Sistêmica.

O presente estudo faz parte de um projeto desenvolvido por Paes (2008), cujo estudo intitula-se “Avaliação da efetividade no controle da hipertensão arterial sistêmica e associação com fatores de risco comparando a atenção do Programa de Saúde da Família e de Unidades Básicas de Saúde de municípios do Nordeste do Brasil”. Este teve como objetivo principal avaliar o impacto do acompanhamento e efetividade do controle da hipertensão arterial sistêmica (HAS), segundo a redução dos níveis de pressão arterial em indivíduos hipertensos, submetidos a ações programáticas para controle da doença. Tal projeto foi estendido a um estudo de coorte compreendendo os anos de 2009, 2010 e 2011 (PAES, 2009).

Utilizou-se nesse estudo (PAES, 2008; PAES 2009) um instrumento componente do *Primary Care Assessment Tool* (PCAT), desenvolvido na Universidade de *Johns Hopkins* e validado para avaliar os aspectos críticos da atenção primária em países industrializados (STARFIELD, 2002). O instrumento foi adaptado e revalidado para o Brasil por Almeida e Macinko (2006) para a tuberculose e, na sequência ele foi readequado por PAES (2008) para a utilização em indivíduos portadores de hipertensão arterial, o qual manteve as oito dimensões, constituídos por itens respectivos, sugeridos originalmente por Starfield (2002).

Refletindo sobre a incompletude de variáveis no cadastro de hipertensos, a ocorrência de dados faltantes (*missing data*) é um problema comum em investigações científicas. Em pesquisas da área de saúde, é muito frequente que o pesquisador defronte-se com o problema de dados faltantes em sua coleta e uma abordagem bastante comum é restringir a análise aos sujeitos com dados completos nas variáveis envolvidas. Essa exclusão de sujeitos pode gerar inferências que não

são válidas, principalmente se os indivíduos que permanecem na análise são diferentes daqueles que foram excluídos (MOONS, SAIJNEN, HARRELL, 2006).

O manuseio apropriado dos valores em falta é importante em todas as análises críticas no processo de qualificação do banco de dados. A manipulação imprópria de valores ausentes poderá distorcer a análise. O objetivo de um procedimento estatístico deve ser fazer inferências válidas e eficientes para a população de interesse. Tentativa de recuperar dados perdidos sem um estudo adequado pode prejudicar a inferência. Há, portanto, a necessidade de preencher dados faltantes com valores plausíveis e então aplicar métodos tradicionais de análise de dados completos para fazer inferências válidas e eficientes.

Para contornar este problema, desde os anos 80, surgiram técnicas estatísticas que envolvem a substituição de dados faltantes por estimativas de valores plausíveis a serem “imputados” aos dados faltantes. Esta técnica denomina-se imputação de dados faltantes na literatura estatística e seu uso vem generalizando-se e estendendo-se a outras áreas (RUBIN, 1996; SCHAFER E GRAHAM, 2002; FRASER, 2007).

Durante os estudos de Paes (2008 2009), foi observado a falta de informações nos questionários aplicados aos hipertensos cadastrados no HIPERDIA. A falta desses dados foi encontrada tanto na área reservada às informações contidas no cadastro dos usuários, quanto nas dimensões referentes às respostas dos hipertensos sobre qualidade de saúde.

1.1 . PROBLEMATIZAÇÃO

A Organização Mundial da Saúde define Sistema de Informação em Saúde – SIS como um mecanismo de coleta, processamento, análise e transmissão da informação necessária para se planejar, organizar, operar e avaliar os serviços de saúde. Considera-se que a transformação de um dado em informação exige, além da análise, a divulgação, e inclusive recomendações para a ação.

A finalidade da informação em saúde é identificar problemas individuais e coletivos da população, servindo de base para análise da situação encontrada e

subsidiando a busca de alternativas para solucionar tais problemas (CHAZAN, PERES, 2008).

No novo modelo de assistência à saúde proposto pelo SUS, é preciso reverter a atual situação de centralização de dados, de limitação do uso dos mesmos, da demora com que são analisados e que retornam para o nível local, idealizando um novo sistema em que os dados passem a ser analisados no próprio município, gerando de forma oportuna subsídios para o planejamento e para as ações em saúde, bem como de ações para a melhoria da qualidade dos dados.

Apesar da importância das informações geradas por esses sistemas, observa-se, de maneira geral, que elas são muito pouco utilizadas no processo de decisão e controle (LIMA; CARVALHO; SCHRAMM, 2006).

Segundo vários estudos (ZILMER et al, 2010) há uma incompletude dos dados no sistema de informação do HIPERDIA. Essa incompletude é em sua maioria pela falta de dados de identificação dos usuários. Nesse sentido, a falta dos documentos de identificação pode causar transtornos que comprometam o processamento do cadastro (MARUITI, BAULI, SCOCHI, 2008).

Para Zilmer et al (2010), alguns fatores parecem ser responsáveis por essa falta de informações preenchidas, tais como: falta de capacitação dos profissionais das equipes de saúde para utilização dos cadastros e fichas de acompanhamento do Sistema HIPERDIA, a dificuldade de utilizar os dados gerados no seu dia a dia para o planejamento de ações em saúde a nível local; e a insuficiência de recursos humanos para o desenvolvimento das funções de forma satisfatória (FRIAS, et al, 2008; FACCHINE, et al, 2005; BARROS, 2006).

Somado a isso, as Unidades de Saúde possuem um sistema precário de informações, que não permite a integração entre os diversos sistemas (MEDEIROS, 2005), fazendo com que tenham diversos cadastros a serem preenchidos, além do déficit em profissionais necessários para a digitação dos dados e retroalimentação do sistema (FRIAS, et al, 2008). Em concordância com outros estudos (CUNHA, 2002; CHAZAN, PEREZ, 2008), estes revelam a carência, na área da saúde, de recursos humanos com treinamento em informática e a dificuldade de acesso à infraestrutura necessária para o uso dessa tecnologia.

Nesse sentido, sabe-se que a falta de dados é um problema nas pesquisas científicas, podendo interferir na análise final da situação de saúde da população.

O impacto da falta de dados prejudica não só pela sua influencia nos resultados, mas também pelo seu impacto prático no tamanho da amostra disponível para análise. Se, por exemplo, a substituição dos dados perdidos não é aplicada, qualquer observação com essa ausência, em qualquer das variáveis, será excluída da análise (HAIR, 2009).

Existem questionamentos inevitáveis com relação aos dados perdidos: os indivíduos que não responderam são diferentes dos que responderam? Sendo assim, essas diferenças têm algum impacto na análise, nos resultados ou em sua interpretação? Preocupações semelhantes a essas também surgem das respostas perdidas para variáveis individuais. São tais questionamentos que direcionam o presente estudo, determinando no segundo passo a substituição das informações perdidas (HAIR, 2009).

1.2. JUSTIFICATIVA

Com o avanço da tecnologia, a demanda por aquisição, armazenamento e processamento de uma enorme quantidade de dados é cada vez maior. Com o crescimento contínuo da disponibilidade de dados de diferentes fontes, surgem questões desafiadoras relacionadas à qualidade dos dados. Um dos grandes problemas da maioria dos bancos de dados no mundo são a imprecisão e a incompletude, ou seja, a presença de dados com vieses e faltantes, respectivamente (PACHECO; CARDOSO, 2005). Há muitas razões que levam a este cenário, entre elas: mal funcionamento de equipamentos, alto custo da coleta de dados e participantes de uma pesquisa que se recusam a responder a certas questões. Em muitas áreas, não é raro encontrar conjuntos de dados que tenham 50% ou mais de dados faltantes (FARHANGEAR; KURGAN; PEDRYCZ, 2007).

Geralmente, três tipos de problemas estão associados aos dados faltantes: (i) perda de eficiência, (ii) restrições operacionais na manipulação e análise dos dados, e (iii) viés resultante das diferenças entre os valores atribuídos aos dados faltantes e os valores reais (FARHANGEAR; KURGAN; PEDRYCZ, 2007). Um tratamento inadequado dos dados faltantes também pode afetar a generalização dos resultados obtidos. Certamente, o melhor cenário seria evitar a ocorrência dos dados faltantes,

e algumas estratégias podem ser adotadas para promover a disponibilidade de dados, como o aumento dos benefícios aos participantes de uma pesquisa, ou repetir o experimento (JUNGER, 2008). No entanto, nem sempre é possível evitar os dados faltantes e, neste caso, não há outra forma de se obter resultados confiáveis, a não ser tratando tais ausências.

Ainda tratando da qualidade de dados, sabe-se que, o desafio que se coloca a um sistema de informações em saúde começa por reconhecer a importância da informação coletada, e que seja possível acessar os dados com facilidade; na medida da necessidade, integrar dados das diversas fontes; garantir a comparabilidade dos dados, ao longo do tempo e entre os diferentes inquéritos, e garantir a qualidade e a completude dos dados (BARROS, 2006).

É necessário avaliar a qualidade dos dados dos sistemas de informação, analisando a completude das variáveis. O HIPERDIA avalia as condições determinantes de saúde de usuários cadastrados na Estratégia Saúde da Família com Hipertensão Arterial Sistêmica e Diabetes. A falta de informações nesse sistema afeta diretamente as análises de comportamento, perfil e dados antropométricos desses usuários e de suas patologias. Dessa forma, é crucial investigar o grau de completude do HIPERDIA.

A carência de pesquisas desenvolvidas no sentido de verificar a qualidade de informações nos Sistemas de Informação em Saúde (VIDRO, 2004) instigou-se a realizar este estudo, para que seja possível identificar problemas no processo de coleta e registro dos dados do HIPERDIA no município de João Pessoa, Paraíba.

Diante da ausência de dados nas respostas dos usuários sobre sua satisfação do serviço, outro desafio que foi proposto é o de avaliar o melhor método de imputação de dados para o estudo, tentando descobrir qual a proporção limiar, ou máxima de informações ausentes para que se possam imputar dados e a fidedignidade das análises seja mantida. As pesquisas recentes mostram que o uso de imputação de dados está sendo difundida, mas ainda são escassos os estudos referentes à saúde, embora o problema de falta de dados seja bastante comum em estudos epidemiológicos e em saúde coletiva.

Dessa forma, o presente estudo tem a finalidade de analisar tais dados perdidos, tentando obter um cenário das informações do HIPERDIA, com a síntese do perfil do tipo de usuário mais prejudicado com a falta de informações, e, identificar as dimensões do questionário que tenham falta de dados utilizando um

método de imputação que tenha melhor desempenho para as informações obtidas no Projeto Paes (2008; 2009).

1.3. OBJETIVOS

1.3.1. Geral

Estudar métodos de imputação de dados sobre a satisfação de usuários hipertensos da Estratégia Saúde da Família do município de João Pessoa/PB.

1.3.2. Específicos:

- Avaliar a qualidade das informações dos hipertensos cadastrados no HIPERDIA;
- Realizar imputações de dados com cenários distintos de tipos de respostas sobre a satisfação dos usuários.

Capítulo 2

2. REFERENCIAL TEÓRICO

2.1. HIPERTENSÃO ARTERIAL SISTÊMICA

A Hipertensão Arterial Sistêmica é uma patologia de alta prevalência no mundo, causadora de sérias complicações cardiovasculares, tornando-se um grande problema de saúde pública, porque vem sendo responsável pela elevação dos índices de morbimortalidade, e pelos altos custos referentes ao tratamento tanto na rede privada quanto na pública e com isso tem suscitado discussões quanto à necessidade de rever as estratégias implementadas, para que se possa mudar esse quadro.

No Brasil, as doenças cardiovasculares aparecem como a primeira causa de morte, e a HAS é um dos mais importantes fatores de risco de desenvolvimento dessas cardiopatias (AMADEO, 2000; BRANDÃO, 2003; BRASIL, 2004; FUNASA, 2000).

Oliveira (2009) conceitua a Hipertensão Arterial Sistêmica (HAS) como “a situação clínica em que os níveis de pressão arterial sistêmica ficam acima do normal definido pela OMS, que é de 140 mmHg de Pressão Arterial Sistólica (PAS) e 80 mmHg de Pressão Arterial Diastólica (PAD)”.

O Quadro 2.1 demonstra os níveis e riscos para as mensurações de pressão arterial segundo a Organização Mundial de Saúde (OMS):

QUADRO 2.1: Classificação do risco estratificado segundo o grau de hipertensão e presença de fatores de risco ou doença.

Fatores de risco ou doença	Grau 1 Hipertensão Leve PAS 140-159 ou PAD 90-99	Grau 2 Hipertensão Moderada PAS 160-179 ou PAD 100-109	Grau 3 Hipertensão Grave PAS ≥180 ou PAD ≥110
I- Sem fatores de risco	Risco baixo	Risco médio	Risco alto
II- 1-2 fatores de risco	Risco médio	Risco médio	Risco muito alto
III- 3 ou mais fatores de risco ou lesões em órgãos-alvo ou diabetes	Risco alto	Risco alto	Risco muito alto
IV- Condições clínicas associadas, incluindo doença cardiovascular ou renal	Risco muito alto	Risco muito alto	Risco muito alto

Fonte: BRASIL, 2000

A hipertensão arterial sistêmica é uma doença crônica que, quando não tratada e controlada adequadamente, possibilita o aparecimento de complicações podendo atingir outros órgãos e sistemas: no Sistema Nervoso Central, podem ocorrer hemorragia e encefalopatia hipertensiva; no coração, pode ocorrer cardiopatia isquêmica (angina), insuficiência cardíaca, aumento do músculo cardíaco e, em alguns casos, morte súbita; nos pacientes com insuficiência renal crônica, associada sempre ocorre nefroesclerose; no sistema vascular, pode ocorrer entupimentos e obstruções das artérias carótidas, aneurisma de aorta e doença vascular periférica dos membros inferiores; e no sistema visual, há retinopatia que reduz a capacidade visual dos pacientes (CARVALHO, 1998).

Apesar da grande variedade e disponibilidade dos agentes anti-hipertensivos disponíveis para o tratamento da HAS, menos de 33% dos pacientes hipertensos adultos tem a sua pressão adequadamente controlada (ANDRADE et al, 2002)

Estudos recentes demonstram que dos novos pacientes com diagnóstico de HAS que iniciam a terapêutica, cerca de 16% a 50% descontinuem a medicação anti-hipertensiva durante o primeiro ano de uso e um número substancial daqueles que permanecem em uso da medicação o fazem de modo inadequado (CHRESTANI, 2008).

A hipertensão arterial sistêmica (HAS) é o principal fator de risco para doenças cardiovasculares, estando associada a 80% dos casos de acidente vascular cerebral (AVC) e a 60% dos casos de doenças isquêmicas do coração. Há evidências suficientes de que, em indivíduos com níveis pressóricos normais, o aumento de 20 mmHg na pressão sistólica e 10 mmHg na pressão diastólica duplica o risco de morte por AVC e infarto do miocárdio (ARAÚJO et al, 2008).

Segundo a Organização Mundial de Saúde (OMS), pelo menos 600 milhões de pessoas são hipertensas em todo o mundo. Trata-se de um fator de risco de alta prevalência afetando aproximadamente 35% dos brasileiros. No Brasil, as doenças cardiovasculares são responsáveis pelas principais causas das internações hospitalares e cerca de 30% de todos os óbitos (ARAÚJO et al, 2008).

Estudo realizado no Rio Grande do Sul, de base populacional, mostrou que um terço dos indivíduos com 20 anos ou mais de idade são hipertensos (ZILMER, 2009). A prevalência de HAS aumenta progressivamente com a idade e também está relacionada à cor da pele, história familiar de hipertensão e hábitos comportamentais (BOING, BOING, 2007).

Mudanças no estilo de vida, tais como, realização de atividade física, perda de peso, redução da ingestão de sal e álcool, além de dieta rica em frutas e vegetais e pobre em gorduras, contribuem para reduzir os níveis da pressão arterial.

A participação ativa do indivíduo é a única solução eficaz no controle da doença e na prevenção de suas complicações. Para que haja esse engajamento, é importante que haja vínculo suficiente entre profissionais de saúde e paciente (MANFROI; OLIVEIRA, 2006). Nos hospitais públicos brasileiros (nível terciário) e em muitos centros de especialidade (nível secundário), os pacientes acabam ficando vinculados ao serviço, e não a um médico ou profissional de saúde específico. Esse cenário tem mudado com o aumento da cobertura da Estratégia de Saúde da Família, que reorganiza a prática assistencial em novas bases e critérios, em substituição ao modelo tradicional de assistência (BRASIL, 2003). Por meio dessa abordagem, não só o paciente, mas toda a família e a comunidade estão vinculadas a uma equipe de saúde específica (ATLAS, et al, 2009).

2.2. SAÚDE PÚBLICA E SISTEMAS DE INFORMAÇÃO DIRECIONADOS AOS USUÁRIOS HIPERTENSOS

Desde a década de 1980, o país tem demonstrado uma maior preocupação em se tomar ações no controle da hipertensão. Segundo Luna (1999), em 1983, o MS publicou, pela primeira vez, um Guia para Controle da Hipertensão Arterial, iniciativa do Cardiologista Aloyzio Achutti. E em 1989, o MS criou o Programa Nacional de Educação e Controle da Hipertensão Arterial (PNE-CHE), ficando claro, dessa forma, que os cuidados com a HA crescem paralelamente com a criação do Sistema Único de Saúde (SUS).

2.2.1. Sistema Único de Saúde e Estratégia Saúde da Família

Foi colocado em debate no Brasil, em meados da década de 1980, a criação de um modelo de assistência à saúde voltada para toda a população de forma integral e universal com diversas características, destacando a responsabilidade do Estado de garantia dos serviços de saúde no país (CORDEIRO, 2004).

Da luta contra a Ditadura Militar e com o tema Saúde e Democracia, tem início no Brasil o Movimento da Reforma Sanitária, no final do ano de 1970, construindo-se a partir da participação das universidades, movimento sindical e experiências regionais de educação e saúde que envolveu toda sociedade civil (ESCOREL, 1998).

Essa Reforma foi direcionada com a 8ª Conferência Nacional de Saúde (CNS) que ocorreu em 1986, com a participação de milhares de representantes sociais para a discussão de um novo modelo de saúde para o país. Resultou-se então em emenda constitucional, nascida de um movimento social, o fato de que a saúde seria um direito do cidadão e dever do Estado e que fosse universal o acesso a todos os bens e serviços que a promovam e recuperem (AROUCA, 1998).

Todo este movimento repercutiu na promulgação da nova Constituição Federal, em outubro de 1988, que incluiu um capítulo dedicado à saúde determinando que a mesma fosse um direito de todos e dever do Estado. Nela ainda foi citado o acesso igualitário e universal às ações e serviços de saúde, além de diretrizes para um novo sistema de saúde (AROUCA, 1998).

Deste modo, foi criado o Sistema Único de Saúde (SUS), que se estabelece como sistema de saúde do país em 1990, através das Leis 8.080 e 8.142/90. A primeira dispendo sobre a organização dos serviços sob os princípios da universalização e integralidade, além das diretrizes norteadoras como a hierarquização, descentralização e regionalização da assistência. A segunda, por sua vez, apresenta as formas de financiamento das ações de saúde e a participação da comunidade (GIACOMOZZI; LACERDA, 2006).

O Sistema Único de Saúde (SUS) é um modelo progressista, que prevê ampliação quantitativa e qualitativa de ação estatal mediante mecanismos políticos e administrativos considerados inovadores, representando a primeira experiência concreta da Reforma Sanitária brasileira que conseguiu romper com a lógica cooperativista e patrimonialista, tradicionalmente mantida em nossa história, e internalizar uma linha de reconhecimento do direito de cidadania (CONASS, 2007).

Segundo esses fatores, o SUS se apresenta como uma proposta de democratização da saúde no país conquistada na Constituição Federal de 1988, através da qual, saúde passa a ser entendida como resultado das condições de vida das pessoas. É uma conquista de responsabilidade do governo e da sociedade através de suas políticas econômicas e sociais. Ao mesmo tempo, representa um

direito humano e de cidadania visto que o Estado deve assegurar o acesso aos bens e serviços (BRASIL, 2007).

O SUS é formado por um conjunto de ações e serviços de saúde sob gestão pública. Esse sistema está organizado em redes regionalizadas e hierarquizadas e atua em todo o território nacional. Possui direção única em cada esfera de governo, é formado pela rede de serviços públicos de saúde dos três níveis federativos - União, Estados e Municípios - e pelo setor filantrópico e privado contratado, sendo este último de forma complementar. É único por ter uma filosofia e sistemática comum a todos entes federativos, devendo atender a todos de maneira integral; descentralizado, racional (de acordo com as necessidades da população) e democrático (ter a participação de todos). É um processo que está sempre em aperfeiçoamento e adaptação (CONASS, 2007).

Para operacionalizar os princípios e as diretrizes do SUS surge em 1994, o Programa Saúde da Família (PSF), que atualmente tem sido denominado Estratégia Saúde da Família (ESF), uma vez que não possui caráter programático, e sim características estratégicas de mudança do padrão de atenção à saúde da população (GIACOMOZZI; LACERDA, 2006).

As práticas da ESF têm o foco principal sob a família, através de ações de caráter preventivo sobre responsabilidade de uma área adscrita sobre determinada população. Assim, constitui-se uma prática menos “curativa” e cartesiana sobre a saúde, avançando para além da simples intervenção médica, buscando a integração com a comunidade, numa atuação interdisciplinar dos profissionais que compõem as equipes de saúde da família (MORAIS; BONFIN, 2005).

As modificações ocorridas na Atenção Primária à Saúde no Brasil, a partir da implantação da Estratégia de Saúde da Família têm repercutido na assistência às doenças cardiovasculares, principalmente sobre a Hipertensão Arterial Sistêmica, que passou de uma assistência pontual com conotações exclusivamente sanitária e intervencionista, a ser parte de um processo de atenção continuado, integral e multidisciplinar no qual se realizam ações assistenciais e sociais, dentro da lógica da vigilância à saúde (MORAIS; BONFIN, 2005).

Diversos estudos como o de Araújo e Guimarães (2007) apontam para a tendência na eficácia da assistência continuada da ESF com o controle dos níveis pressóricos. O aumento na frequência de consultas e conseqüentemente no

acompanhamento e avaliação dos usuários, vinculados à adesão desses ao tratamento da hipertensão é determinante para a normalização da pressão arterial Sistólica e Diastólica. Outra característica da ESF que possibilita controle desses níveis é a ação de uma equipe multidisciplinar (MANO, PIERIN, 2005),

2.2.2. Sistema de Informação em Saúde

O termo “Sistema de informação” (SI) é definido como grupo integrado de arquivos, procedimentos e equipamentos para o armazenamento, manipulação e recuperação de informações. É utilizado na área de Saúde Pública para definir toda a rede de produção, registro, análise, interpretação e retroalimentação de dados que possibilitem a tomada de decisões. Um sistema de informação contém um conjunto de variáveis integradas e articuladas entre si, obtendo e selecionando dados para transforma-los em informação. (TEIXEIRA, 2005).

O termo “informação” é definido como o significado que o homem atribui a um determinado *dado*, por meio de convenções e representações. Desta maneira, a informação constitui-se em suporte básico para toda atividade humana e que todo o cotidiano é um processo permanente de informação. E, no caso de instituições, empresas, organizações, conhecer seus problemas, buscar alternativas para solucioná-los, atingir metas e cumprir objetivos requerem conhecimento e, portanto, informação. Por isso, pode-se dizer que há um consenso de que não é possível exercer gerência em nenhum setor se não houver um sistema de apoio à decisão que se sustente na informação.

Da mesma forma, a informação em saúde deve ser entendida como um instrumento de apoio decisório para o conhecimento da realidade socioeconômica, demográfica e epidemiológica, para o planejamento, gestão, organização e avaliação nos vários níveis que constituem o Sistema Único de Saúde (CARVALHO, EDUARDO; 1998).

As teorias sobre planejamento em saúde, visando aprimorar as técnicas de reordenação das instituições de saúde e racionalização de suas atividades, incorporam as noções sobre sistemas de saúde e sistemas de informação, com base na “Teoria de Sistemas”. Segundo essa teoria, a constituição de um sistema implica uma interação entre todos os componentes da realidade que deverá ser captada por ele. Busca-se, através do sistema, a recomposição de um todo, que

será possível mediante o conhecimento e a comunicação (fluxos) entre as partes (BERTALANFF, 1975).

Na gerência de serviços de saúde, é básica a necessidade de cadastros de pacientes, cadastro da população, cadastros de estabelecimentos, produção das atividades de saúde, conhecimento do perfil de doenças atendidas, da mortalidade, número de profissionais de saúde, número de consultórios, leitos, medicamentos utilizados, gastos efetuados e tantas outras informações.

A Câmara Técnica de Informação e Informática (CTIIS) do Conselho Nacional dos Secretários de Saúde (CONASS) realizou entre os anos de 2004 e 2005 um levantamento em que foi registrado um número de 797 sistemas de informação utilizados na Secretarias Estaduais de Saúde. Encontraram cerca de 11 (onze) SI que se repetiam em basicamente todas as Secretarias Estaduais, sendo então configuradas de “base nacional”. Desta forma, surgiu a nomenclatura de “sistemas de informação de base nacional”, tendo sua relevância para o setor público de Saúde no Brasil, sendo então imposto pelo Ministério da Saúde como de utilização obrigatória em todos os estados e municípios, sendo sua base de dados consolidada e mantida pelo nível federal (CONASS, 2007).

Ao Departamento de Informática do SUS (DATASUS), órgão da Secretaria Executiva do Ministério da Saúde é atribuída a responsabilidade de coletar, processar e disseminar informações sobre saúde. Normalmente, tais ações são executadas por solicitação e em conjunto com outras áreas técnicas, voltadas a atividades finalísticas, principais usuárias dos sistemas desenvolvidos. As informações por eles veiculadas são disponibilizadas no domínio do DATASUS, com os números de forma bruta e na forma de indicadores. Por ser de fácil manipulação e extremamente útil para os gestores, é importante destacar que o DATASUS mantém em seu sitio, sempre atualizado, um “Caderno de Informações de Saúde”, com diversos tipos de dados – demográficos, epidemiológicos, financeiros – sobre cada estado e município do Brasil.

Esse Departamento responde de fato pelo desenvolvimento e manutenção da maioria dos SI do Ministério da Saúde, embora muitos dos existentes tenham sido desenvolvidos e sejam mantidos de forma autônoma por algumas áreas técnicas, com pessoal próprio ou terceirizado, como por exemplo, os da Vigilância Sanitária. Independente de sua origem, segundo Neves (2006), uma das características mais criticadas desses sistemas é a sua incapacidade de operar em conjunto, uns com os

demais. O Ministério da Saúde tem feito algumas tentativas para modificar essa situação, sendo exemplo de tais iniciativas a definição dos cadastros de usuários e de estabelecimentos de saúde, o Cartão Nacional de Saúde (CNS) e o Cadastro Nacional de Estabelecimentos de Saúde (CNES), em princípio únicos para todos os SI, e, mais recentemente, a proposta de unificação das tabelas do Sistema de Informações Ambulatoriais (SIA-SUS) e do Sistema de Informações Hospitalares (SIH-SUS) (CONASS, 2007).

Tal como acontece nas Secretarias Estaduais e Municipais de Saúde, o Ministério da Saúde mantém sistemas administrativos voltados a atividades intrínsecas à sua organização e funcionamento, mas, no que diz respeito às ações finalísticas da Saúde, os SI a elas relacionados dependem de dados coletados na “ponta”, ou seja, em estabelecimentos de saúde, direta ou indiretamente sob gestão estadual ou municipal. O fluxo desses dados é bastante variável, sendo que em alguns casos os municípios os enviam para os estados e estes, após consolidá-los, para o MS e, em outros, estados e municípios os enviam diretamente para o MS (NEVES, 2006).

A quantidade – em relação ao esperável – e a qualidade – em relação ao desejável – dessas informações é extremamente variável em todo o território nacional, sendo tão melhores quanto mais utilizadas pelos gestores locais em seus processos de trabalho rotineiros (MORAES, 2004).

Outros aspectos que merecem destaque, por sua relevância para o setor, foram: 1) o percentual importante de sistemas que operam sem qualquer controle de qualidade, formal ou informal, e, principalmente, 2) o grande número de respostas “não” (existem) e “não sei” (se existem) mecanismos para a garantia da privacidade dos dados. Foi também observada, no nível estadual, a repetição de muitos dos problemas apontados para o nível federal, em especial no que se refere à ausência de “comunicação” entre os sistemas de informação (CONASS, 2007).

2.2.3. Sistema de informação HIPERDIA: origem e contemporaneidade

A formação de grupos específicos para hipertensos teve início na década 1970 nos Estados Unidos, com o objetivo principal de monitorar a adesão ao tratamento com efetivo controle da pressão arterial. A troca de experiências entre os

membros do grupo, identificados pela condição comum da doença, exerceria ainda grande efeito terapêutico sobre eles (GOMES; SILVA; SANTOS, 2010).

Com a implantação do Sistema Único de Saúde (SUS) no Brasil, houve a necessidade de se criar Sistemas de Informação em Saúde (SIS) que fossem capazes de acompanhar toda a produção de dados e que garantisse a realização de avaliações periódicas e rotineiras da situação de saúde no país. Deste modo, a estruturação dos SISs é de fundamental importância para o SUS monitorar as condições de saúde da população (SILVA; LAPREGA, 2005).

O SUS dispõe de uma complexa gama de sistemas de informação em saúde, de base individual e abrangência nacional. A natureza política, universal e gratuita do SUS permite que seus sistemas de informação alcancem, senão a totalidade, boa parte de trabalhadores do país (FACCHINI, 2005).

Os SIS que abrangem atendimentos ambulatoriais e hospitalares, os de estatísticas vitais e de vigilância epidemiológica e sanitária. Dentre estes, tem-se os de abrangência nacional como o Sistema de Informação de Mortalidade (SIM), Sistema de Informação de Nascidos Vivos (SINASC), Sistema de Agravos de Notificação (SINAN), Sistema de Informação Ambulatorial (SIA), Sistema de Informação Hospitalar (SIH) e o Sistema de Informação da Atenção Básica (SIAB) e entre outros que são considerados elementos importantes para ações de controle, monitoramento e avaliação da atenção à saúde.

O SIAB foi implantado em 1998 para substituir o Sistema de Informação do Programa de Agentes Comunitários de Saúde (SISPACS). O SIAB visa o acompanhamento das equipes da Estratégia Saúde da Família (ESF) e, por meio desse, obtêm-se dados sobre cadastros de famílias, condições de moradia e saneamento, situação de saúde, produção e composição das equipes de saúde, ou seja, oferece informações sobre a situação de saúde em que as famílias se encontram (THAINES, et al, 2009). Esse sistema de informação possui oito fichas de acompanhamento, entre elas, a ficha de hipertensos (B-HA), e a ficha de diabéticos (B-DIA), sendo que ambas tem como finalidade cadastrar e acompanhar mensalmente os usuários com diagnóstico confirmado, e a cada visita domiciliar realizada pelo Agente Comunitário de Saúde (ACS) os dados deverão ser atualizados (BRASIL, 2003).

Na Atenção Básica, há um sistema de informação que também cadastra, acompanha e monitora os usuários com hipertensão arterial e *Diabetes mellitus*

conforme as diretrizes do Plano de Reorganização da Atenção a Hipertensão Arterial e *Diabetes Mellitus* (BRASIL, 2001).

O Plano de Reorganização da Atenção à Hipertensão Arterial Sistêmica e ao *Diabetes Mellitus* foi lançado no ano 2000, tendo o intuito de minimizar os problemas das doenças crônicas não transmissíveis como a hipertensão e diabetes. Diversas ações foram implementadas nos Estados e municípios, como capacitações profissionais na atenção básica, pactuação de normas e metas entre as três esferas da gestão de saúde, atenção à assistência farmacêutica e dispensação de medicamentos de uso contínuo, e promoção de atividades educativas (BRASIL, 2001).

Definidas como ações prioritárias aos grupos de risco, foram determinadas pela Norma Operacional Básica de Assistência à Saúde (NOAS), de janeiro de 2001, as ações desenvolvidas nas Unidades de Saúde da Família dirigidas principalmente aos pacientes hipertensos, incluindo acompanhamento rigoroso e assistência multiprofissional. A abordagem por uma equipe interdisciplinar contribui para oferecer ao paciente e à comunidade uma visão ampla do problema, dando-lhes motivação para adotarem mudanças nos hábitos de vida e aderirem ao tratamento.

O Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos (SisHiperdia) foi desenvolvido em 2002 em consequência do Plano de Reorganização, permitindo cadastrar e acompanhar os portadores de hipertensão arterial sistêmica (HAS) e/ou diabetes *Mellitus* (DM), em todas as Unidades Básicas do SUS; possuindo também o propósito de gerar informações para os trabalhadores e gestores de saúde, para aquisição, dispensação e distribuição de medicamentos a todos os usuários cadastrados (CONTIERO, et al, 2009). Por meio desse sistema informatizado, tornou-se possível a descrição do perfil epidemiológico dos pacientes cadastrados. Tal potencialidade, no entanto, pode ser mais explorada no país e agregada ainda aos indicadores de controle de hipertensão presentes no Pacto de Indicadores da Atenção Básica. Por sua vez, a informação é uma importante ferramenta para gestores e planejadores de saúde na formulação e avaliação de políticas públicas (BOING, BOING, 2007).

O HIPERDIA é disponibilizado pelo DATASUS, sendo de uso obrigatório para os municípios que aderiram à Estratégia Saúde da Família. Esse programa é “alimentado” pelos dados de consulta médica e de enfermagem, com atuação do

agente comunitário de saúde com um acompanhamento sistemático do usuário hipertenso pela equipe de Saúde da Família (ZILMER; SCHWARTZ; MUNIZ, 2010).

Uma meta primordial no direcionamento das ações da equipe de saúde ao hipertenso, fundamentado pelas diretrizes do Plano de Reorganização da Atenção à Hipertensão Arterial Sistêmica e ao *Diabetes Mellitus* é garantir a adesão do indivíduo ao tratamento. O desenvolvimento de estudos que analisem o controle da pressão arterial, bem como o conhecimento de seus resultados e a qualidade do acompanhamento desses pacientes, torna-se uma ferramenta indispensável ao trabalho do profissional de saúde que atua nessa área e à gestão de saúde do município. Entretanto, embora vários autores tenham publicado artigos sobre o tema em diversos países, na América Latina e, em especial, no Brasil, esses estudos ainda são muito escassos (SILVA, 2011).

A importância de todos estes sistemas está na informação gerada para a saúde, orientando a análise, o planejamento e a gestão. Ou seja, na informação gerada para a ação. Para se ter uma ação eficaz, a primeira condição é que se tenha uma informação confiável. A qualidade da informação depende da coleta adequada, da correta estruturação e da correta interpretação (fidedignidade) dos dados (TEIXEIRA, et al, 2005).

Igualmente importante em todo o processo, a retroalimentação das fontes de dados locais é componente essencial de todo o sistema de informações, pois além de ser o local da ação, é também através da retroalimentação que o serviço poderá acompanhar o produto de seu trabalho de maneira integral e aperfeiçar suas demais ações (TEIXEIRA, et al 2005). A partir dos dados coletados nos principais bancos é possível também estudar e planejar ações no nível local pela própria equipe de saúde da família.

A avaliação de políticas e programas é essencial em saúde pública, contribuindo para os esforços em busca de uma sociedade saudável e prevenindo desperdícios de recursos com a implementação de programas ineficazes. Na avaliação dos benefícios das políticas de saúde à população, o conhecimento dos arranjos e peculiaridades locais dos serviços de saúde é tratado como requisito básico. Entretanto, ao avaliar o efeito de uma política de saúde na atuação dos serviços e nas condições de saúde da população, é preciso valorizar o sinergismo entre os determinantes vinculados às políticas de saúde, aos serviços de saúde

(estruturas, recursos humanos e processos) e ao estado de saúde das populações (BOING; BOING, 2007).

Avaliar a qualidade das informações em saúde, nos Sistemas de Informação de Saúde (SIS), é imprescindível, na medida em que eles são instrumentos importantes para o diagnóstico da situação de saúde, pois caracterizam populações em risco e possibilitam planejar estratégias terapêuticas de acordo com as necessidades e especificidades de cada grupo populacional (ZILMER; SCHWARTZ; MUNIZ, 2010).

2.3. INCOMPLETUDE DE DADOS EM SISTEMAS DE INFORMAÇÃO EM SAÚDE E EM PESQUISAS EPIDEMIOLÓGICAS

Diversos estudos abordam a avaliação da qualidade dos dados de Sistemas de Informação em Saúde (FRIAS et al., 2008; VIDRO, 2004; ZILMER, 2009; OLIVEIRA; PALHA, 2008).

Melo et al. (2003) aborda as condições de preenchimento das fichas SINASC no Estado de São Paulo em 2002 apontando prejuízos quanto a falta de dados nestas fichas em variáveis consideradas indispensáveis para a avaliação do recém nascido nas unidades de saúde, como o índice de Apgar e duração da gestação.

Um estudo realizado objetivando conhecer o perfil de pacientes portadores de TBMR (tuberculose multirresistente) inscritos em um Centro de Referência para a TB (tuberculose) do Estado de São Paulo (2002/2009) e cadastrados no Sistema de Informação da TBMR. O estudo identificou altos índices de incompletude de variáveis chegando a 87% em alguns itens. Especificamente ao HIPERDIA, autores relatam as limitações encontradas para análise da distribuição de indicadores e perfil sociodemográfico referente aos portadores de HA (CHAZAN; PEREZ, 2008; ZILMER et al, 2009).

Boing e Boing (2007) relatam do problema de que os dados oriundos do HIPERDIA não podem ser tomados como representativos da prevalência de hipertensão nas regiões brasileiras, pelo fato de ser comum a inscrição das pessoas no programa por atendimento de demanda, quando o usuário procura o serviço público de saúde, havendo uma subrepresentação dos hipertensos. Vários

portadores de hipertensão não procuram as Unidades Básicas de Saúde e tampouco são cadastradas mediante busca ativa por parte dos serviços de saúde. Além disso, o registro dos pacientes depende da sistematização e remessa dos dados pelos municípios ao Ministério da Saúde. Muitos municípios, entretanto, não fazem o registro ou o fazem de maneira parcial.

A subnotificação, assim como a incompletude de dados, compromete a análise e qualidade das informações, assim como prejudica a real avaliação da qualidade da assistência prestada (ZILMER et al, 2009).

As Unidades de Saúde possuem precário sistema de informações, que não permite a integração entre os diversos sistemas e o cruzamento dos dados (MEDEIROS, 2005), fazendo com que tenham diversos cadastros a serem preenchidos, além do déficit em profissionais necessários para a digitação e retroalimentação do sistema.

Sobre a falta de dados ou de preenchimento de variáveis, tal problema é vastamente encontrado em estudos epidemiológicos (JUNGER, 2008). Geralmente, a situação é resolvida com a simples exclusão de variáveis ou dos sujeitos do estudo. Porém, determinar a abordagem analítica adequada para as circunstâncias de banco de dados com observações faltantes ou incompletas é uma questão que merece cuidados. A utilização de métodos inadequados pode levar a conclusões erradas sobre a análise do conjunto de dados (NUNES, 2009).

Diante dessa situação, desde a década de 80 há um crescente desenvolvimento de técnicas estatísticas direcionadas a solucionar tais problemas, especialmente na área da saúde (RUBIN, 1992; SCHAFER, 1999; HAREL, 2007; KENWARD, 2007)

2.4. SOLUÇÃO PARA VARIÁVEIS COM DADOS FALTANTES: IMPUTAÇÃO DE DADOS

A imputação de dados faltantes tem sido uma estratégia comum para a análise de dados com problemas de falta de dados ou incompletude de variáveis. Entende-se por imputação a técnica de preencher os dados faltantes com valores plausíveis (DALLACOSTA; DALLACOSTA; NUNES, 2010).

Sobre este problema de dados perdidos, critérios básicos para avaliar procedimentos de estimação foram estabelecidos por Neyman e Pearson (1973). Segundo Schafer; Grahan (2002):

Seja Q uma quantidade populacional a ser estimada, e seja $\hat{Q}$ uma estimativa de Q . Se a amostra contém dados faltantes, então o método para lidar com eles deve levar em conta o procedimento estatístico que será usado para calcular $\hat{Q}$. Se o procedimento funciona bem, então $\hat{Q}$ estará perto de Q tanto na média, para amostras repetidas, como para uma amostra em particular. Isto é, deseja-se que o viés – a diferença entre a média de $\hat{Q}$ e o verdadeiro valor de Q – seja pequeno e também se deseja que a variância ou o desvio padrão de $\hat{Q}$ seja pequeno.

2.4.1. Mecanismos de Informações Perdidas

Os dados perdidos são uma realidade em análises epidemiológicas. Raramente pode-se evitar problemas com a falta de dados. Desta forma, o desafio das pesquisas, segundo Hair (2008) é abordar as questões geradas pelos dados perdidos que podem afetar a generalidade dos resultados. A idéia primária do pesquisador é determinar as razões inerentes à tais dados. Há uma necessidade de se concentrar nos motivos dos dados perdidos, tentando compreender os processos que conduzem tais dados a fim de selecionar o direcionamento de ação apropriado.

Um processo de dados perdidos é qualquer evento sistemático externo ao respondente (como problemas na coleta de dados ou erro na entrada desses). Desse modo, quaisquer resultados estatísticos baseados nesses dados seriam tendenciosos, uma vez que as variáveis incluídas na análise são influenciadas pelo processo de dados perdidos. A preocupação em entender os processos de dados perdidos é análogo à necessidade de se entenderem as causas da falta de respostas no processo da coleta de dados.

Partindo da idéia de que os mecanismos que geram os dados faltantes sejam identificados e levados em conta nas análises dos dados, Rubin (1987) faz uma revisão sobre os mecanismos de não-resposta e os métodos de inferência que se utilizam na presença de tal modelo:

a) Dado faltante completamente aleatório – MCAR (Missing Completely at Random)

Dados faltantes têm essa denominação quando as razões para as perdas não são relacionadas a quaisquer respostas dos sujeitos, incluindo o valor faltante. É o mecanismo mais restritivo, impondo a mesma probabilidade de não-resposta para todas as situações.

b) Dado faltante não aleatório – (NMAR)

Pode ser denominado também como sendo de omissão não ignorável (NIM). Tal dado ausente é não aleatório se está relacionado com os valores não observados, mesmo se controlado para outras variáveis da análise. NIM pode ser considerado como o oposto de MCAR no sentido de que a probabilidade de ter valores ausentes em uma variável depende da variável em si.

Dados mais predispostos a serem faltantes, geralmente, são aqueles situados nos extremos da distribuição, com valores muito baixos ou altos comparados ao padrão da amostra.

c) Dado faltante aleatório – MAR (Missing at Random)

É quando o padrão de perda de uma variável é previsível a partir de outras variáveis no banco de dados, não sendo devido à variável específica na qual os dados são perdidos. Sendo assim, a probabilidade de ter valores em falta, não depende da variável em si, mas sobre os valores de alguma outra variável. A disponibilidade de variáveis auxiliares é que vai determinar a habilidade de se analisar corretamente os dados MAR. A maioria dos métodos de tratamento para dados faltantes assume que os dados são MAR (HARRELL, 2001). Nesse caso, o dado faltante pode ser considerado ao acaso, mas, condicional às outras variáveis observadas que determinam essa perda, estando disponíveis para análise (DONDEERS, 2006). MAR pode ser considerada como uma situação intermediária entre MCAR e NIM.

Essa suposição de perda de dados representa a condição mais geral sob a qual inferências válidas podem ser obtidas sem ser necessário citar o tipo de

mecanismo de não-resposta. Embora a suposição de que os dados sejam MAR não possa ser testada, como no caso de normalidade dos dados, têm sido abalizados em diversos artigos que, se fossem incluídas variáveis suficientes no modelo de imputação, mais próximo se chegaria a suposição do real (HAREL, 2007; KENWARD, 2007).

Segundo Nunes (2007) os dados com característica MCAR e MAR são também denominados dados faltantes ignoráveis, enquanto os NMAR são os não-ignoráveis.

2.4.2. Imputação Única

Também conhecido como Método de Substituição Simples, são métodos que substituem os dados perdidos uma única vez por métodos como os descritos abaixo, segundo conceitos de Engels e Diehr (2003):

I) Substituição por um Valor de Tendência Central:

Como o próprio nome apresenta, há uma substituição dos valores perdidos por um valor de tendência central resultante dos valores completos (média, moda, mediana).

II) “Hot deck”:

É um método cuja técnica visa localizar o indivíduo com dado observado mais parecido com o indivíduo com dado faltante em relação às variáveis auxiliares e substitui-se o dado faltante pelo valor do respondente pareado. Assim, cria-se uma matriz de “doadores” a partir dos itens preenchidos e de outras variáveis auxiliares, tipo sexo e faixa etária, ou seja, uma matriz de padrões de respostas. Verifica-se qual indivíduo respondente tem o mesmo padrão do não respondente em relação a determinadas variáveis. Aquele de mesmo padrão terá a resposta cedida ao que não respondeu.

III) Estimativa de Máxima Verossimilhança:

É um método comum de imputação. Consiste no seguinte processo: saber primeiramente a proporção de cada resposta para determinada variável e, em seguida, associar os dados faltantes proporcionalmente a cada resposta.

IV) Métodos para dados longitudinais:

Segundo Engels e Diehr (2003), as técnicas de imputação mais utilizadas são:

i) Para *grupos populacionais*:

- Método de Coluna Média: onde se utiliza a média de todas as pessoas no conjunto de dados para um determinado ano;
- Método da mediana da coluna: que corresponde a utilizar a mediana de todas as pessoas no conjunto de dados para um determinado ano;

ii) Para “Baseline” ou linha de base:

- Método da Classe Média: que utiliza a média de outras pessoas na classe correspondente;
- Método de Classe Mediana: que utiliza a mediana de outras pessoas na classe correspondente;
- Método Hot Deck: utilizando um valor de uma pessoa aleatória na classe correspondente.
- Regressão: onde há uma previsão do valor de um modelo de regressão;
- Regressão com erros: igual a Regressão com a adição de um termo de erro;

iii) Com *dados anteriores*:

- Linha média anterior: utilizando a média dos valores conhecidos do indivíduo anterior.
- Linha mediana anterior: com a mediana dos valores conhecidos do indivíduo anterior;
- LOCF: utilizando a última observação do indivíduo anterior.

iv) Com *dados “antes” e “depois”*:

- Linha média: média dos valores da pessoa “antes e depois”;
- Linha mediana: a mediana dos valores da pessoa “antes e depois”;
- Última e próxima: média dos últimos e próximos aos valores conhecidos;

v) *Regressão (média predita)*:

Os valores imputados são preditos através de regressão simples ou múltipla, que pode ser usada simplesmente utilizando uma ou mais variáveis altamente correlacionada com as anteriores.

Sentas e Angelis (2006) denominam imputação por regressão quando os valores em falta são estimados através da aplicação de regressão em que a variável com os dados em falta é considerada como dependente e todas as demais variáveis como preditores.

2.4.3. Imputação Múltipla (IM)

Segundo Engels e Diehr (2003), a Imputação Múltipla (IM) consiste em três passos:

- I) São obtidos vários bancos de dados completos por meio de técnicas adequadas de imputação;
- II) Separadamente, os vários bancos são analisados por um método estatístico tradicional, como se realmente fossem conjuntos completos de dados;
- III) Os vários resultados encontrados no passo “II” são combinados de um jeito simples e apropriado para obter a chamada inferência da imputação repetida.

O primeiro passo é a parte fundamental da IM, pois as técnicas de imputação utilizadas têm de preservar a relação das observações faltantes e presentes e ainda levar em conta o mecanismo de ausência e o padrão dos dados faltantes.

A partir das m imputações realizadas, o passo “II” da IM pode ser realizado, ou seja, os m bancos de dados são analisados por métodos tradicionais de análise. Finalmente, os m resultados obtidos podem ser combinados usando-se as regras propostas por Rubin (1987).

As regras de Rubin (*Rubin rules*) estão amplamente divulgadas na literatura que aborda a IM, pois são normas simples que resolvem o passo 3 da IM. Essas regras podem ser usadas independentemente do método utilizado para fazer a IM. A idéia é que a partir de cada análise sejam obtidas as estimativas para o parâmetro de interesse Q , ou seja, Q_j para $j = 1, 2, \dots, m$. Segundo Schafer (1999), Q pode ser qualquer medida escalar a ser estimada, tal como média, correlação, coeficiente de

regressão ou razão de chances. Então, a estimativa combinada será a média das estimativas individuais:

$$Q = \frac{1}{m} \sum_{j=i}^m \hat{Q}_j \quad (1)$$

Para a variância combinada, primeiramente calcula-se a variância dentro das imputações:

$$\bar{U} = \frac{1}{m} \sum_{j=i}^m U_j \quad (2)$$

E a variância entre imputações:

$$B = \frac{1}{m-1} \sum_{j=i}^m (\hat{Q}_j - \bar{Q})^2. \quad (3)$$

Então, a variância total, que é a variância combinada, será:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B \quad (4)$$

A imputação múltipla produz estimativas da variância levando em consideração que a distribuição da variável com valores faltantes não foi realmente observada; o que se determina intervalos de confiança mais precisos (Schafer, 1997).

Para a realização da análise computacional, alguns aplicativos têm sido com frequência citados na literatura, pois disponibilizam o uso dos métodos de IM. Dentre os mais utilizados, pode-se citar o SAS, S-Plus, SOLAS, NORM, BMDP e MICE, sendo que o MICE é de domínio público, pois é operado dentro do ambiente do aplicativo R (The R Foundation for Statistical Computing, Viena, Áustria; <http://www.r-project.org>).

A principal desvantagem dos procedimentos baseados em imputação é que em sua maioria a imprecisão devida à imputação não é contemplada na análise e,

portanto, a variância dos estimadores é subestimada. Os procedimentos mais simples de imputação freqüentemente levam a estimativas viesadas das associações (DONDERS et al, 2006).

2.5. PROJETO SOBRE A HIPERTENSÃO ARTERIAL SISTÊMICA (PAES, 2008;2009)

Motivado pela necessidade de conhecer o perfil dos usuários hipertensos e o grau de satisfação destes com os serviços prestados por gestores e profissionais de saúde (Agentes Comunitários de Saúde) do programa HIPERDIA, Paes (2008) desenvolveu seu projeto de pesquisa intitulado: *“Avaliação da Efetividade no controle da hipertensão arterial sistêmica e associação com fatores de risco comparando a atenção do Programa de Saúde da Família e de Unidades Básicas de Saúde de municípios do Nordeste do Brasil”* (PAES, 2008). Esse projeto foi desenvolvido no Laboratório de Estudos Demográficos - LED do Departamento de Estatística da UFPB no período de 2009/2010 e recebeu financiamento do CNPq (Edital: MCT/CNPq/MS – SCTIE – DECIT/MS No. 37/2008).

A partir desse projeto, deu-se seguimento a outro no ano de 2010, formando uma coorte, tendo como título: *“Desempenho do Programa Saúde da Família comparando com os das Unidades Básicas de Saúde no controle de Hipertensão Arterial Sistêmica e fatores associados em municípios do Estado da Paraíba: um estudo de coorte”* (PAES, 2009), também financiado pelo CNPq (Edital MCT/CNPq/CT-Saúde nº 67/2009)

O primeiro estudo foi realizado nos municípios de João Pessoa, Campina Grande – Paraíba, e Natal – Rio Grande do Norte. Já o segundo, apenas com os dois primeiros municípios citados. O universo em ambos os estudos foi constituído pelos usuários cadastrados no HIPERDIA nas Unidades de Saúde da Família, entre o período de 01 de Janeiro de 2006 até o dia 31 de Dezembro de 2007 (ANEXO 2).

Como instrumentos de avaliação foram utilizados questionários, baseados em três fontes de coleta: formulário do HIPERDIA, prontuários e questões formuladas para a entrevista (PAES, 2008;2009) (ANEXO 1). O espaço destinado às informações secundárias, oriundas do HIPERDIA e prontuários, encontra-se na parte inicial do questionário, já as questões referentes à entrevista (dados primários),

encontram-se estruturadas em noventa e seis perguntas subdivididas em oito dimensões essenciais da atenção básica, classificadas como: *Fatores de Risco e Doenças Concomitantes*; *Saúde do Caso Confirmado de HAS*; *Acesso ao Diagnóstico*; *Acesso ao Tratamento*; *Adesão/vínculo*; *Elenco de Serviços*; *Coordenação*; *Enfoque na Família*; *Orientação para a comunidade*. O propósito dessas dimensões consiste em conhecer o grau de satisfação dos hipertensos cadastrados no programa HIPERDIA, cujas respostas foram mensuradas na escala tipo Likert, envolvendo os seguintes itens: “Nunca”, “Quase Nunca”, “Às Vezes”, “Quase Sempre” e “Sempre”, quantificadas de 1 a 5. As perguntas correspondentes a cada dimensão foram adaptadas a partir de um estudo proposto por Bárbara Starfield (2002), para avaliação da qualidade da Atenção Básica de Saúde.

Para a realização das entrevistas, participaram entrevistadores previamente treinados e com qualificação para a verificação de medidas antropométricas (peso, altura, cintura e quadril) e da pressão arterial (PA) de cada paciente. Os instrumentos para verificação dessas medidas foram submetidos à calibração por uma empresa especializada segundo as normas do Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (INMETRO).

Desta forma, a coleta foi organizada considerando o cadastramento nos anos 2006/2007 e observaram-se os prontuários desses usuários no ano de 2008, para verificar o desempenho desses no decorrer do ano. Em 2009 foi aplicado o questionário estruturado (ANEXO 2), o qual foi repetido em 2010, com os mesmos usuários constituindo-se, desta forma, em uma coorte.

As informações coletadas no projeto Paes (2008; 2009) suscitaram o questionamento sobre sua qualidade, tanto daquelas provenientes do cadastro do programa HIPERDIA, quanto do questionário referente ao grau de satisfação dos usuários hipertensos com os serviços oferecidos pelo programa.

Conforme já discutido, a qualidade das informações é freqüentemente medida pela incompletude, ou seja, pelos dados perdidos, ou não informados. Desse modo, recorreu-se aos métodos de imputação de dados para suprir as eventuais perdas de dados e/ou informações.

Capítulo 3

3. MATERIAL E MÉTODOS

3.1. FONTES E BASES DOS DADOS

O instrumento para a pesquisa foi composto por 96 itens, divididos em dois blocos: dados primários (n=77) e secundários (n=19). Os dados secundários correspondem às informações extraídas dos formulários do programa HIPERDIA (ANEXO 1) e do prontuário (consultas que apresentem registro de mensuração pressórica).

Os dados primários abordaram questões que envolveram e avaliaram o grau de satisfação dos usuários com os serviços de saúde, quanto ao tratamento, e avaliação da situação de saúde dos indivíduos (ANEXO 2).

Para o estudo, abalizou-se o município de João Pessoa, capital Paraibana, que possuía uma população de aproximadamente de 674.762 habitantes, segundo dados do Instituto Brasileiro de Geografia e Estatística – IBGE (2008). Segundo a base de informações do SIAB (2008), o município de João Pessoa possuía 180 equipes de saúde distribuídas em cinco Distritos Sanitários, com um total de 43.953 hipertensos cadastrados até o ano de 2008.

Para construção da base de dados do Projeto PAES (2008), foi utilizado um tamanho amostral representativo para o município de João Pessoa, com base no processo de amostragem casual simples, considerando um população de hipertensos cadastrados $N = 43\ 953$, um parâmetro de sucesso com prevalência $p=0,119$, nível de confiança de 95% e margem de erro amostral máximo $\varepsilon = 0,035$ (3,5%), resultando em uma amostra de 343 usuários hipertensos, tendo como fonte de dados a Estratégia Saúde da Família fornecida pelo SIAB (2008).

Os critérios para inclusão do usuário do estudo foram: ter idade maior que 19 anos e ser cadastrado no programa nos anos 2006/2007. Os dados secundários foram retirados da ficha do HIPERDIA e do prontuário de cada usuário. A classificação de o paciente ser acompanhado ou não, foi utilizado seguindo o critério da classificação adotado pelo Ministério da Saúde (2002), em que foi considerado acompanhado o paciente que teve três ou mais consultas, contendo os valores da pressão arterial sistêmica, na unidade de saúde, e não-acompanhado foi definido como aquele usuário que teve nenhuma ou menos de três consultas ao ano.

No primeiro momento, 2009/2010, foram entrevistados 343 usuários, e no segundo momento, em 2010, houve perdas diminuindo o tamanho da amostra por diversos motivos, obtendo-se o número de 306 usuários. Dessa forma, utilizou-se o banco de dados relativo ao segundo momento do estudo, dando seguimento à qualificação dos bancos de dados da coorte.

Para a construção do banco de variáveis neste trabalho, ele foi dividido por dados secundários, obtidos em 2009 (cadastrados em 2006/2007), e primários, obtidos pelas entrevistas em 2010. Os dados secundários correspondem às informações contidas no cadastro HIPERDIA, constituído por três blocos (para hipertensos cadastrados em 2006/2007) e outro para as informações do prontuário:

I - DADOS SECUNDÁRIOS

a) Informações do HIPERDIA

1º BLOCO - *Dados Antropométricos*: contendo dados do usuário mensurados no momento do cadastro com: *Pressão Arterial Sistólica, Pressão Arterial Diastólica, Peso, Altura e Cintura*.

2º BLOCO - *Informações Socio-demográficas*: abordando características do perfil do usuário: *Sexo, Idade, Situação Familiar/Conjugal, Escolaridade, Raça/Cor*.

3º BLOCO - *Fatores de Risco e Doenças Concomitantes*, em que foram classificados com diagnóstico positivo:

- *Antecedentes Familiares Cardiovasculares* – Pessoas cujos familiares já apresentaram problemas cardiovasculares.
- *Diabetes Tipo 1* – São pessoa que fazem uso de insulinas para o controle da glicose desde o momento do diagnóstico (insulino-dependentes).
- *Diabetes Tipo 2* – São pessoas geralmente de idade acima de 40 anos e que possuem excesso de peso (insulina-não dependentes).
- *Tabagismo* – são considerados aqueles indivíduos que consomem 01 ou mais cigarros por dia.

- *Sedentarismo* – pessoas que realizam menos que 30 minutos de exercício, 03 vezes por semana e não faz em esforço físico pesado em casa ou no trabalho.
- *Sobrepeso ou Obesidade* – a classificação é dada de acordo com a tabela de referência da ficha de cadastro dos hipertensos.
- *AVC* – Pacientes que já foram acometidos por Acidente Vascular Cerebral.
- *Infarto agudo no miocárdio* – Pacientes que já apresentaram algum evento de infarto.
- *Outras coronariopatias* – Pacientes que apresentaram algum outro problema do coração.
- *Pé diabético* – Pacientes que tiveram alterações anatomopatológicas e neurológicas periféricas que ocorrem nos pés de pessoas acometidas pelo Diabetes Mellitus.
- *Amputação por diabetes* – Pacientes que tiveram algum tipo de amputação por conta da diabetes.
- *Doença Renal* – Pacientes que já tiveram algum problema nos rins.

b) Informações do Prontuário

- *Etilismo* – A partir da realização da entrevista foi gerada a variável, onde se perguntou ao paciente sobre a utilização de bebidas alcoólicas, classificando-o em alcoolista aquele que bebe com frequência e não alcoolista aquele que bebe ocasionalmente.
- *Menopausa* – Usuárias que apresentaram a última menstruação espontânea;
- *Uso de anticoncepcivo hormonal* – usuárias que utilizam métodos hormonais para impedir ou reduzir significativamente o processo de fecundação.

II – DADOS PRIMÁRIOS

A segunda parte do banco de dados constituiu-se de dados primários, coletados a partir do questionário desenhado para atender às dimensões de avaliação de serviços de atenção básica (PAES, 2009). Esses dados englobam oito dimensões, denominadas (ANEXO 2):

F - *Saúde do Caso Confirmado de HAS;*

G - *Acesso ao Diagnóstico;*

H - *Acesso ao Tratamento;*

I - *Adesão/vínculo;*

J - *Elenco de Serviços;*

L - *Coordenação;*

M - *Enfoque na Família;*

N - *Orientação para a comunidade.*

3.2. PERFIL DESCRITIVO DOS USUÁRIOS COM DADOS FALTANTES NO HIPERDIA

Construiu-se uma matriz de variáveis para demonstrar aquelas com maior número de falta de informações, obtendo o cenário desses dados perdidos. Dentre as variáveis presentes no cadastro HIPERDIA, foram escolhidas aquelas com maior número de dados faltantes, e que fazem parte do bloco de Informações gerais: Distrito Sanitário, Sexo, Escolaridade e Idade. As variáveis da matriz foram cruzadas para a identificação daquelas denominadas como “chaves”, para que se tivesse um número de usuários com a condição de presença de dados faltantes em todas essas variáveis.

3.3. MÉTODOS DE IMPUTAÇÃO

De acordo com as características das informações geradas pelas dimensões organizadas no questionário (dados em escala ordinal), escolheram-se quatro métodos de imputação única. Segundo critérios estabelecidos para esses tipos de dados, Nunes (2007) alega que existem dificuldades operacionais em se utilizar a

Imputação Múltipla: é necessário mais espaço para armazenar um número maior de dados e resultados obtidos, propiciando assim, mais trabalho em se analisar um banco completo do que pela Imputação Única, e, a Imputação Múltipla não está vastamente implementada nos aplicativos estatísticos, sendo escassos os suportes computacionais que apresentem tal método.

A escolha dos quatro métodos se deu pela sua praticidade de aplicação, disponibilidade de aplicativos estatísticos, por serem de mais amplo conhecimento e viabilidade técnica.

3.3.1 Método 1: Substituição por um Valor de Tendência Central

Os valores faltantes são substituídos pela média da variável: média geral, ou a média de um grupo mais similar ao do caso com o faltante, identificado por uma ou mais variáveis categóricas presentes no banco de dados. Pode-se substituir também pela mediana ou moda da variável ou de outro grupo de casos mais similares.

Sempre que existirem *outliers*, ou valores extremos, na amostra, é recomendado utilizar o valor da mediana ao invés do valor da média ou moda. A imputação com a mediana pode produzir melhores resultados no caso de variáveis com distribuição assimétrica (Miettinen, 1985).

Desta forma, foi adotado nesse estudo o valor da moda para a imputação dos dados faltantes, mais indicada para variáveis de nível ordinal.

3.3.2 Método 2: *Hot deck*

Os valores dos respondentes, similares em relação às variáveis auxiliares, foram selecionados para a imputação, doando características definidoras para os dados. Assim, localiza-se o indivíduo com o dado observado mais parecido com o indivíduo com dado faltante em relação às variáveis auxiliares e substitui-se tal dado faltante pelo valor do respondente pareado. Havendo mais de um respondente pareado, deverá ser usado o método de imputação do “vizinho mais próximo”, onde algum critério de classificação é desenvolvido para determinar o registro mais semelhante àquele com o dado faltante e determinado registro se torna “doador” desses dados. Um exemplo de variáveis, doadoras para determinar o perfil usado na determinação de tal imputação, são as variáveis sócio-demográficas, como sexo, idade e localidade de domicílio. Assim, verifica-se qual indivíduo respondente tem o

mesmo perfil do não-respondente em relação a tais dados sócio-demográficos. Aquele do mesmo padrão, ou perfil, é o doador. Desta forma, o dado faltante no item será preenchido com a resposta do doador.

No estudo, foram utilizadas como variáveis chave dos usuários doadores informações de: Distrito Sanitário, Tipo de Acompanhamento, Idade, Faixa Etária, Sexo e Escolaridade. A Faixa Etária foi estabelecida segundo os critérios de idade “maior ou igual a 60 anos” e “menor que 60 anos”. Essas variáveis foram escolhidas por serem as mais frequentes com respostas positivas.

3.3.3 Método 3: Estimativa de Máxima Verossimilhança

Esse método faz referência ao algoritmo EM (*Expectation-Maximization*) e é atualmente um método bastante comum de imputação. Tal algoritmo é utilizado quando se deseja estimar parâmetros a partir de um conjunto de dados incompletos. É um processo iterativo em que se repetem dois passos até que haja convergência: E (Estimação) e M (Maximização). No passo E se estima os dados faltantes para completar a matriz dos dados. No passo M, com os dados completados, há uma aprendizagem das probabilidades e então essas probabilidades são usadas para fazer a inferência no passo E, e assim, sucessivamente. O algoritmo é processado até que haja convergência.

É no contexto dos procedimentos baseados em modelos que será introduzido o algoritmo EM (*Expectation-Maximization*), um método aplicável a diversos padrões de dados faltantes, proposto por Dempster, Laird e Rubin (1977).

O algoritmo EM é um processo iterativo para estimação da máxima verossimilhança em problemas de dados incompletos. O algoritmo consiste na formalização da idéia intuitiva de lidar com dados incompletos: (1) substituir os valores faltantes por valores estimados, (2) estimar os parâmetros, (3) reestimar os valores faltantes considerando que os novos parâmetros são corretos, (4) reestimar os parâmetros. Este processo é repetido até que um critério de convergência seja alcançado (Little; Rubin, 1986).

A aplicação do algoritmo EM em problemas que envolvem dados incompletos está restrita aos casos em que o mecanismo de formação dos dados faltantes é do

tipo faltando ao acaso (MAR). Esta restrição elimina os casos onde os valores estariam faltando devido aos valores observados (Dempster; Laird; Rubin, 1977).

O uso do termo “dados incompletos” sugere a existência de dois espaços amostrais X e Y e um mapeamento de X para Y . Os dados observados $\mathbf{y}$ são uma realização do espaço amostral Y . O correspondente $\mathbf{x}$ em X é observado através de uma função $\mathbf{y}$. Considera-se, então, que existe um mapeamento $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ de X para Y , e $\mathbf{x}$ pertence a $X(\mathbf{y})$ que é o subconjunto de X determinado pela equação $\mathbf{y} = \mathbf{y}(\mathbf{x})$, onde $\mathbf{y}$ são os dados observados (Dempster; Laird; Rubin, 1977). O vetor $\mathbf{x}$ é o vetor de dados completos, pois representa o conjunto de dados que teria sido observado em uma situação sem a ocorrência de dados faltantes. Se $\mathbf{z}$ denota o vetor de dados faltantes ou dados não observados, então,

$$\mathbf{x} = (\mathbf{y}, \mathbf{z})^T \quad (5)$$

Seja $f(\mathbf{x} | \boldsymbol{\theta})$, onde $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_d)$ é um vetor de parâmetros desconhecidos pertencentes ao espaço paramétrico d -dimensional Θ , a função de densidade de probabilidade conjunta do vetor aleatório $\mathbf{X}$, dependente do parâmetro $\boldsymbol{\theta}$, correspondente ao vetor de dados completos $\mathbf{x}$. E $g(\mathbf{y} | \boldsymbol{\theta})$ a função de densidade de probabilidade conjunta do vetor aleatório $\mathbf{Y}$ correspondente aos dados observados $\mathbf{y}$. Então, as densidades dos dados completos e observados, respectivamente $f(\mathbf{x} | \boldsymbol{\theta})$ e $g(\mathbf{y} | \boldsymbol{\theta})$, estão relacionados pela equação:

$$g(\mathbf{y} | \boldsymbol{\theta}) = \int_{\mathbf{x}(\mathbf{y})} f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} \quad (6)$$

O objetivo do algoritmo EM é encontrar o valor de $\boldsymbol{\theta}$ que maximiza $g(\mathbf{y} | \boldsymbol{\theta})$ dados os valores observados $\mathbf{y}$ utilizando $f(\mathbf{x} | \boldsymbol{\theta})$ (Dempster, Laird e Rubin, 1977).

Cada iteração do algoritmo EM envolve dois passos que são o *expectation* (passo E) e o *maximization* (passo M). Suponha que $\boldsymbol{\theta}^{(k)}$ representa o atual valor de $\boldsymbol{\theta}$ após k iterações, no passo E da iteração $(k+1)$ é calculada a esperança condicional da função de log-verossimilhança de $\boldsymbol{\theta}$ dados $\mathbf{y}$ e $\boldsymbol{\theta}^{(k)}$, e no passo M, é determinado $\boldsymbol{\theta}^{(k+1)}$ pertencente ao espaço Θ que maximiza a esperança condicional da log-verossimilhança.

A função de log-verossimilhança dos dados completos $\mathbf{x}$ para o parâmetro $\boldsymbol{\theta}$, se todo o vetor $\mathbf{x}$ tivesse sido observado, é dada por

$$\log L_c(\boldsymbol{\theta}) = \log (f(\mathbf{x}|\boldsymbol{\theta})) \quad (7)$$

Os passos E e M se repetem alternadamente até que a diferença entre os valores da verossimilhança dos dados incompletos na k -ésima e na $(k+1)$ -ésima iteração seja tão pequena quanto se deseja, ou seja,

$$L(\boldsymbol{\theta}^{(k+1)}) - L(\boldsymbol{\theta}^{(k)}) < I \quad (8)$$

onde I , arbitrário, é a maior diferença para se considerar a convergência da seqüência de valores da verossimilhança $\{L(\boldsymbol{\theta}^{(k)})\}$ (McLachlan e Krishnan, 1996).

Dempster, Laird e Rubin (1977) mostram que a função de verossimilhança dos dados incompletos $L(\boldsymbol{\theta})$ não decresce após uma iteração do algoritmo EM, isto é, sua convergência é monótona:

$$L(\boldsymbol{\theta}^{(k+1)}) \geq L(\boldsymbol{\theta}^{(k)}) \quad (9)$$

para $k = 0, 1, 2, \dots$

3.3.4 Método 4: Regressão Logística Multinomial

Este método consiste em uma generalização da regressão logística (RL), que é usado para a relação entre um modelo dicotômico (binário) variável dependente e um conjunto de k variáveis preditoras $\{x_1, x_2, \dots, x_k\}$, que podem ser categóricas (fatores) ou numéricas (covariáveis). Como a variável dependente binária sempre pode ser interpretada como a ocorrência ou não para um evento E , o modelo de regressão logística é expressado de forma que:

$$\log \left(\frac{\text{prob}(E)}{1 - \text{prob}(E)} \right) = b_0 + \sum_{i=1}^k b_i x_i \quad (10)$$

onde os b_i denotam os coeficientes desconhecidos da regressão logística desconhecido (b_0 é o intercepto), enquanto que $prob(E)$ denota a probabilidade E do evento ocorrer. A quantidade no lado esquerdo da equação (10) é chamada de logit. Assim, o modelo simples de RL pode ser utilizado para prever a probabilidade de ocorrência de um evento.

O modelo pode ser generalizado, no caso em que a variável dependente tenha valores em mais de duas categorias, ou seja, politômica. Nesse caso, assume-se que as categorias possíveis são q , e o modelo será definido por:

$$\log\left(\frac{prob(categ_j)}{prob(categ_q)}\right) = b_0^{(f)} x_i, j = 1, \dots, q-1 \quad (11)$$

Em (11), pode-se ver que uma das categorias é usada como referência, sendo chamada de categoria de base. Após estimar os coeficientes do modelo (6) pelo método de máxima verossimilhança, pode-se facilmente calcular os logits e, portanto, as probabilidades de cada uma das categorias. A previsão final é a categoria com a máxima probabilidade. A Regressão Logística Multinomial pode ser utilizada para a imputação ao considerar a variável dependente categórica com os valores em falta e todos os outros preditores (SENTAS; ANGELIS, 2006).

3.4 ESTUDO DE SIMULAÇÃO

Após a imputação dos dados perdidos é necessário examinar o desempenho das estimativas obtidas usando medidas resumos. Mas, para que a comparação das estimativas não fosse baseada em uma única amostra, recorreu-se ao processo relativo de simulação com várias amostras.

Uma das dificuldades na avaliação de procedimentos de imputação de dados faltantes é que geralmente não há como comparar os valores imputados com os valores originais. Por isto, neste trabalho, optou-se por usar dados reais.

A fim de avaliar a validade dos métodos de imputação propostos neste estudo e possibilitar a comparação com outros existentes na literatura foi conduzido um estudo de simulação baseado na geração de padrões representativos de diversos

cenários de dados faltantes. Os padrões de dados faltantes considerados foram os resultantes da combinação de mecanismos de geração de dados faltantes, proporções de dados faltantes e configurações das lacunas ausentes.

Foram investigadas proporções iguais a 5%, 10%, 15%, 30% e 40% de dados faltantes. Alguns autores (Greenland e Rothman, 1998; Miettinen, 1985) indicam que, para uma pequena proporção de dados faltantes e um grande número de observações, a análise de dados completos produz bons resultados, portanto o cenário de 5% foi incluído como referência. A proporção de 40%, por outro lado, serve para avaliar os métodos de imputação sob condições extremas de informações perdidas. O intervalo de 5% e 15% é considerado com condições factíveis de imputação

Foram simulados 500 bancos de dados pelo programa SAS®, de forma que apresentasse dados faltantes de forma aleatória e com um conjunto de amostras diferentes a fim de que fosse executada e comparada diversas versões de amostras para verificar se de fato existe uma melhor técnica de imputação para as situações propostas.

Segundo Pacheco e Cardoso (2005), o processo de simulação se resume na execução ou manipulação dinâmica de um modelo de um sistema-objeto com um objetivo qualquer. A simulação é uma técnica de resolução de problemas que se dá pela observação do comportamento sobre o tempo, de um modelo dinâmico de um sistema (GORDON, 1978).

O processo de simular permite que se possa fazer identificação de problemas, a comparação com o desempenho de outros sistemas, estudos sobre a utilização da capacidade instalada, níveis de inventário, lógica de controle, refinamento de projeto, integração, seqüenciamento, testes de inicialização de equipamentos, entre outras aplicações (PACHECO; CARDOSO, 2005). Colmanetti (2001) afirma que o uso de simulação pode proporcionar uma melhor compreensão do sistema real, poupar tempo no processo de aprendizado. A simulação permite ainda testar hipóteses sem interferir na operação do sistema real, é mais aderente que modelos matemáticos e assim proporciona análises mais realistas que a modelagem matemática.

De uma forma genérica podemos descrever o processo de simulação de forma que: o analista constrói um modelo do sistema de interesse, escreve um programa de computador que representa o modelo e usa o computador para imitar o

comportamento do sistema quando sujeito a uma variedade de políticas de operação. Através de análise dos resultados, escolhe-se a melhor política de operação a ser adotada.

Pensando dessa forma, o estudo trouxe como alternativa do estudo para comparações técnicas de imputação, dois processos distintos de simulação de dados, para diferentes proporções de dados faltantes: “Cenário 1” e “Cenário 2”, embora sabe-se que são infinitas as possibilidades de cenários.

3.4.1. Simulação para Cenário 1

Analisando a validade dos quatros estimadores para a imputação (Moda, *Hot-deck*, Máxima Verossimilhança e Regressão Logística Multinomial) foi feito um estudo de simulação utilizando o item F4 – O (a) Senhor(a) pratica atividade física, com que frequência? – do questionário (ANEXO 1). Como citado na metodologia, tal item foi escolhido por ter um índice de completude de 100%, já que há uma necessidade de uma quantidade máxima de dados originais disponíveis, visto que se utilizará como “padrão ouro” para comparação entre as amostras.

O estudo foi baseado na geração de padrões representativos de diversos cenários de falta de dados. Investigou-se proporções iguais a 5%, 10%, 15%, 30% e 40%, resultando em um número de 15, 30, 45, 90 e 120 dados faltantes gerados respectivamente. Tais amostras sem reposição foram usadas a fim de alcançar a quantidade de valores faltantes desejada em cada rodada. Pressupôs-se que os dados fossem do tipo MCAR e dessa forma foram gerados como uma amostra aleatória de todas as observações do conjunto de dados utilizado.

Cada cenário de dados faltantes foi replicado 500 vezes e imputado segundo as estimativas estudadas num trabalho computacional intensivo. A Construção da simulação de amostras aleatórias foi feita pelo software SAS® - *Statistical Analysis System*, versão 9.1.3, executando na plataforma Windows, sob uma configuração de um Pentium IV com processador Intel *Duo-core* de 4 Gb de memória RAM.

A Tabela 3.1 demonstra a análise descritiva da variável objeto da simulação, tendo o dado mais frequente a resposta “1” (52,8%). Desta forma, visto que um dos métodos ou estimativas utilizadas seria a Tendência Central pela moda, tal valor foi substituído nas lacunas de dados ausentes gerados aleatoriamente pelo estudo de simulação.

Os estimadores *Hot Deck* e Máxima Verossimilhança foram utilizados segundo as delimitações propostas nos métodos desse estudo (seção 3.4.2 e 3.4.3). Para cada simulação, foram utilizadas padrões de amostras de valores observados.

Tabela 3.1: Análise descritiva do item F4

F4		
Respostas	Frequencia	Percentual (%)
0	1	0,3
1	161	52,8
2	28	9,2
3	53	17,4
4	15	4,9
5	47	15,4
Total	305	100,0

Para a imputação pela Regressão Logística Multinomial utilizou-se o modelo metodológico proposto na seção 3.4.4, equação 11.

Como foram utilizados vários bancos de dados faltantes, com várias combinações de valores da amostra, resultou-se em modelos dinâmicos, no sentido de que os valores estimados dos parâmetros ($b_0, b_1, \dots, b_q$) eram alterados a cada apresentação de uma nova amostra.

Desta forma foi possível identificar e construir as amostras simuladas para imputação das situações encontradas. Assim, foram definidos valores para os seguintes parâmetros: tamanho da amostra (n), proporção de simulação de dados faltantes (x); número de amostras simuladas (k); métodos para imputação (z):

- a) $n = 305$ (o n inicial era de 306, mas excluiu-se um sujeito por ter alto índice de dados incompletos – 87,8%)
- b) $x = 0,05; 0,10; 0,15; 0,30; 0,40$
- c) $k = 500$
- d) $z = 4$

A Figura 3.1 demonstra o fluxograma do esquema de simulação para construção no SAS®.

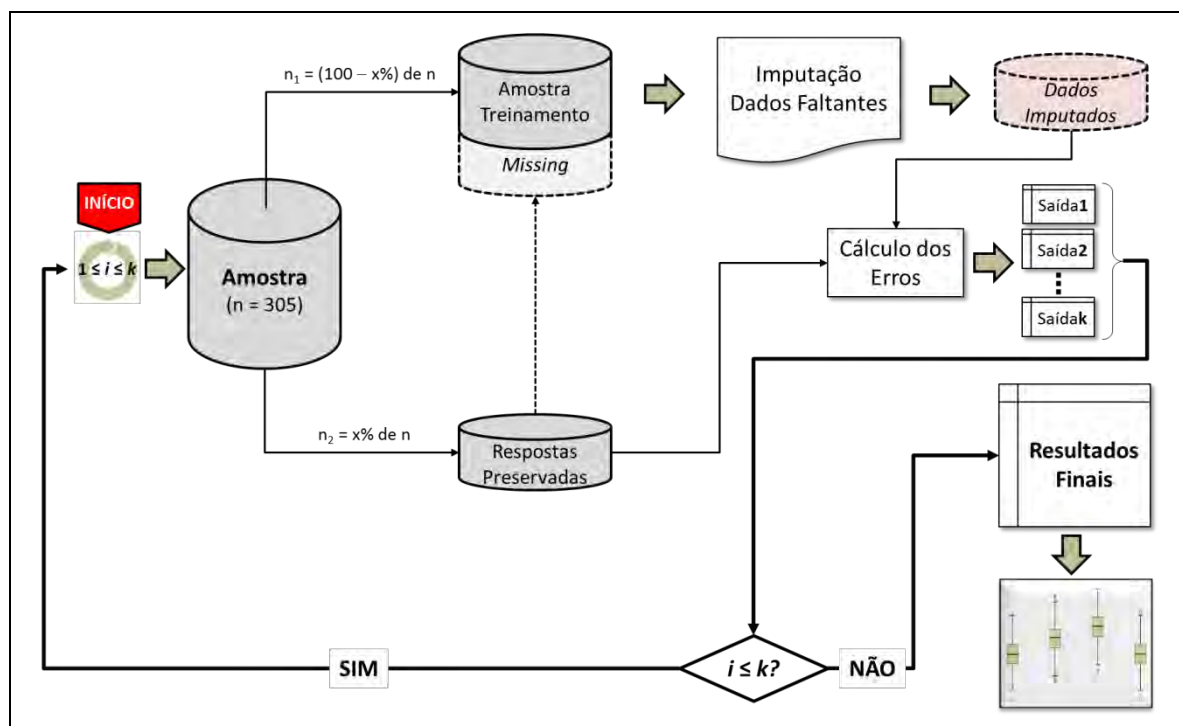


Figura 3.1: Esquema do Processo de Simulação para o Cenário 1.

Nota: k representa o total de simulações;
 $x\%$ representa o percentual de dados faltantes.

A combinação desses parâmetros abordou 20 estruturas diferentes para cada amostra diferente simulada, ou seja, para cada amostra, retirou-se a proporção necessária de dados e se imputou pelas quatro técnicas propostas. Para cada uma das 20 estruturas diferentes, criou-se um laço de repetição de 500 simulações, gerando assim 10.000 amostras distintas para análise de desempenho preditivo, a fim de evitar que um experimento amostral não favoreça a um ou a outro método de imputação.

Desta forma, da amostra inicial ($n=305$), foi retirada a proporção de valores considerados, deixando preservadas as informações reais. A amostra com a falta de dados é dessa forma imputada pelos métodos de substituição única sugeridas (TC, HD, MV, RLM). Quando o processo de simulação atingiu o número de 500 amostras aleatórias com cada amostra imputada pelos 4 diferentes métodos, usou-se a comparação pelas medidas resumos – Quadrado Médio do Erro (QME) e Erro Percentual Médio Absoluto (MAPE) e pela Proporção de Acerto (%Acerto).

Depois da disposição dos valores de erros, os resultados finais foram apresentados por *Box-plot* para melhor visualização, interpretação e comparação das implicações.

3.4.2. Simulação para o Cenário 2

O segundo momento para simulação apresentou a construção de uma variável com quantidade de respostas ou dados homogêneos. A amostra simulada teve 306 dados com aproximadamente 16,66% de preenchimento para cada tipo de resposta, cujos valores variavam de 0 a 5. Essa é uma situação mais rara e extrema de acontecer, mas que difere da particularidade da simulação com a variável real utilizada no primeiro momento (Cenário 1).

A Tabela 3.2 descreve os valores utilizados para gerar amostras com padrões de respostas na construção do segundo cenário:

Tabela 3.2: Análise descritiva do item simulado para construção do Cenário 2.

Item Simulado		
Respostas	Frequência	Percentual (%)
0	21	16,66
1	21	16,66
2	21	16,66
3	21	16,66
4	21	16,66
5	21	16,66
Total	306	100,0

Desta forma, foram definidos valores para os parâmetros, da mesma forma que no Cenário 1: tamanho da amostra (n), proporção de simulação de dados faltantes (x); número de amostras simuladas (k); métodos para imputação (z):

- a) $n = 306$
- b) $x = 0,05; 0,10; 0,15; 0,30; 0,40$
- c) $k = 500$
- d) $z = 4$

O esquema de simulação, demonstrado na Figura 4.2 foi o mesmo que no Cenário 2, diferindo apenas no tamanho da amostra inicial que passou a ser de 306 indivíduos, argumenta-se tal número para que as respostas tenham frequência igual na amostra, ou seja, cada uma se apresentará 21 vezes.

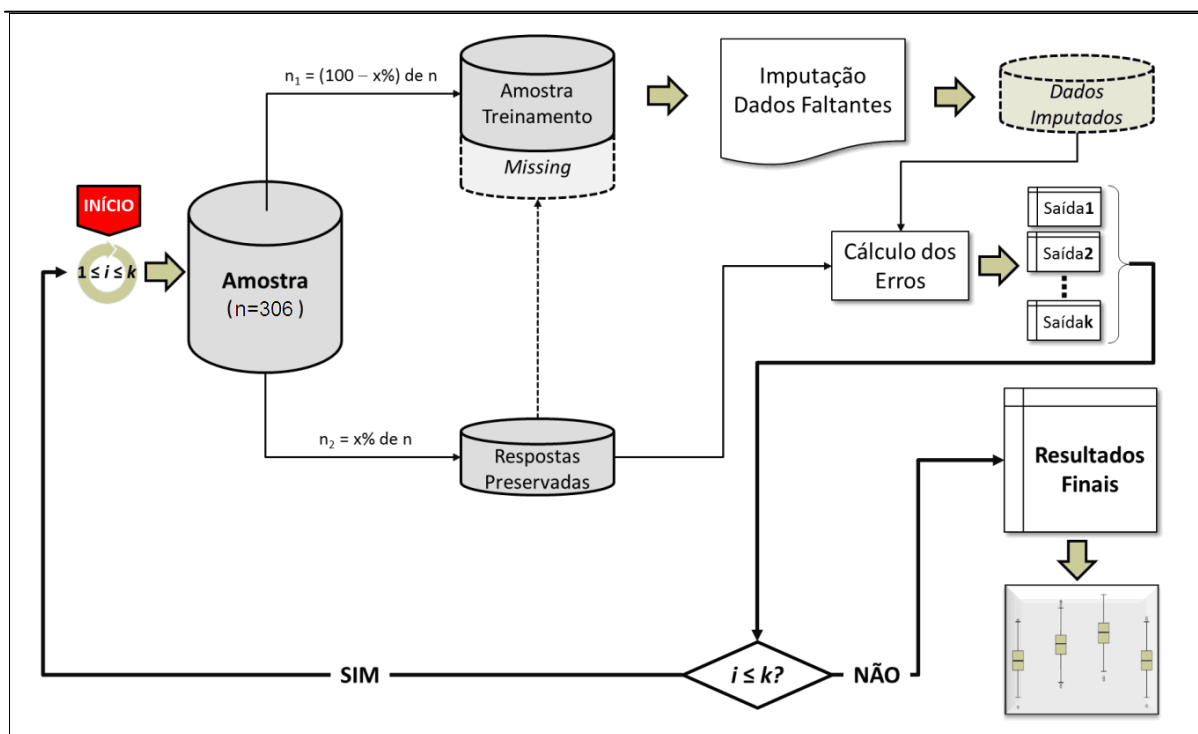


Figura 3.2: Esquema do Processo de Simulação para o Cenário 2.

Nota: k representa o total de simulações;
 $x\%$ representa o percentual de dados faltantes.

Como no Cenário 1, construiu-se para cada uma das 20 estruturas diferentes, um laço de repetição de 500 simulações, gerando 10.000 amostras distintas para análise de desempenho preditivo. A disposição dos valores de erros e os resultados finais foram apresentados por *Box-plot* para melhor análise e comparação das implicações.

3.5 ACURÁCIA DOS MÉTODOS

Depois de imputados os valores perdidos, foi examinado o desempenho das estimativas geradas pela aplicação dos métodos através de três medidas resumo (ENGELS ; DIEHR , 2003).

Para se escolher um método de previsão ou outro é importante utilizar uma medida de erro com a finalidade de se encontrar aquele método que melhor atenda os resultados esperados. Existem na literatura diversas medidas de erro disponíveis dentre as quais pode-se citar comparação gráfica, diagramas de dispersão, coeficiente de correlação, erro quadrático acumulado, raiz do erro médio quadrático

(RMSE), erro percentual médio absoluto (MAPE), entre outros. Neste trabalho em específico foram usadas duas medidas do erro, RMSE e MAPE, e o percentual de acerto da imputação comparado aos dados reais, para avaliar as previsões realizadas. O cálculo das medidas utilizadas pode ser visualizado nas equações abaixo. Para o RMSE tem-se que:

$$RMSE = \left(\sum \frac{(y - \hat{y})^2}{m} \right)^{\frac{1}{2}} \quad (12)$$

Onde $\hat{y}$ é o valor imputado, y é o valor verdadeiro, e m é o número de dados faltantes.

Essa medida demonstra o quão perto os valores estimados estão os verdadeiros valores (KALTON, 1993). O RMSD tende a penalizar outliers porque a medida de diferença é quadrática. Esperou-se estimativas baseadas em valores conhecidos do próprio indivíduo para superar aqueles que usam apenas os dados ou relacionamentos de outras pessoas.

A segunda medida resumo utilizada foi o Erro Percentual Absoluto Médio (MAPE), que é a média da diferença absoluta entre os valores previstos e atuais, expressa em percentagem. Assim, se existem previsões e valores reais para n períodos, o erro médio percentual absoluto é:

$$MAPE = \frac{\sum_{i=1}^s \left| \frac{O_i - E_i}{O_i} \right|}{s} \times 100 \quad (13)$$

onde s é o tamanho do arquivo de valores faltantes (5%, 10%, ..., 40%), O_i corresponde aos valores observados e E_i representa os valores estimados.

A última medida utilizada foi a porcentagem de acerto, medindo a frequência de acertos de cada amostra.

A variável ou item utilizado pelo banco real para a simulação das imputações e utilizações dos métodos, foi a F4 (*O senhor pratica atividade física? Com que frequência?*), pelo fato de que tal variável tem um índice de 100% de completude.

Foi observado que o item F4 apresentava uma frequência superior de uma categoria de resposta em relação às outras ("1" - 52,8%), como essa particularidade poderia sugerir um método de imputação para apenas esse fato, construiu-se um processo de simulação que obteve homogeneidade de respostas. Depois de construída, passou-se a simulação de dados faltantes com respectivas imputações de acordo com o utilizado pela variável real.

3.6 CRITÉRIOS DE DECISÃO

A confrontação dos métodos foi disposta pela apresentação dos resultados em Box Plot, uma vez que tal método visa comparar tanto a precisão quanto a dispersão de diferentes conjuntos de dados. A técnica Box Plot (ou diagrama em caixa) é útil para revelar o centro, a dispersão e a distribuição dos dados, além da presença de *outlier* (DANIEL, 2009).

O método que apresentar menor dispersão dos dados observados em relação aos estipulados, ou seja, o que minimiza as diferenças será considerado como aquele de melhor desempenho, igualmente relacionado com aquele que obteve melhor porcentagem de acertos. Essa é uma análise de sensibilidade de forma tradicional.

Desta forma, sugere-se que o método com melhor desempenho seja utilizado também para a imputação dos dados secundários (variáveis antropométricas).

3.7 ORGANIZAÇÃO DO BANCO DE DADOS

Os dados foram organizados a partir de uma máscara criada na planilha eletrônica do Microsoft Office Access 2007. As variáveis selecionadas foram transferidas para um banco de dados no Microsoft Office Excel 2007, e posteriormente para o pacote *Statistical Package for the Social Sciences* (SPSS 11.0) e o pacote Statistical Analysis System (SAS®).

3.8 CONSIDERAÇÕES ÉTICAS DA PESQUISA

O projeto PAES (2008) atendeu às normas e recomendações emanadas pela Comissão Nacional de Ética em Pesquisa (CONEP), para a pesquisa envolvendo seres humanos (Resolução 196/96 do Conselho Nacional de Saúde/ MS), que incorpora, sob a ótica do indivíduo e das coletividades, os quatro referenciais básicos da bioética: autonomia, não maleficência, beneficência e justiça, entre outros, e visa assegurar os direitos e deveres que dizem respeito à comunidade científica, aos sujeitos da pesquisa e ao Estado. Tendo sido aprovado pelo Comitê de Ética do CCS (Centro de Ciências da Saúde) da UFPB para sua execução (ANEXO 3), além de ter aprovação prévia da Secretaria Municipal de Saúde de João Pessoa (ANEXO 4). As entrevistas foram coletadas somente após a assinatura pelos participantes do Termo de Consentimento Livre e Esclarecido, o qual esclareceu aos mesmos os objetivos e assegurou o sigilo da sua identificação e o pleno direito de retirada da participação da pesquisa a qualquer instante.

Capítulo 4

4. RESULTADOS

Este capítulo descreve e analisa os resultados de tal estudo, que são apresentados em três momentos distintos. O primeiro momento prioriza os dados secundários colhidos nas segundas vias dos cadastros do HIPERDIA dos usuários hipertensos adicionado de dados dos prontuários dos respectivos usuários, no município de estudo. O segundo momento descreve os cenários de dados faltantes dos itens colhidos dos questionários sobre a satisfação da qualidade do serviço. O estudo de simulação e análise da acurácia ou performance dos métodos de imputação serão descritas no terceiro momento.

4.1. CENÁRIO DOS DADOS PERDIDOS NO HIPERDIA

A Tabela 4.1 descreve as frequências das variáveis preenchidas e não preenchidas conforme os registros dos cadastros do HIPERDIA dos usuários hipertensos no município de João Pessoa em 2006/2007.

Observa-se que as variáveis sócio-demográficas foram as maiores prejudicadas com a falta de dados, sendo a idade a de maior porcentagem de perda de dados, embora pudesse ser presumida analisando a data de nascimento.

O que primeiramente chama a atenção é que não houve nenhuma variável com total preenchimento. De fato, o máximo de preenchimento foi de 97% nas medidas pressóricas. Embora os percentuais das medidas pressóricas alcancem cerca de 3% de dados faltantes, é inadmissível que o programa omita tal informação na principal variável desfecho. Na verdade, espera-se que a porcentagem de dados faltantes seja nula.

Sexo e Situação Conjugal são variáveis que ultrapassaram 15% de não resposta, valores considerados elevados. É importante identificar que as variáveis do bloco de “fatores de risco” também apresentaram frequências maiores que 5% de informações perdidas. Tais variáveis são extremamente relevantes para o entendimento das associações com as doenças cardiovasculares, segundo inúmeros estudos sobre a temática (DALLACOSTA; DALLACOSTA; NUNES, 2010; ZILMER, et al, 2010; COSTA, 2003).

Tabela 4.1: Completude no preenchimento das variáveis do cadastro HIPERDIA em João Pessoa/PB, 2006/2007.

Informações do HIPERDIA	Dados				Total	
	Preenchidos		Faltantes			
	n	%	n	%	n	%
Dados Antropomédicos						
Pressão Arterial Sistólica	331	97,1	10	2,9	343	100
Pressão Arterial Diastólica	331	97,1	10	2,9	343	100
Peso	330	96,2	13	3,8	343	100
Altura	327	95,4	16	4,6	343	100
Cintura	319	93,0	24	7,0	343	100
Informações Sociodemográficas						
Sexo	270	78,8	73	21,2	343	100
Idade	20	5,9	323	94,1	343	100
Situação Familiar/Conjugal	291	84,9	52	15,1	343	100
Escolaridade	311	90,7	32	9,3	343	100
Raça/Cor	312	91,0	31	9,0	343	100
Fatores de Risco e Doenças Concomitantes						
Antecedente Familiar Cardiovascular	317	92,7	25	7,3	343	100
Diabetes Tipo 1	321	93,6	22	6,4	343	100
Diabetes Tipo 2	318	92,2	25	7,8	343	100
Tabagismo	319	93,0	24	7,0	343	100
Sedentarismo	315	91,9	28	8,1	343	100
Sobrepeso/Obesidade	318	92,7	25	7,3	343	100
Infarto Agudo do Miocárdio	316	92,1	27	7,9	343	100
Outra Coronopatia	317	92,4	26	7,6	343	100
AVC	317	92,4	26	7,6	343	100
Pé Diabético	316	92,1	27	7,9	343	100
Amputação por Diabetes	318	92,7	25	7,3	343	100
Doença Renal	311	90,7	32	9,3	343	100

Fonte: HIPERDIA (2006/2007)

A Tabela 4.2 demonstra as variáveis colhidas dos prontuários dos usuários no ano de 2009. Tal fonte determina o tipo de acompanhamento:

Tabela 4.2: Completude no preenchimento das variáveis no prontuário dos usuários em João Pessoa/PB, 2009.

usuários em São Paulo, 2008.							
Informação do Prontuário	Dados				Total		
	Preenchidos		Faltantes				
	N	%	n	%	N	%	
Etilismo		0	0	343	100	343	100
Menopausa		1	0	342	100	343	100
Uso de anti-contraceptivo hormonal		1	0	342	100	343	100
Fonte: Paes (2008)							

Nota-se que não há disponibilização dessas informações no prontuário, tornando difícil o diagnóstico e busca ativa desses usuários pelas unidades de saúde com relação aos serviços de saúde mental e de saúde da mulher, por exemplo, tais variáveis tendem a serem eliminadas das análises do estudo pelo elevado número de informações faltantes.

Ainda observando as informações dos cadastros do HIPERDIA dos usuários sujeitos da pesquisa, obtiveram-se informações descritivas dos cadastrados que continham pelo menos um dado faltante em sua ficha, demonstrado na Tabela 4.3.

Tais variáveis foram consideradas “chaves” para a construção do perfil dos usuários com maior índice de incompletude, resultando na informação do risco de se ter uma ficha com maior prejuízo de perda ou não preenchimento de informações do HIPERDIA.

Verificou-se um elevado índice de indivíduos não acompanhados (68,1%) e do sexo masculino (71,4%) com dados faltantes. Notavelmente, o sexo masculino é responsável por altas taxas de não acompanhamento dos serviços de atenção primária de saúde, segundo diversos estudos relativos à adesão do usuário ao programa HIPERDIA (SILVA, 2011).

O Distrito 3 apresentou um maior número de cadastro que apresenta algum item incompleto (32,8%) seguido do Distrito 1 (25%) e 2 (21,1%).

No item “Situação Conjugal”, os usuários que convivem com companheiro (a) e filho(s) apresentaram um índice bem superior de cadastro com item(s) incompleto (s) que nas demais situações, sobrepondo-se aos demais: 46,4%.

Tabela 4.3: Informações descritivas dos usuários com pelo menos um dado ausente no cadastro HIPERDIA em João Pessoa/PB, 2006/2007.

Informações do Hiperdia	n	%	Total (%)
<i>Acompanhamento</i>			
Não acompanhado	225	68,1	
Acompanhado	106	31,9	100
<i>Distrito Sanitário</i>			
1	83	25,0	
2	70	21,1	
3	109	32,8	
4	42	12,7	
5	28	8,4	100
<i>Sexo</i>			
Masculino	236	71,4	
Feminino	95	28,6	100
<i>Situação Conjugal e Familiar</i>			
Convive com companheiro e filhos	154	46,4	
Convive com companheiro sem filhos	58	17,5	
Convive com familiares sem companheiro	38	11,4	
Convive com companheiro e outros familiares	57	17,1	
Mora sozinho	18	5,4	100
<i>Escolaridade</i>			
Não sabe ler/escrever	66	20,0	
Alfabetizado	53	16,0	
Ensino Fundamental Incompleto	117	35,3	
Ensino Fundamental Completo	32	9,7	
Ensino Médio Incompleto	11	3,3	
Ensino Médio Completo	37	11,3	
Ensino Superior Incompleto	3	1,0	
Ensino Superior Completo	9	2,7	
Especialização/Residência	2	0,7	100
<i>Raça</i>			
Branca	121	36,5	
Preta	142	42,9	
Parda	58	17,6	
Amarela	9	2,7	
Indígena	1	0,3	100

Fonte: HIPERDIA (2006/2007)

No quesito “Escarlaridade” os indivíduos com baixo nível de instrução foram os mais prejudicados pela falta de dados no HIPERDIA, somando 71,3% de usuários com cadastros apresentando dados perdidos.

Quanto à “Raça”, os usuários de raça preta ou negra apresentaram-se em maior número, totalizando quase metade de todos os usuários com cadastro HIPERDIA apresentando pelo menos uma variável incompleta (42,9%).

Dessa forma, observa-se que os indivíduos mais prejudicados com a incompletude de algumas informações que compõem o HIPERDIA, apresentam as seguintes características:

- Não acompanhado;
- Pertencente ao Distrito Sanitário 3;
- Sexo Masculino;
- Que convive com Companheiro (a) e filho(s);
- De baixo nível de instrução, ou escolaridade, com no máximo Ensino Fundamental Incompleto;
- Raça negra.

4.2 CENÁRIO DOS DADOS PERDIDOS NAS DIMENSÕES DO QUESTIONÁRIO

Com relação às dimensões, para a segunda fase desse estudo, referente ao ano de 2010, são identificados valores faltantes em vários itens, como mostrado na Tabela 4.4. Não foram apresentados na tabela aqueles itens cuja completude foi total ou maior que 95%.

Segundo critérios encontrados na literatura (HARRELL, 2001; NUNES, 2007), quando é encontrada uma proporção menor que 5% de falta de dados, não há necessidade de imputação de dados, podendo-se analisar somente o banco completo. Quando é encontrada uma proporção entre 5 e 15%, recomenda-se utilizar a imputação única, sendo indicada também a imputação múltipla. Quando se tem mais que 15%, a imputação múltipla é indicada na maior parte dos casos, embora a fidedignidade dos dados imputados seja duvidosa a partir de 30% de dados faltantes.

Tabela 4.4: Completude de dimensões do estudo no município de João Pessoa/PB, 2009.

Itens das Dimensões	Dados				Total	
	Preenchidos		Faltantes			
	N	%	n	%	n	%
Acesso ao Diagnóstico						
G2 ¹	281	91,5	26	8,5	305	100
G3 ²	292	95,0	15	5,0	305	100
Acesso ao Diagnóstico						
H3 ³	274	89,3	33	10,7	305	100
Adesão/Vínculo						
I1_1b ⁴	287	93,5	20	6,5	305	100
I3 ⁵	277	90,2	30	9,8	305	100
I4 ⁶	276	89,9	31	10,1	305	100
I5 ⁷	278	90,6	29	9,4	305	100
I10 ⁸	270	87,9	37	12,1	305	100
I11 ⁹	285	92,8	22	7,2	305	100
I12_a ¹⁰	284	92,5	23	7,5	305	100
I12_c ¹¹	275	89,6	32	10,4	305	100
Elenco de Serviços						
J1 ¹²	289	94,1	18	5,9	305	100
J2 ¹³	288	93,8	19	6,2	305	100
J3 ¹⁴	285	92,8	22	7,2	305	100
J8 ¹⁵	283	92,2	24	7,8	305	100

¹ G2- Quando o senhor (a) começou a ter os sintomas da HAS (dor de cabeça, tontura), quantas vezes precisou procurar a unidade de saúde para descobrir que era hipertenso?

² Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura), procurou o serviço de saúde mais próximo da sua casa?

³ Se o(a) Sr(a) passar mal por causa da medicação ou da HAS, consegue uma consulta no prazo de 24hs na unidade de saúde que faz tratamento?

⁴ O(A) Sr(a) recebe apoio por parte: Colegas trabalho?

⁵ Se o(a) Sr.(a) tem alguma dúvida sobre o seu tratamento, consegue falar com o mesmo profissional da unidade de saúde que o atende?

⁶ Quando o(a) Sr.(a) faz alguma pergunta ao profissional da unidade de saúde sente que é compreendido?

⁷ O profissional da unidade de saúde que o(a) Sr.(a) está fazendo o tratamento responde às suas perguntas de maneira clara?

⁸ Os profissionais da unidade de saúde relacionam-se bem com as pessoas da comunidade?

⁹ O(A) Sr.(a) já pensou em mudar da unidade de saúde por causa dos profissionais?

¹⁰ Numa escala de 1 a 5, considerando 1 (pior situação) e 5 (melhor situação), qual seria o seu grau de satisfação em relação aos seguintes profissionais que o atende? Médico

¹¹ Numa escala de 1 a 5, considerando 1 (pior situação) e 5 (melhor situação), qual seria o seu grau de satisfação em relação aos seguintes profissionais que o atende? Auxiliar Enfermagem

¹² Informação sobre a HAS e seu controle?

¹³ Informação sobre os medicamentos e seus efeitos?

¹⁴ Educação em saúde (informação sobre outros temas de saúde)?

¹⁵ Os Horários de funcionamento do serviço de saúde para atendimento são sempre respeitados?

Continuação

Coordenação

L1 ¹⁶	292	95,0	15	5,0	305	100
L2 ¹⁷	292	95,0	15	5,0	305	100

Orientação para a Comunidade

N1 ¹⁸	269	87,6	38	12,4	305	100
N2 ¹⁹	232	75,6	75	24,4	305	100
N3 ²⁰	268	87,3	39	12,7	305	100

Fonte: Paes (2009)

Do total de 77 itens que compuseram as dimensões, 20 deles apresentaram falta de dados. Das oito dimensões, apenas a “F” – *Saúde do caso confirmado* e a “M” – *Enfoque na Família* apresentaram preenchimento total dos itens.

A dimensão “N” – *Orientação para a Comunidade* foi a que apresentou maior falta de preenchimento dos itens ou variáveis, sendo que a questão N2 – *Com que frequência os serviços de saúde desenvolvem ações sobre HAS com as Igrejas, Associações de Bairro, escolas, etc.*, alcançou mais que 20% de falta de dados.

Destaca-se que segundo o SUS, a participação da comunidade é defendida por lei (Lei nº8142/90) e está diretamente relacionada ao grau de consciência política e de organização da própria sociedade civil. Alerta ainda que os serviços de saúde devem identificar o usuário como membro de uma comunidade, com direitos e deveres, e não como receptor passivo de benefícios do Estado (CONASS, 2007).

Outras variáveis prejudicadas que envolvem questões inerentes à qualidade dos serviços de saúde, foram:

a) Dimensão “H” – *Adesão/vínculo*

- H3 – *Se o senhor (a) passa mal por causa da medicação ou da HAS, consegue uma consulta no prazo de 24hs na unidade de saúde que faz tratamento?;*

¹⁶ Os profissionais da unidade de saúde pegam seu prontuário/ficha do(a) Sr(a) durante a consulta?

¹⁷ Durante seu atendimento o profissional da unidade de saúde anota as suas queixas no seu prontuário?

¹⁸ Os profissionais da unidade de saúde conversam sobre a importância da sua participação e da sua família em instituições da comunidade (igrejas, associação de bairro, etc.) como apoio para resolver seus problemas de saúde?

¹⁹ Com que frequência os serviços de saúde desenvolvem ações sobre HAS com as Igrejas, Associações de Bairro, escolas, etc.?

²⁰ Os profissionais da unidade de saúde conversam sobre a influência da família/amigos/colegas no seu tratamento?

b) Dimensão “I” - Acesso ao Tratamento

- I4 - Quando o (a) Sr.(a) faz alguma pergunta ao profissional da unidade de saúde sente que é compreendido?;
- I10 - Os profissionais da unidade de saúde relacionam-se bem com as pessoas da comunidade?;
- I12_c - Numa escala de 1 a 5, considerando 1 (pior situação) e 5 (melhor situação), qual seria o seu grau de satisfação em relação aos seguintes profissionais que o atende? Auxiliar Enfermagem;

c) Dimensão “N” - Orientação para a comunidade

- N1 - Os profissionais da unidade de saúde conversam sobre a importância da sua participação e da sua família em instituições da comunidade (igrejas, associação de bairro, etc.) como apoio para resolver seus problemas de saúde? ;
- N3 - Os profissionais da unidade de saúde conversam sobre a influência da família/amigos/colegas no seu tratamento?, tendo proporção maior que 10% de dados faltantes.

Com os dois cenários descritivos sobre a perda de dados, tanto na parte secundária (HIPERDIA), quanto na primária (questionários), pode-se perceber a problemática colocada em que surge como solução viável a imputação de dados faltantes.

4.3 SIMULAÇÃO E ACURÁCIA DOS MÉTODOS

Os métodos utilizados foram comparados utilizando três medidas resumos distintas: Erro Quadrado Médio (RMSE), Erro Percentual Absoluto Médio (MAPE) e percentagem de acerto. Cada uma dessas medidas foi utilizada em cada uma das proporções de dados faltantes e para cada um dos métodos utilizados em todas as amostras com simulação de dados faltantes.

Em seguida, procedeu-se à simulação de dados faltantes para os dois cenários traçados no Capítulo 4, ou seja, simulações com o Cenário 1 e 2. No

primeiro, apenas o item F4 (variável resposta utilizada) foi utilizado e no segundo, foi feita uma simulação para uma situação de categorias de respostas homogêneas.

Antes de mostrar os resultados é importante discutir como o desempenho de um determinado método de imputação deverá ser avaliado. Um método de imputação é considerado adequado quando os valores imputados apresentam pequena variabilidade e, sobretudo, quando tais valores coincidem com os valores reais, ou seja, quando os valores estimados pelo método coincidem com os valores da variável resposta se esses tivessem sido observados.

Sob esse ponto de vista, um método de imputação que imputa muitos valores que coincidem com os valores reais da variável resposta é dito um método eficiente, já que o método acerta muitas vezes o alvo (valor real da variável resposta), fazendo um paralelo com o jogo de tiro ao alvo. Essa condição de eficiência do método necessariamente ajuda a diminuir as medidas de erro, uma vez que as distâncias entre os valores imputados e o alvo (valores reais) serão nulas.

Por outro lado, um método de imputação poderá acertar o alvo algumas vezes, mas ser muito instável (volátil). Isso significa que o método de imputação, apesar de acertar o alvo algumas vezes, apresenta uma variabilidade muito grande em relação à região dos possíveis valores que poderão ser imputados. A Figura 4.1 apresenta as relações entre a eficiência e a estabilidade no que tange à qualidade dos métodos de imputação de dados.

Segundo a Figura 4.1, o método de imputação desejável será aquele que for eficiente e, ao mesmo tempo, estável. É sob essa linha de raciocínio que os métodos de imputação serão avaliados e comparados.

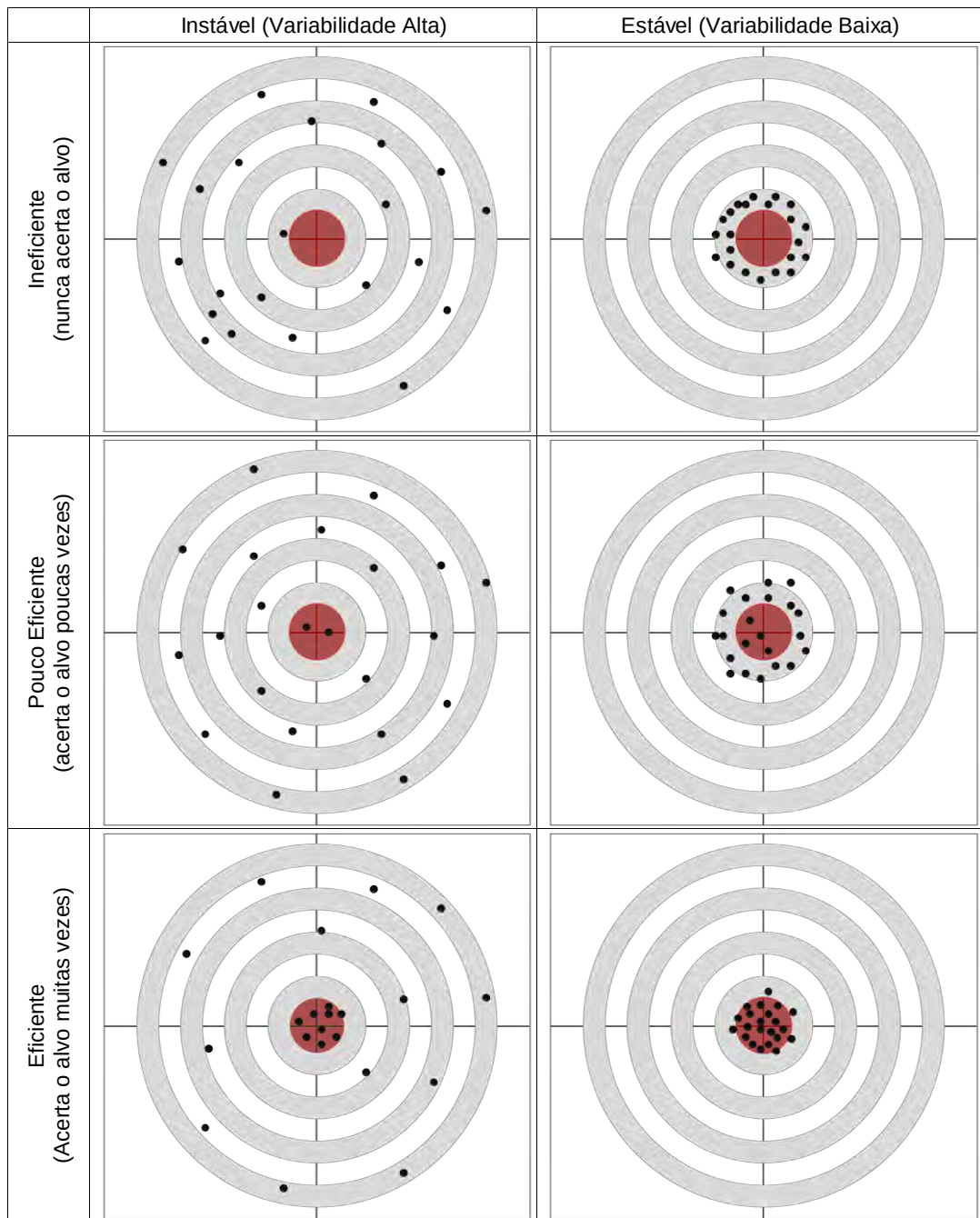


FIGURA 4.1: Avaliação dos métodos de imputação com relação à eficiência e à estabilidade

Na Tabela 4.5 são apresentadas as médias e desvios padrão das medidas de desempenho segundo os graus de dados faltantes e os quatro tipos de métodos de imputação avaliados. Esses resultados se referem à estratégia adotada para o Cenário 1.

TABELA 4.5: Médias e desvios padrão para os erros de imputação (RMSE e MAPE) e para o percentual de acerto segundo o método e proporção de valores faltantes para o Cenário 1.

Percentual de Valores Faltantes	Medida	Métodos								Total	
		Tendência Central		Hot-Deck		Máxima Verossimilhança		Regressão Logística Multinomial			
		Média	DP	Média	DP	Média	DP	Média	DP	Média	DP
5%	RMSE	1,813	0,478	1,972	0,288	2,127	0,384	2,137	0,381	2,012	0,408
	MAPE (%)	10,151	3,344	29,385	9,023	31,954	9,843	25,330	7,018	24,205	11,426
	Acerto (%)	55,333	14,396	32,109	10,512	34,533	14,735	28,267	11,708	37,588	16,643
10%	RMSE	1,861	0,230	2,000	0,232	2,161	0,267	1,861	0,230	1,971	0,269
	MAPE (%)	10,281	1,829	30,031	6,259	32,074	8,598	10,281	1,829	20,667	11,765
	Acerto (%)	54,800	7,742	30,733	8,365	33,333	9,110	54,800	7,742	43,417	14,081
15%	RMSE	1,904	0,228	2,035	0,178	2,122	0,256	1,906	0,228	1,992	0,241
	MAPE (%)	10,788	1,879	28,880	4,965	30,011	8,304	10,802	1,892	20,120	10,600
	Acerto (%)	52,533	7,889	31,956	6,520	33,956	8,376	52,400	8,082	42,711	12,462
30%	RMSE	1,909	0,119	2,045	0,124	2,128	0,137	1,911	0,118	1,998	0,155
	MAPE (%)	10,698	0,978	29,023	3,554	30,392	4,225	10,721	0,973	20,209	9,945
	Acerto (%)	53,222	4,171	32,489	4,100	34,267	3,825	53,000	4,135	43,244	10,699
40%	RMSE	1,917	0,105	2,047	0,091	2,116	0,096	1,915	0,105	1,999	0,131
	MAPE (%)	10,827	0,737	27,868	3,473	29,675	3,723	10,818	0,732	19,797	9,381
	Acerto (%)	52,450	3,115	34,083	3,013	34,400	3,512	52,200	3,043	43,283	9,598
Total	RMSE	1,881	0,269	2,020	0,197	2,131	0,248	1,946	0,253	1,994	0,260
	MAPE (%)	10,549	1,981	29,038	5,826	30,821	7,377	13,590	6,781	20,999	10,763
	Acerto (%)	53,668	8,460	32,275	7,064	34,098	8,852	48,133	12,519	42,053	13,102

Com o Cenário 1, no geral os métodos que obtiveram menor RMSE foram o de TC com média de 1,881 ($\pm 0,269$) e RLM com média de 1,946 ($\pm 0,253$).

Para o MAPE, também com menor discrepância, a TC obteve média geral de erro de 10,549 ($\pm 1,981$) e RLM com média de 13,590 ($\pm 6,781$).

Seguindo a tendência abordada anteriormente, os maiores percentuais de acertos foram encontrados nos métodos TC e RLM, com 53,668 ($\pm 8,46$) e 48,133($\pm 12,519$), respectivamente.

Observando a eficiência e estabilidade média das imputações para as diferentes proporções de dados faltantes, nota-se que, a partir de 10% de falta de dados, as imputações produzem menores erros — RMSE variando entre 1,971($\pm 0,269$) e 1,999($\pm 0,131$); MAPE variando entre 20,667($\pm 11,765$) e 20,999($\pm 10,763$) — ou seja, indicam maior estabilidade. Quanto à eficiência, paralelamente aos resultados há uma maior performance a partir de 10% de dados faltantes, com a percentagem de acerto variando entre 42,053($\pm 13,102$) e 43,417($\pm 14,081$).

Analisando os quatro métodos de imputação, percebe-se que a TC obteve menores médias de erros — RSME de 1,881($\pm 0,269$) e MAPE de 10,549($\pm 1,981$) — sendo mais estável e mais eficiente, pois a média da percentagem de acerto foi maior: 53,668, seguido pela RLM com 48,133.

Para 5% de dados faltantes, método da TC obteve maior estabilidade e eficiência: RMSE de 1,813($\pm 0,478$), MAPE de 10,151($\pm 3,344$) e percentual de acerto de 55,333.

Com 10%, os métodos de TC e RLM obtiveram iguais performances, tanto na eficiência, quanto na eficácia.

Já para 15%, o método da TC mostrou maior estabilidade e eficiência — RMSE de 1,904($\pm 0,228$), MAPE de 10,788($\pm 1,879$) e percentual de acerto de 52,533.

Com 30%, foi mais estável e eficiente, com média de 1,909($\pm 0,119$) para o RMSE e 10,698($\pm 0,978$) para o MAPE, a média do percentual de acerto foi de 53,222.

Para 40% de falta de dados, o método mais estável foi o da RLM — RMSE de 1,915($\pm 0,105$) e MAPE de 10,818($\pm 0,732$), mas a TC ainda foi a mais eficiente pois acertou em média 52,200% das imputações.

Observando os erros para as proporções, avaliou-se o RMSE para 5% de dados faltantes, tendo média de 2,012($\pm 0,408$) diminuindo para 1,971($\pm 0,269$) com 10% de dados faltantes. Abaliza-se que a partir desse limiar de dados perdidos, a média do RMSE para as outras proporções manteve a mesma tendência, variando entre 1,971 ($\pm 0,269$) e 1,999 ($\pm 0,131$).

Com o MAPE, obteve-se uma maior média de erro com a proporção de 5% de imputações para dados faltantes ($24,205 \pm 11,426$). Seguiu-se a tendência do RMSE que manteve índices semelhantes para as demais proporções, variando entre 19,797 ($\pm 9,381$) e 20,667 ($\pm 1,765$), mostrando que as imputações tiveram melhor êxito com mais de 10% de dados faltantes.

O mesmo ocorreu com a medida percentual de acerto que passou de uma média de 37,588 (para 5%) para 43,283 (para 40%).

Analisando as simulações para o Cenário 2, com valores de frequência homogênea, a Tabela 4.6 mostra as médias e desvios padrão das medidas para cada proporção de dados faltantes e métodos de imputação. Observa-se claramente, para a medida RMSE, que o método com melhor performance em todos os tipos de percentual de dados faltantes foi a RLM com uma média de $2,211 (\pm 0,249)$. Entretanto, o MAPE apresentou menores médias no método MV para as proporções 5%, 10% e 30% de dados faltantes (média de $39,6 (\pm 7,267)$).

Quanto à percentagem de acertos, o método RLM demonstrou uma melhor performance (eficiência), apresentando média de $24,379 (\pm 7,339)$. É importante apontar que para 40% de dados perdidos, a RLM foi a que obteve melhor êxito na imputação (estabilidade), mostrando menores índices de RMSE e MAPE.

Avaliando a medida RMSE para as diferentes proporções, a menor média foi detectada com imputações para 5% de dados perdidos. Esse índice cresce discretamente à medida que se aumenta as proporções de dados faltantes.

O MAPE apresentou as mesmas características do RMSE, oferecendo menor média de erro entre os métodos, à medida que se aumenta as proporções de falta de dados.

TABELA 4.6: Médias e desvios padrão para os erros de imputação (RMSE e MAPE) e para o percentual de acerto por método e proporção segundo os valores faltantes para o Cenário 2.

Percentual de Valores Faltantes	Medida	Métodos								Total	
		Tendência Central		Hot-Deck		Máxima Verossimilhança		Regressão Logística Multinomial			
		Média	DP	Média	DP	Média	DP	Média	DP	Média	DP
5%	RMSE	2,420	0,578	2,323	0,359	2,403	0,463	2,137	0,381	2,321	0,464
	MAPE (%)	39,598	12,715	38,405	10,587	37,555	10,796	37,995	10,527	38,388	11,134
	Acerto (%)	7,356	2,066	13,659	7,881	19,130	11,121	28,267	11,708	18,474	12,120
10%	RMSE	2,455	0,563	2,362	0,269	2,325	0,237	2,269	0,238	2,353	0,358
	MAPE (%)	42,740	15,807	41,077	9,360	39,744	8,351	40,529	8,244	41,023	10,873
	Acerto (%)	8,163	3,196	9,861	5,144	17,667	7,537	22,400	7,709	14,602	8,466
15%	RMSE	2,469	0,564	2,361	0,185	2,397	0,243	2,215	0,228	2,360	0,350
	MAPE (%)	43,310	15,881	39,995	6,688	40,837	6,587	41,193	6,327	41,334	9,755
	Acerto (%)	10,089	3,019	11,467	4,196	16,933	5,272	23,778	5,202	15,567	7,018
30%	RMSE	2,579	0,540	2,392	0,118	2,379	0,141	2,243	0,159	2,398	0,317
	MAPE (%)	45,264	13,602	41,053	3,924	40,099	3,624	41,123	5,475	41,885	7,997
	Acerto (%)	12,333	1,476	12,844	3,420	17,156	3,383	23,600	3,665	16,483	5,481
40%	RMSE	2,523	0,528	2,420	0,109	2,419	0,145	2,190	0,161	2,388	0,313
	MAPE (%)	42,137	15,773	40,695	3,994	39,765	4,410	38,827	4,291	40,356	8,719
	Acerto (%)	13,067	1,403	13,383	3,968	16,183	3,304	23,850	4,005	16,621	5,476
Total	RMSE	2,489	0,553	2,372	0,229	2,385	0,272	2,211	0,249	2,364	0,365
	MAPE (%)	42,610	14,811	40,245	7,434	39,600	7,267	39,933	7,366	40,597	9,826
	Acerto (%)	10,473	3,197	12,209	5,187	17,386	6,711	24,379	7,339	16,280	7,997

Em uma análise geral das medidas usadas pelo conjunto de métodos de imputação, observa-se que a média do percentual de acerto não se altera com o aumento da porcentagem de dados perdidos, exceto para 5%. Ou seja, acima de 10% de dados faltantes, os métodos continuam sendo eficientes.

Neste caso, os métodos de imputação para a situação de particular homogeneidade nas frequências de respostas diferentes, teve como melhor performance em ordem decrescente: RLM, MV, HD e TC.

Como ferramenta útil no diagnóstico da utilização das diversas medidas para avaliação da acurácia dos métodos nos diversos cenários de proporção de dados perdidos, ilustrou-se tais resultados através de *Boxplot*, são exibidas nas Figuras 4.1, 4.2 e 4.3.

O Box-plot ou gráfico de caixas dá uma idéia da posição, dispersão, assimetria, caudas e possíveis dados discrepantes no erro quadrado médio. Observa-se que a distribuição dessa medida resumo é simétrica usando todas as metodologias de imputação.

A menor variabilidade obtida para a simulação com a variável real, ou seja, no Cenário 1, foi obtida com 40% de dados faltantes, observando os valores para RMSE (Figura 4.1). A variabilidade das médias do RMSE com proporções acima de 15% tendeu a permanecerem mais estreitas e com menores medianas. Pode-se concluir que a maior diferença de erro foi alcançada com a predição de interações faltantes através da Máxima Verossimilhança, enquanto a de menor erro (mais próximo de zero “0”) foi usando estimativas para Tendência Central e Regressão logística Multinomial.

A segunda medida resumo utilizada foi o MAPE (Erro Percentual Médio Absoluto) (Figura 4.2). As estimativas de menor erro foram as de TC e RLM, assim como foi verificado pelo MAPE.

A Figura 4.3 mostra o Erro Percentual Médio Absoluto. As de menores variabilidades e taxas de erros foram observadas para TC e RLM, com maiores fidedignidades na imputação. Esses métodos apresentaram visivelmente maior condensabilidade dos dados comparado ao HD e MV em todas as proporções de dados faltantes.

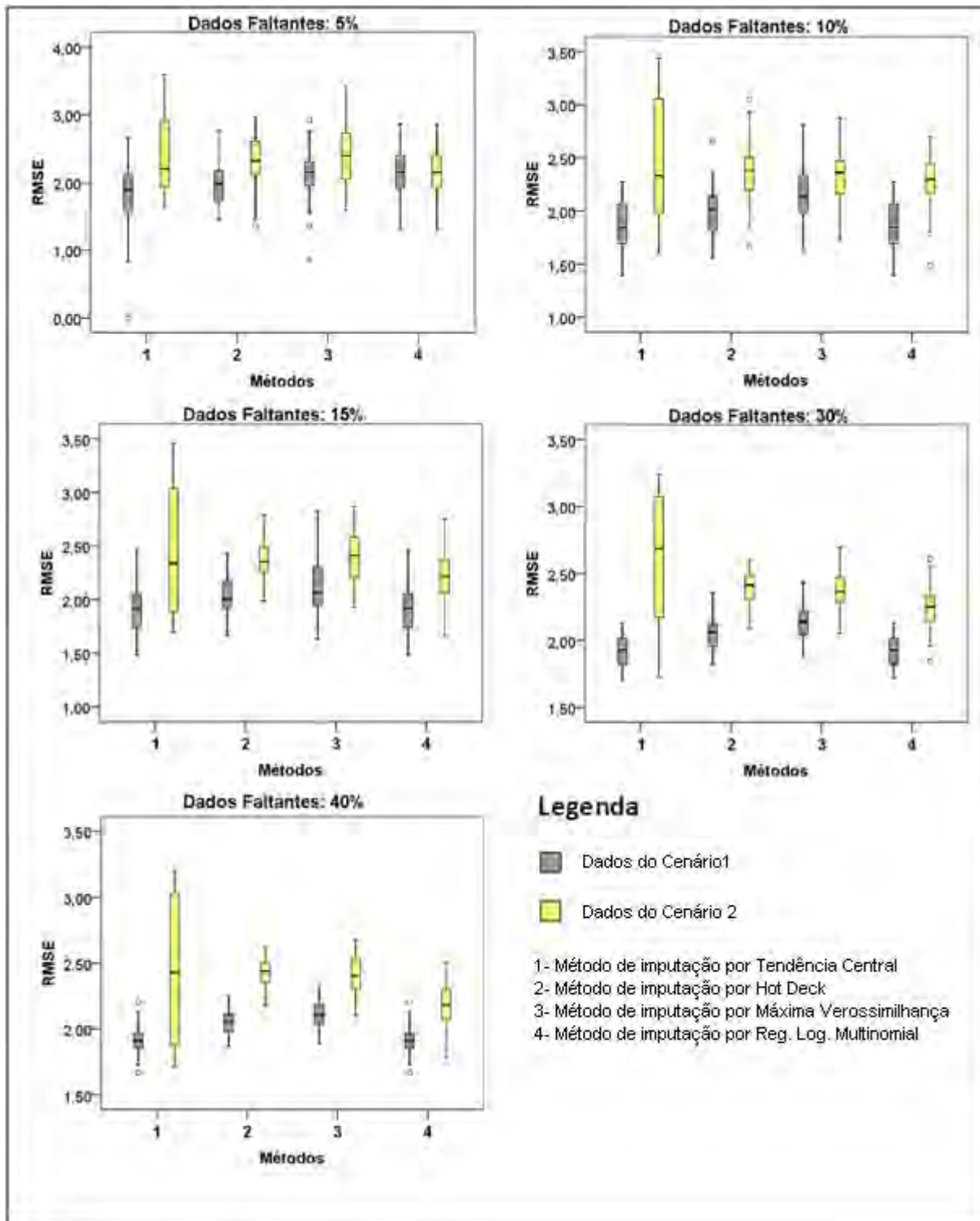


FIGURA 4.2: Distribuição amostral do RMSE por método e cenários, segundo as proporções de valores faltantes

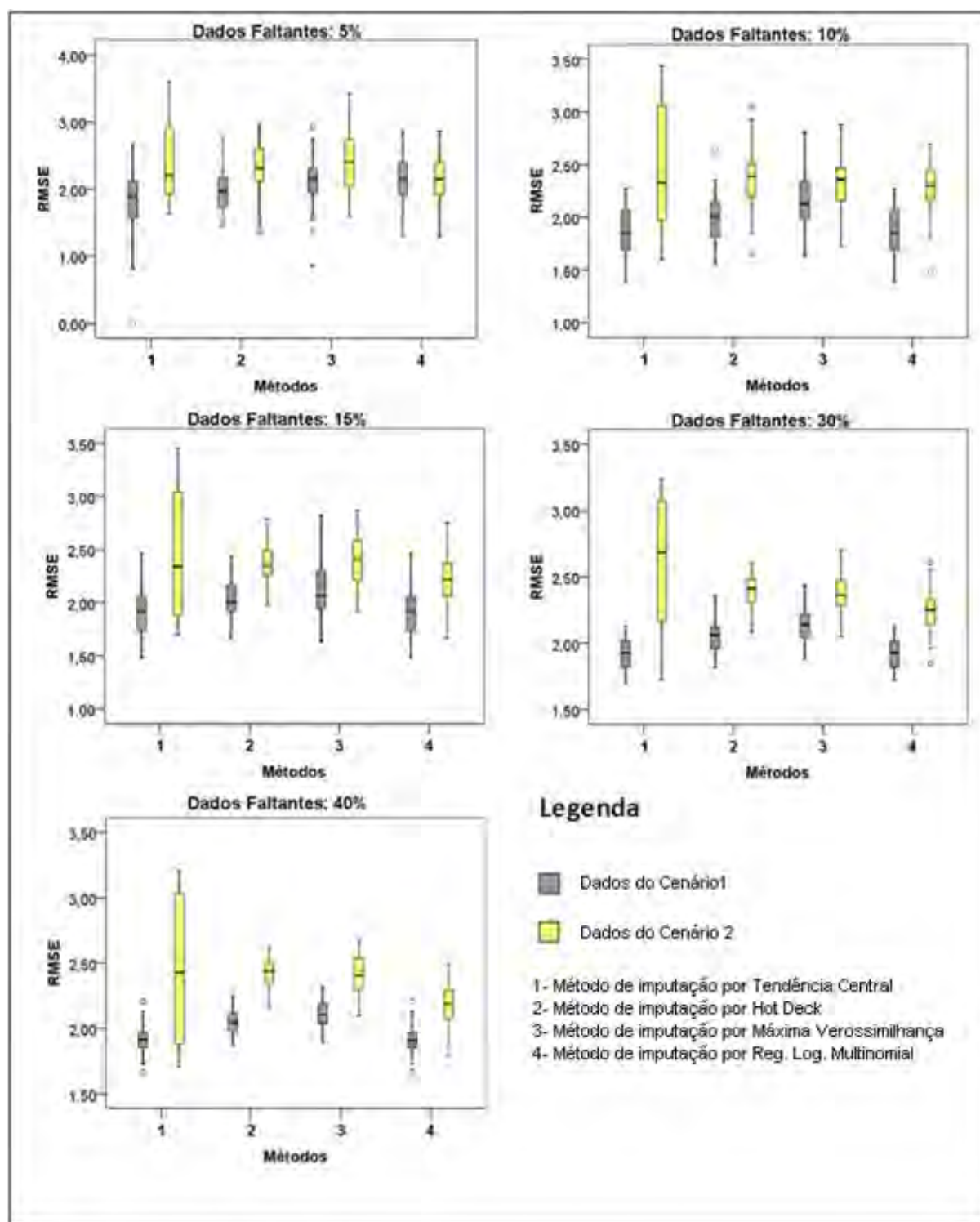


FIGURA 4.3: Distribuição amostral do MAPE por método e cenários, segundo as proporções de valores faltantes

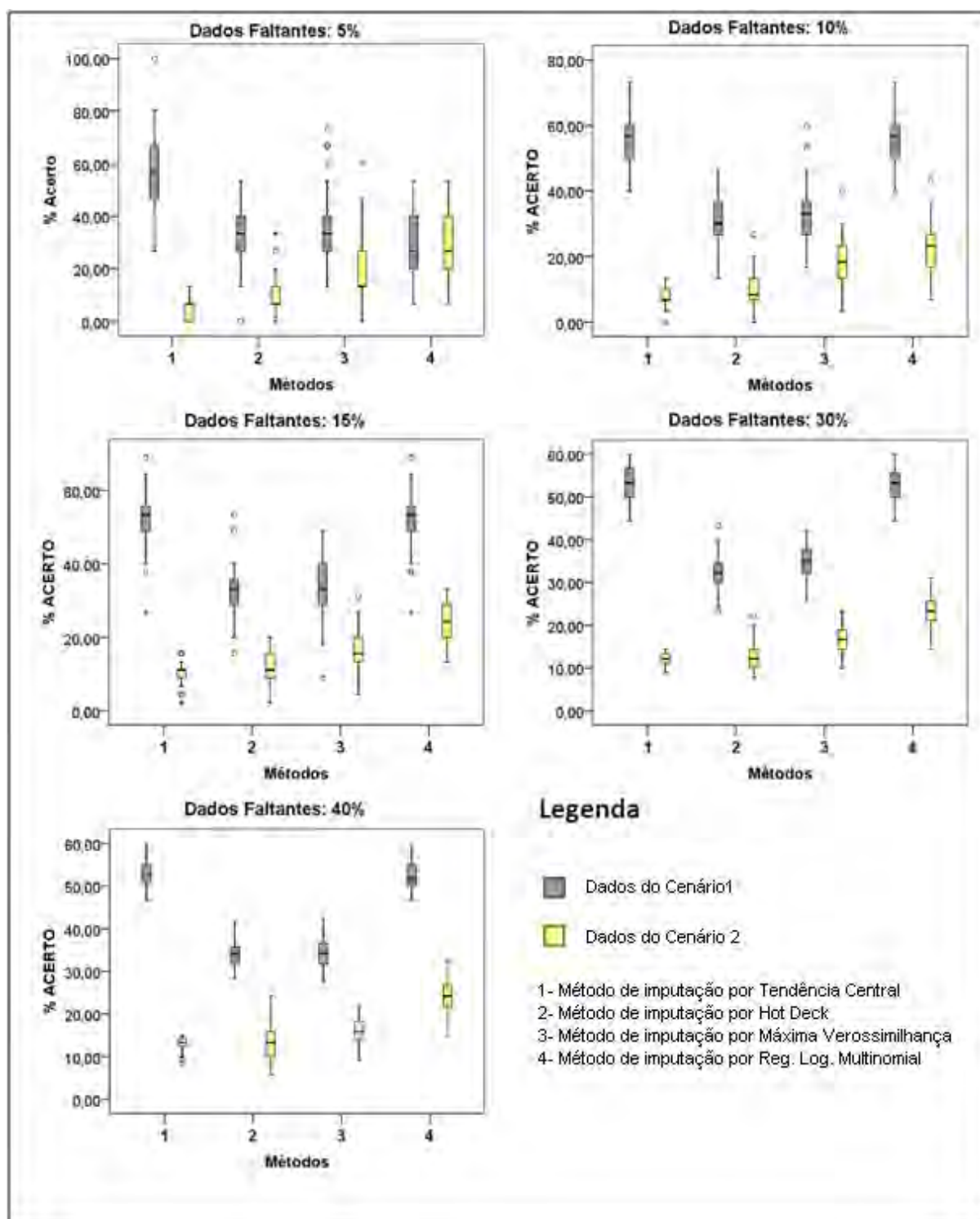


FIGURA 4.4: Distribuição amostral do Percentual de Acerto por método e cenários, segundo as proporções de valores faltantes

Quanto ao Cenário 2, o método de TC obteve maior variabilidade de erro, tanto para o RMSE quanto para o MAPE. O percentual de acerto foi maior para o método de RLM.

Dessa forma, verificou-se que não houve diferenças importantes entre a performance de imputações com 5%, 10%, 15%, 30% ou 40% de dados faltantes em ambos os cenários, ou seja, até 40% de dados faltantes. Os métodos RLM e TC foram os que apresentaram melhor desempenho nos dois cenários estudados.

Para exemplificar as possíveis situações ou itens que apresentam as características de um ou outro cenário, a Tabela 4.7 e a Tabela 4.8 apresentam as frequências e percentuais das respostas ou valores dos diversos itens que compõem o questionário, para o Cenário 1 e Cenário 2, respectivamente.

Tabela 4.7: Percentagem das frequências das respostas aos itens das dimensões características ao Cenário 1

Itens	Respostas					
	0	1	2	3	4	5
G2	9,6	12,5	5,3	17,4	11,0	43,8
G3	7,5	15,4	3,8	11,0	4,5	57,9
H3	14,2	18,2	3,6	14,2	6,2	43,5
I1_1b	77,0	7,0	0,7	2,1	0,3	12,9
I3	17,7	1,8	2,5	7,2	4,3	66,4
I4	15,9	2,2	0,4	5,4	5,8	70,3
I5	15,5	2,5	0,7	4,0	4,0	73,4
I10	9,3	1,1	1,1	8,1	13,0	67,4
I11	2,8	1,8	2,8	6,0	4,6	82,1
I12_a	16,5	2,1	4,6	7,4	5,6	63,7
I12_c	4,7	2,2	2,5	4,7	13,8	72,0
J1	8,7	9,0	2,8	13,1	7,6	58,8
J2	8,7	23,3	2,8	8,3	7,3	49,7
J3	8,1	14,4	2,1	10,9	7,7	56,8
J8	5,7	2,8	1,4	11,3	4,6	74,2
L1	7,2	2,1	0,0	2,7	1,0	87,0
L2	8,2	2,7	0,3	2,1	2,7	83,9
Total	100,0	100,0	100,0	100,0	100,0	100,0

A Tabela 4.7 mostra apenas os itens passíveis de imputação descritos na seção 5.2. desse trabalho, que denota o cenário de dados faltantes nos questionários. Desta forma, a tabela comprova que os itens apresentam apenas uma resposta predominante sobre todas as outras. Das 17 variáveis, a grande maioria manifesta um comportamento de resposta para o valor “5”, determinando um alto grau de satisfação em relação ao serviço de saúde da qual ele é usuário.

Ilustrando o segundo cenário, encontraram-se padrões homogêneos de respostas em alguns itens do questionário, conforme apresenta a Tabela 5.8.

Tabela 4.8: Percentagem das frequências das respostas aos itens das dimensões característicos ao Cenário 2

Itens	Respostas					
	0	1	2	3	4	5
H13	20,9	27,6	6,7	17,5	5,4	21,9
N1	16,0	31,6	5,6	20,8	7,1	19,0
N2	26,3	11,2	2,6	23,7	11,6	24,6
N3	14,2	42,2	4,1	15,7	4,9	19,0
Total	100,0	100,0	100,0	100,0	100,0	100,0

Observa-se que dos itens apresentados, apenas N1, N2 e N3 são citados na seção 5.2. como passíveis de imputação. O item H13 exemplifica a problemática ressaltada com a simulação para o Cenário 2, sendo que em outras variáveis do questionário, dois ou três valores de resposta têm destaque dentre todos os itens.

Capítulo 5

5. DISCUSSÃO

O HIPERDIA é uma importante ferramenta para avaliação de pacientes hipertensos e diabéticos cadastrados na Estratégia Saúde da família. É um sistema de informações que apresenta dados indispensáveis para a avaliação da situação de saúde de tais usuários. Além disso, a utilização e análise desses dados é passo importante e fundamental para a sua melhoria, permitindo que gestores tenham em mãos ferramentas mais adequadas para subsidiar decisões e o meio científico para avaliar políticas de prevenção e controle dos agravos em saúde. No entanto, muitos dos preditores têm observações faltantes.

Frequentemente, os pacientes têm vários preditores sem informações, além de não ser incomum alguns itens importantes terem mais que 50% de dados faltantes.

Como observado nesse estudo, todas as variáveis encontradas nos cadastros HIPERDIA dos usuários examinados para o município de João Pessoa apresentaram incompletude. Indivíduos com perfil caracterizado por ser: homem, negro, de escolaridade baixa, residentes no Distrito 3 e que convivem com companhia e filhos predominaram com dados incompletos no cadastro

Quanto à falta de informações sobre a variável “Raça”, (que apresentou 42,9% de incompletude de cadastro para negros), Lopes et al. (2002) estudou a relação de tal variável com as doenças cardiovasculares, notando que os resultados mostravam tendência de HA mais grave, relacionando-a também com lesões renais, em negros e pardos comparado com os brancos. O III Consenso Brasileiro de Hipertensão Arterial (1998) constatou que a variável raça é um dos itens mais importantes da história clínica do paciente com hipertensão.

Com relação ao pouco nível de instrução dos usuários com déficit de preenchimento no cadastro, diversos autores trazem a variável sócio-demográfica “Escolaridade” como definição e construção para o perfil dos hipertensos. A baixa escolaridade apresenta associação significativa com a Hipertensão Arterial (MONTEIRO et al, 2005; AVEZUM et al , 2005; LESSA et al, 2004).

Lessa et al. (2004) concluíram em estudo sobre HA e DM que raça negra, baixa escolaridade e baixo nível socioeconômico são descritos como de elevado risco cardiovascular na literatura internacional”.

Mesmo o estudo apontando o sexo masculino como de maior índice de incompletude dos dados, tal variável apresenta uma dura realidade do perfil de incompletude, visto que as unidades locais de saúde trabalham com demanda espontânea, assim, as mulheres procuram mais os serviços de saúde e, conseqüentemente, maiores são as suas chances de serem inscritas no HIPERDIA, sendo, portanto, maiores as chances de terem o cadastro mal preenchido. Nesse trabalho foi encontrado o inverso, reafirmando a dificuldade dos serviços frente à tais questões de gênero (TRAVASSOS et al., 2002).

Quanto ao estado civil, não existem informações e estudos que demonstrem claramente sua relação com a HA, embora Avezum et al. (2005) sinalizem a hipótese de que o nível de estresse mental pode ser alterado dependendo do estado civil, relacionando então a um potencial risco associado ao Infarto Agudo do Miocárdio. Tais autores ainda citam que tal hipótese foi confirmada em estudos europeus com a associação de diminuição da expectativa de vida por eventos coronarianos em pessoas solitárias ou com a experiência de viuvez.

Desta forma, percebe-se que informações e dados sobre variáveis sócio-demográficas não devem faltar em qualquer investigação clínica realizada por qualquer profissional de saúde ou pela avaliação de situações de saúde abordadas por gestores, resultando em práticas e política de saúde. Tais informações ajudam na definição de fatores de risco associados, sendo estes mutáveis ou não-mutáveis.

O Plano de Reorganização da Atenção à Hipertensão Arterial Sistêmica e ao *Diabetes Mellitus* no Brasil foi um importante passo, porém necessita de constante avaliação para medir os impactos gerados na população.

O inquérito sobre a satisfação dos usuários com os serviços e atenção prestada mostrou algumas facetas sobre a perda de dados. Foram feitas comparações dos possíveis métodos de imputação para solucionar a falta de dados no estudo sobre a satisfação dos usuários hipertensos na ESF.

Segundo a comparação dos métodos pelas medidas resumos ou erros: RMSE e MAPE, e pelo percentual de acerto, os métodos de melhor desempenho e fidedignidade para imputar dados categorizados nas condições particulares de maior frequência de um tipo de resposta dos itens, seria a medida de Tendência Central e a Regressão Logística Multinomial. Esses métodos apresentaram melhores

resultados considerando o menor erro (maior proximidade de “0”), a mediana da distribuição, além de minimizar a dispersão.

Baseado nesses critérios encontrou-se que a predição de interações em experimentos ou estudos com dados categorizados como medidas de satisfação, é mais recomendável imputar com medidas de TC por apresentar maior praticidade, entendimento e rapidez de resultados.

Já para variáveis ou amostras com homogeneidade de tipos de respostas (simulação com o Cenário 2), sugere-se que o método utilizado seja a RLM, por apresentar performance superior aos demais métodos.

Abaliza-se primeiramente que o estudo é bastante particular, ao se propor a verificar a satisfação de usuários no serviço de saúde. Dessa forma, a categorização dos dados em Escala Likert para essa análise limita a possibilidade de utilização de diversos métodos para imputação de dados, inclusive a Imputação Múltipla. Sugerem-se assim métodos mais simples para estimar dados faltantes em estudos com tais características, os quais são bastante comuns, principalmente na área de saúde.

Estudos sugerem uma maior fidedignidade de resultados na utilização da Imputação Múltipla em detrimento à por Substituição Única, embora seja um método de maior complexidade computacional (RUBIN, 1996; NUNES, 2009; NUNES, 2010; HARRELL, 2001; SCHAFER, 1999; WHITE; ROYSTON, 2007). Apesar de serem menos complexos em aplicabilidade, nota-se uma escassez de estudos epidemiológicos que utilizam métodos de Substituição Única para imputar.

A variável utilizada para o estudo de simulação e posteriormente comparação das estimativas, no Cenário 1, obteve uma maior quantidade de resposta “1” definido no estudo como não satisfação ou nula/baixa frequência de hábitos saudáveis, sobrepondo-se em mais de 50% em quantidade de valores comparado às outras possíveis respostas. Da mesma forma, verificou-se que a maioria dos diversos outros itens das dimensões desse estudo, apresentaram maior frequência de respostas com o mesmo valor. Sendo assim, essa situação particular justifica-se por se mesma, como uma possibilidade real freqüente de respostas.

Diversos estudos sobre qualidade de serviços, estratégias de políticas de saúde e satisfação de usuários abalizam para uma grande ocorrência de respostas em extremos (AVEZUM et al, 2011; BOING, BOIG, 2007; CONTIERO et al, 2009).

Apesar dessa situação particular, averigou-se também outra situação na qual variáveis se apresentam com homogeneidade de respostas.

Dessa forma, o estudo é indicado como parâmetro ou modelo na utilização das técnicas que reproduzem com maior fidedignidade os dados reais em situações de maior frequência de determinada resposta, ou de frequências semelhantes. Observou-se também que existem itens ou variáveis do banco de dados do estudo que apresentam as características propostas no Cenário 2, justificando o uso da RLM como método de imputação mais fidedigno.

Diante das averiguações realizadas nesse trabalho, entende-se que o melhor método para imputar dependerá do objetivo da análise e característica da variável ou amostra estudada. Abalizou-se que as medidas RMSE e MAPE concordam em decisão, não em valores absolutos, por obterem medidas distintas ou diferentes unidades destas.

A Figura 5.1 demonstra o resumo das decisões desse estudo, resultante das análises dos métodos que apresentaram as melhores performances de acordo com a quantidade de dados faltantes das amostras (Cenário 1):

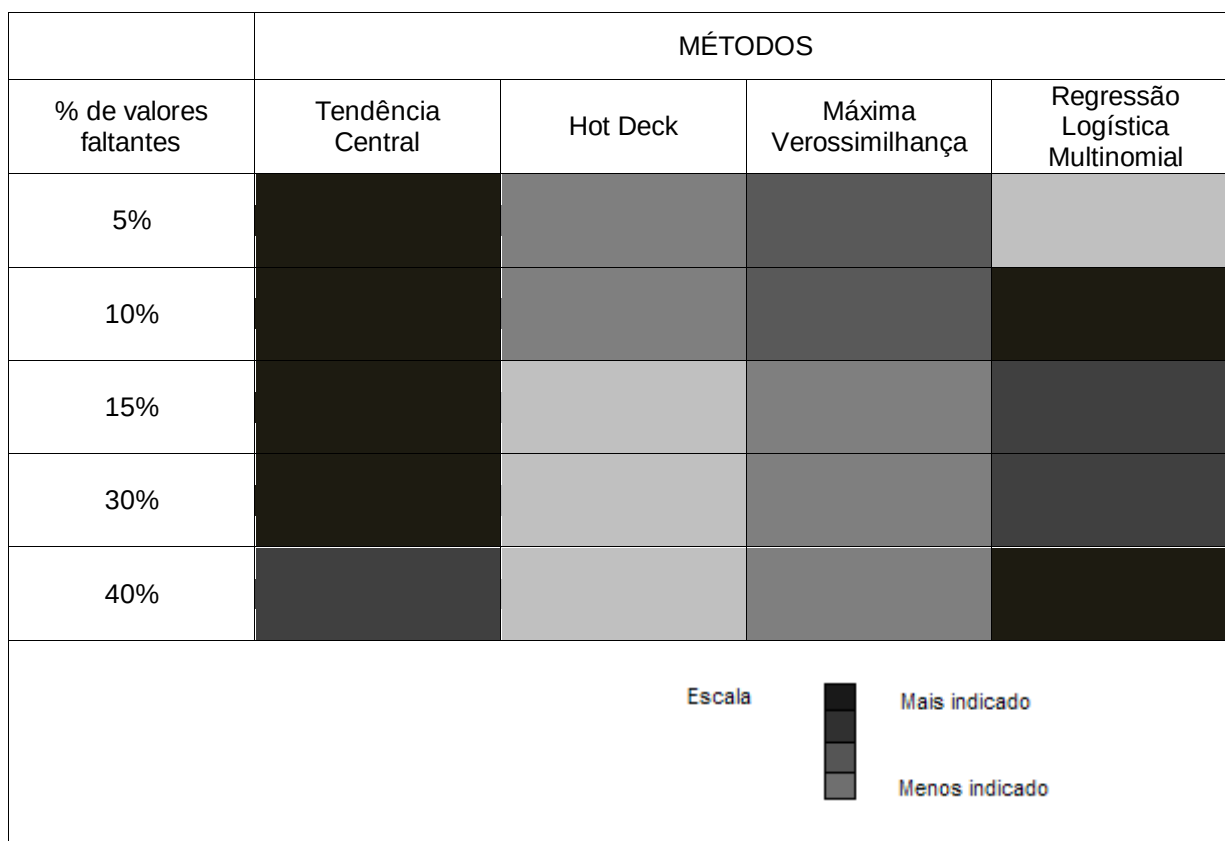


Figura 5.1: Decisão para escolha de métodos de imputação em amostras ou variáveis com presença de frequência de tipo de resposta superior às demais.

A Figura 5.2 apresenta o resumo das decisões, resultado das análises dos métodos que apresentaram as melhores performances de acordo com a quantidade de dados faltantes das amostras para variáveis que apresentem homogeneidade na frequência de respostas (Cenário 2).

	MÉTODOS			
% de valores faltantes	Tendência Central	Hot Deck	Máxima Verossimilhança	Regressão Logística Multinomial
5%				
10%				
15%				
30%				
40%				
Escala  Mais indicado Menos indicado				

Figura 5.2: Decisão para escolha de métodos de imputação em amostras ou variáveis com presença de frequência com tipo de respostas homogêneas.

Dessa forma, percebe-se que, de acordo com as situações ou características dos dados das variáveis incompletas, podem-se tomar decisões para o tratamento adequado dos valores faltantes, evitando manipulação indevida. Assim, é possível solucionar o problema que a falta de dados impõe, como o tamanho reduzido da amostra e a análise de conjunto de dados remanescentes tendenciosos.

Capítulo 6

6. CONSIDERAÇÕES FINAIS

Os resultados encontrados mostram que o HIPERDIA, no município de João Pessoa/PB, possui uma boa completude na maioria das variáveis analisadas — Fatores de Risco e Doenças Concomitantes, já as Sócio-Demográficas apresentaram maior incompletude.

A interpretação desses dados de origem secundária, oriundos dos Sistemas de Informações em Saúde, sempre requer uma análise crítica.

Possíveis limitações com relação à qualidade dos dados do HIPERDIA dizem respeito à metodologia de cadastramento, particularmente em relação à capacitação dos profissionais responsáveis pelo preenchimento das fichas. Alerta-se ainda para o armazenamento da segunda via do cadastro na UBSF, pois, em muitos casos, não há local específico para tal, ficando as fichas dispersas, o que pode oportunizar perdas de fichas, dificultando a avaliação da qualidade dos dados.

Uma vez corretamente utilizados, pela riqueza de dados que podem reunir, os sistemas de informação possibilitam a comprovação da necessidade de alocação de recursos financeiros para a execução de ações para o cuidado adequado da população acometida. Além disso, concorrem para a implementação de medidas preventivas e de promoção da saúde da população assistida, mas, como discutido anteriormente, depende do processo de trabalho das equipes e da capacidade de formulação de perguntas de quem tem acesso às bases de dados (CUNHA, 2002).

Os Sistemas de Informação são ferramentas úteis para os profissionais e gestores em saúde, nos diferentes níveis de atenção, no enfrentamento de problemas de saúde encontrados pela população, na sua totalidade, possibilitando uma visão ampliada dos agravos, permitindo que ações sejam planejadas de acordo com características do grupo. A partir disso, necessita-se desenvolver mais estudos que tenham como objetivo avaliar as informações contidas nas fichas de cadastro do HIPERDIA, assim como de outros sistemas, permitindo criar estratégias para melhor utilizá-los. Mesmo com tais limitações, o Ministério da Saúde poderá utilizar tais métodos na qualificação dos sistemas de informações de saúde e seus respectivos bancos de dados.

Para o problema de falta de dados este trabalho também mostrou a importância de se considerar métodos de imputação para dados faltantes, em

especial a Imputação Única. No entanto, deve-se ter cuidado na generalização dos resultados obtidos com este trabalho, já que eles foram obtidos para situações particulares, usando o banco de dados como uma simples ferramenta para o exercício de se divulgar a técnica da imputação, em particular, os métodos Tendência Central e Regressão Logística Multinomial.

Os itens relacionados ao tamanho amostral, tipo de variável e estrutura de relação entre as variáveis envolvidas devem ser sempre levados em consideração. Mais trabalhos são necessários para que seja avaliado o desempenho dos métodos de imputação, sejam eles para conjuntos de dados com maior proporção de preditores contínuos ou categóricos, com possível relação não-linear, seja para conjuntos de dados com desfechos binários ou não-binários. Estudos sobre análises de sensibilidade para as simulações e estudo de risco para incompletude também são sugeridos para trabalhos futuros.

Tal estudo apresenta uma carta ou guia de decisão para imputação em situações particulares, sugerindo os métodos mais eficazes que retratem com mais fidedignidade a resposta perdida. Existem outras situações que fogem aos cenários estudados, mais que ficam como sugestão para estudos futuros, assim como o desenvolvimento de outros métodos e metodologias para a imputação.

Referências

7. REFERÊNCIAS

ALMEIDA, C.; MACINKO, J. **Validação de uma metodologia de avaliação rápida das características organizacionais e do desempenho dos serviços de atenção básica do Sistema Único de Saúde (SUS) em nível local**. Brasília: Organização Pan-Americana da Saúde, 2006.

AMADEO, C. **Hipertensão Arterial: prognóstico e epidemiologia**. In: ZILMERMAN, A.; CÉSAR, L. A. M. **Manual de Cardiologia**. São Paulo: Socesp, Ateneu, 2000.

ANDRADE, J.P. et al. **Aspectos Epidemiológicos da Aderência ao Tratamento da Hipertensão Arterial Sistêmica**. Arquivo Brasileiro de Cardiologia, vol. 79 (4): 375-379, 2002.

ARAÚJO, A. P. S. et al. **Prevalência dos fatores de risco em pacientes com acidente vascular encefálico atendidos no setor de neurologia da clínica de fisioterapia da UNIPAR, campus sede**. Arquivos de Ciências da Saúde Unipar, Umuarama, vol. 12(1): 35-42, jan./abr, 2008.

AROUCA, S. **Reforma Sanitária**. 1998. Disponível em <http://bvarouca.cict.fiocruz.br/sanitarista05.html>. Acessado em 20 de janeiro de 2011.

ATLAS, S.J., et al. **Patient-physician connectedness and quality of primary care**. Ann Intern Med, vol.150:325-35, 2009.

AVEZUM, A. et al. **Fatores de risco associados com infarto agudo do miocárdio na região metropolitana de São Paulo: uma região desenvolvida em um país em desenvolvimento**, 2005. Disponível em: <<http://www.scielo.br>>. Acesso em: 16 dez 2011.

BARROS, AJD. **São grandes os desafios para o Sistema Nacional de Informações em Saúde**. Ciências e Saúde Coletiva. Vol.11(4):870-86, 2006.

BERTALANFFY, L. V. **Teoria Geral dos Sistemas**. Rio de Janeiro, Vozes/MEC, 1975.

BOING, A. C.; BOING, A. F. **Hipertensão Arterial Sistêmica: o que nos dizem os sistemas brasileiros de cadastramentos e informações em saúde**. Revista Brasileira de Hipertensão, vol 14(2): 84-88, 2007.

BRANDÃO, A. et al. **Epidemiologia da Hipertensão Arterial**. Revista da Sociedade de Cardiologia do Estado de São Paulo, São Paulo, vol.13(1): 7-16, jan.fev. 2003.

BRASIL, Ministério da Saúde. Secretaria de Políticas Públicas. **Guia prático do Programa de Saúde da Família**. Brasília, 2001.

BRASIL. Ministério da Saúde. **Plano de Reorganização da Atenção à Hipertensão Arterial e ao Diabetes mellitus: programa de educação permanente em**

Hipertensão Arterial e Diabetes mellitus para os municípios com população acima de 100 mil habitantes. Brasília, 2002.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **SIAB – Manual do Sistema de Informação da Atenção Básica.** Brasília, DF, 2003.

BRASIL, Ministério da Saúde. **Avaliação do Plano de Reorganização da atenção à Hipertensão Arterial Sistêmica.** Organização Pan-Americana da Saúde. Brasília, 2004.

BRASIL. Ministério da Saúde. **Sistema de Informações sobre Mortalidade.** 2007. Disponível em: <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt.br>>. Acessado em 08/12/2010.

CARVALHO, A. O. **Sistemas de Informação em Saúde para Municípios.** Série Saúde & Cidadania. São Paulo: Faculdade de Saúde Pública da Universidade de São Paulo, 1998.

CARVALHO, A. O.; EDUARDO, M. B. P. **Sistemas de Informação em Saúde para Municípios.** São Paulo: Série Saúde e Cidadania. Faculdade de Saúde Pública da Universidade de São Paulo, 1998.

CHAZAN, A. C.; PEREZ, E. A. **Avaliação da implementação do Sistema Informatizado de Cadastramento e acompanhamento de Hipertensos e Diabéticos (HIPERDIA) nos municípios do Estado do Rio de Janeiro.** Rev APS. 11(1):10-16, 2008.

CHRESTANI, M. M. A. D. **Hipertensão Arterial Sistêmica auto-referida: validação diagnóstica em estudo de base populacional.** Pelotas/RS, 2008. Dissertação (Mestrado em Epidemiologia) - Universidade Federal de Pelotas. 2008.

COLMANETTI, M. S. **Modelagem de sistemas de manufatura orientada pelo custeio das atividades e procesos.** Dissertação (Mestrado). Escola de Engenharia de São Carlos; São Carlos. 2001.

CONASS. **Ciência e Tecnologia em Saúde/ Conselho Nacional dos Secretariados de Saúde.** Brasília: CONASS, 2007.

CONTIERO, A. P. *et al.* **Idoso com hipertensão arterial: dificuldades de acompanhamento na Estratégia Saúde da Família.** Rev. Gaúcha Enferm. 30(1):62-70, 2009.

CORDEIRO, T. **O Instituto da Medicina Social e a Luta pela Reforma Sanitária: contribuição à história do SUS.** Physis, vol 2(14), 2004.

CUNHA, R. E. **Cartão Nacional de Saúde: os desafios da concepção e implantação de um sistema nacional de captura de informações de atendimento em saúde.** Ciênc Saúde Colet., vol 7(4):869-78, 2002.

DALLACOSTA, F. M.; DALLACOSTA, H.; NUNES, A. D. **Perfil de hipertensos cadastrados no Programa do Hiperdia de uma unidade básica de saúde.** Unoesc & Ciência – ACBS, vol 1(1):45-52, 2010.

DANIEL, W. W. **Biostatistics.** 9 ed. IE-Wiley. EUA: A Foudation for Analisys in the Health Sciences. 2009.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. **Maximum likelihood from incomplete data via the EM algorithm (with discussion).** J. Roy. Statist. Soc. B 39, 1–38, 1977.

DONDERS, A.R. *et al.* **Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example.** J Clin Epidemiol; vol. 59:1102-9, 2006.

ENGELS, J. M.; DIEHR, P. **Imputation of missing longitudinal data: a comparison of methods.** Journal of Clinical Epidemiology, vol.56(10): 968-76, 2003.

ESCOREL, S. **Reviravolta da Saúde: origem e articulação do movimento sanitário.** Rio de Janeiro: FioCruz,1998.

FACCHINI, L.A. *et al.* **Sistema de Informação em Saúde do Trabalhador: desafios e perspectivas para o SUS.** Ciências e Saúde Coletiva, vol. 10(4):857-67, 2005.

FRASER, G. **Guided multiple imputation of missing data: using a subsample to strengthen the missing-at-random assumption.** Epidemiolgy, vol. 18(2):246-52, 2007.

FRIAS PG, PEREIRA PMH, ANDRADE CL, SZWARCOWALD CL. **Sistema de informações sobre mortalidade: estudo de caso em municípios com precariedade dos dados.** Cadernos de Saúde Pública, vol. 24(10):2257-66, 2008.

OLIVEIRA, C. A.; PALHA, P. F. **Sistema de Informações Hiperdia, 2002-2004, adequação das informações.** Cogitare Enfermagem, vol.13(3): 395-402, 2008.

FRIAS, P.G. **Sistema de informações sobre mortalidade: estudo de caso em municípios com precariedade dos dados.** Cadernos de Saúde Pública, vol. 24(10):2257-66, 2008.

FUNASA. **Proposta de monitoramento dos agravos não transmissíveis e seus fatores de risco.** In: FUNASA. II Reunião para implantar o monitoramento de doenças não transmissíveis de seus fatores de risco e de outros agravos à saúde, 2000. Disponível em <<http://www.funasa.go.br>>. Acesso em: 07 de julho de 2011.

GIACOMOZZI, C. M.; LACERDA, M. R. **A prática da assistência domiciliary dos Profissionais da Estratégia de Saúde da Família.** Florianópolis: Revista Texto e Contexto de Enfermagem, vol. 15(4), 2006.

GOMES, T. J. O.; SILVA, M. V. R.; SANTOS, A.A. **Controle da pressão arterial em pacientes atendidos pelo programa Hiperdia em uma unidade de Saúde da Família.** Revista Brasileira de Hipertensão, vol. 17(3):132-139, 2010.

GREENLAND, S.; ROTHMAN, K. J. **Modern epidemiology.** 2 ed. Philadelphia: Lippincott-Raven, 1998.

HAIR, J. R. *et al.* **Análise Multivariada de dados.** 6 ed. Porto Alegre: Bookman, 2009.

HAREL, O. Z. **Multiple imputation: review of theory, implementation and software.** Statistics in Medical, vol. 1(16): 3057-77, 2007.

HARRELL, F. E. **Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.** New York: Springer-Verlag, 2001.

IBGE, (Fundação Instituto Brasileiro de Geografia e Estatística), 2008. **Censo Demográfico: Brasil, 2008.** Rio de Janeiro:IBGE.

JUNGER, W. L. **Análise . imputação de dados e interfaces computacionais em estudos de series temporais epidemiológicos.** Rio de Janeiro, RJ, 2008. Tese (Doutorado em Medicina Social) – Universidade do Estado do Rio de Janeiro. 2008

KENWARD, M. G. **Multiple imputation: current perspectives.** Stat Methods Med. Res, vol 16: 199-218, 2007.

KALTON, G. **Compensating for missing survey data.** Ann Arbor, MI: Institute for Social Research, 1993.

LESSA, I. *et al.* **Simultaneidade de fatores de risco cardiovascular modificáveis na população adulta de Salvador (BA), Brasil.** Ver. Panamerica de La Salud Publica, vol. 16(2) 13-33, ago, 2004.

LIMA, C.A.R.; CARVALHO, M.S.; SCHRAMM, J.M.A. **Financiamento público em saúde e confiabilidade dos bancos de dados nacionais. Um estudo dos anos de 2001 e 2002.** Cadernos de Saúde Pública, vol. 22(9):1855-1864. Rio de Janeiro, 2006.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data.** 2nd ed. New York: Wiley; 1986.

LOPES, et al. **Associação entre raça e incidência de doença renal secundária a glomerulonefrite: influencia do tipo histológico e da presença de hipertensão arterial.** Revista Associação Médica Brasileira, São Paulo, v,47, n.1, p.78-84, jan. mar. 2002.

LUNA, R. L. **Aspéctos históricos da hipertensão no Brasil.** HiperAtivo, vol. 6(1), jan, 1999.

MANFROI, A.; OLIVEIRA, F.A. **Dificuldades de adesão ao tratamento na hipertensão arterial sistêmica: considerações a partir de um estudo qualitativo em uma Unidade de Atenção Primária à Saúde.** Revista Brasileira de Medicina da Família e Comunidade, vol. 2(7):165-76, 2006.

MARUITI, A.M.P.; BAULI, J.D.; SCOCHI, M.J. **A importância do Sis-Hiperdia como ferramenta de busca ativa para a prevenção de doenças cardiovasculares.** In: Simpósio Internacional de Estratégias de Cuidados a Indivíduos, Grupos e Famílias e as Doenças Crônicas, 12º Encontro do Laboratório de Ensino Interdisciplinar de Famílias e Saúde; 2008 set 1-4; Pelotas, Brasil. Maringá: Universidade Estadual de Maringá, 2008.

MEDEIROS KRJ. **O Sistema de Informação em Saúde como instrumento da política de recursos humanos: um mecanismo importante na detecção das necessidades da força de trabalho para o SUS.** Ciências e Saúde Coletiva, vol. 10(2):433-40, 2005.

MELO, J. M. P. *et al.* **Avaliação do sistema de informação sobre nascidos vivos e o uso de seus dados em epidemiologia e estatística de saúde.** Revista Saúde Pública, vol. (37)6, 2003.

MIETTINEN, O. S. **Theoretical epidemiology.** Principle of occurrence research in medicine. New York: Wiley, 1985.

MONTEIRO, et al. **Monitoramento de fatores de risco para doenças crônicas por entrevistas telefônicas.** Revista Saúde Pública, São Paulo, v.39, n.1, fev. 2005.

MOONS, K.G.; STIJNEN, T.; HARRELL, J.R.F.E. **Using the outcome for imputation of missing predictor values was preferred.** Journal Clinical Epidemiology, vol. 59:1092-101, 2006.

MORAES, I. H. S. **Informações em saúde: da prática fragmentada ao exercício da cidadania.** São Paulo: Hucitec, 2004.

MORAIS, T. C.; BONFIN, M. **Assistência domiciliar em saúde: subsídios para um projeto de atenção básica brasileira.** Revista Ciência e Saúde Coletiva, vol. 10(sup): 231-242, 2005.

NEVES, M. A. T. **Sistemas de informação em saúde: aspectos históricos e médico-sanitários da produção técnico-científica brasileira.** Dissertação de mestrado – Departamento de Medicina Preventiva, FMUSP, 2006.

NEYMAN, J.; PEARSON, E. S. **On the problem of most efficient tests of statistical hypotheses.** Philosophical Transactions of the Royal Society of London, series A, vol. 231: 289-337, 1973.

NUNES, L. N. **Métodos de imputação de dados aplicados na área da saúde.** Tese de doutorado em medicina: Epidemiologia. Faculdade de medicina. Universidade federal do Rio Grande do Sul, Porto Alegre. 2007.

NUNES, L. N. et al. **Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos.** Caderno de Saúde Pública, Rio de Janeiro, vol. 25(2): 268-278, fev, 2009.

OLIVEIRA, C.A. **O Sistema de informações HIPERDIA em Guarapuava/PR 2002-2004, implantação e qualidade das informações.** 2005. 100F. Dissertação (Mestrado em Enfermagem em Saúde Pública) – Escola de Enfermagem de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2005.

OMS. **Cuidados Inovadores para condições crônicas: componentes estruturais de ação – Relatório Mundial.** Brasília: 2003.

PACHECO, R. F., CARDOSO, T. M. **Utilização de simulação na análise de sistemas de saúde: um estudo de caso no ambulatório da santa casa de misericórdia de Goiânia.** XII SIMPEP - Bauru, SP, 2005.

PAES, N. A. **Avaliação da efetividade do controle da hipertensão arterial sistêmica e associação com os fatores de risco comparando a atenção do Programa de Saúde da Família e de Unidades Básicas de Saúde de municípios do nordeste do Brasil.** Projeto CNPq. Edital: MCT/CNPq/MS – SCTIE – DECIT/MS No. 37/2008. Tema: G. DOENÇAS DO APARELHO CIRCULATÓRIO , 2008.

PAES, N. A. **Desempenho do Programa Saúde da Família comparando com os das Unidades Básicas de Saúde da Família no controle da Hipertensão Arterial Sistêmica e fatores associados em municípios do Estado da Paraíba: um estudo de coorte.** Projeto CNPq Edital MCT/CNPq/CT-Saúde nº 67/2009.

RUBIN, D.B. **Multiple imputation for nonresponse in surveys.** New York: Wiley, 1987.

RUBIN, D.B. **Inference and missing data.** Biometrika. 63(3): 581-92, 1976.

RUBIN, D.B. **Multiple imputation after 18+ years.** J Am Stat Assoc; 91:473-89, 1996.

SCHAFER, J. L. **Multiple Imputation: a primer.** Statistical Methods in Medical Research, vol. 8(1): 3-15, 1999.

SCHAFER, J.L., GRAHAM, J.W. **Missing data: our view of the state of the art.** Psychol Methods; 7:147-77, 2002.

SENTAS, P.; ANGELIS, L. **Categorical missing data imputation for software cost estimation by multinomial logistic regression.** The Journal of Systems and Software, vol. 79(1): 404-414, 2006.

SILVA, C. S. **Análise da dimensão adesão/vínculo dos hipertensos com níveis pressóricos não controlados nas unidades de Saúde da Família do município de João Pessoa.** João Pessoa, PB, 2011. Dissertação (mestrado em Modelos de Decisão em Saúde) – Universidade Federal da Paraíba, 2011.

SILVA AS, LAPREGA MR. **Avaliação crítica do sistema de informação da atenção básica (SIAB) e de sua implantação na região de Ribeirão Preto, São Paulo, Brasil.** Cadernos de Saúde Pública, vol. 21(6):1821-8, Nov-Dez; 2005.

STARFIELD, B. **Atenção primária: equilíbrio entre necessidades de saúde, serviços e tecnologia.** Brasília: UNESCO, Ministério da Saúde, 2002.

TEIXEIRA, F, S. *et al* . **Sistemas de informação: ferramenta no Planejamento de ações em saúde na CASSIRJ.** Anais do VII Congresso Brasileiro de MFC, AMMFC: 2005.

THAINES, G.H.L.S, *et al* . **Produção, fluxo e análise de dados do sistema de informação em saúde: um caso exemplar.** Revista Texto Contexto Enfermagem, vol. 18(3): 466-74, jul-set, 2009.

TIGRE, C. *et al* . **La Práctica Epidemiológica em Los Sistemas de Servivios de Salud.** Educ. Médica Salud, vol. 24(3): 306320, 1990.

TRAVASSOS, C. *et al* . **Utilization of health care services in Brazil: gender, family characteristics, and social status.** Revista Panamericana de la Salud Publica / Pan Am J Pub Health, vol.11(5-6):365-73, 2002.

VIDRO, A.C. **Sistemas de informação em saúde: situação atual em municípios de pequeno porte.** Porto Alegre: Universidade Federal do Rio Grande do Sul; 2004.

WHITE, I. A.; ROYSTON, P. **Editorial: Multiple imputation in practice.** Stat Methods Medical Res, vol. 16; 195-7, 2007.

ZILMER, J. G. V. **Avaliação da qualidade de informações contidas nas fichas do Hiperdia na Unidade Básica de Saúde Simões Lopes, em Pelotas, Rio Grande do Sul.** Monografia. Pelotas: Universidade Federal de Pelotas; 2009.

ZILLMER, J. G. V.; SCHWARTZ, E.; MUNIZ, R. M. **Avaliação da completude das informações do hiperdia em uma Unidade Básica do Sul do Brasil.** Rev. Gaúcha Enferm. [online], vol.31, n.2, pp. 240-246. ISSN 1983-1447,2010.

Anexos

ANEXO 1



MS – HIPERDIA
PLANO DE REORGANIZAÇÃO DA ATENÇÃO
À HIPERTENSÃO ARTERIAL E AO DIABETES MELLITUS

1.ª Via: Enviar para digitação
CADASTRO DO HIPERTENSO
E/OU DIABÉTICO

Nome da Unidade de Saúde (*)		Cód. SIA/SUS (*)		Número do Prontuário																																																									
IDENTIFICAÇÃO DO USUÁRIO (*)																																																													
Nome (com letra de forma e sem abreviaturas)			Data Nascimento / /	Sexo ☐ M ☐ F																																																									
Nome da Mãe (com letra de forma e sem abreviaturas)			Nome do Pai																																																										
Raça/Cor (TV)	Escolaridade (TV)	Nacionalidade ☐ Brasileira ☐ Estrangeira	País de Origem		Data Naturalização / /																																																								
Nº Portaria	UF Munic. Nasc.	Nome Munic. Nascimento	Sit. familiar/Conjugal (TV)	Nº Cartão SUS																																																									
DOCUMENTOS GERAIS																																																													
Título de Eleitor	Número		Zona	Série																																																									
CTPS	Número		Série	UF	Data de Emissão / /																																																								
CPF	Número		PIS/PASEP	Número																																																									
DOCUMENTOS OBRIGATÓRIOS (**)																																																													
Identidade	Número	Complemento	Órgão (TV)	UF	Data de Emissão / /																																																								
Certidão (TV)	Tipo	Nome do Cartório			Livro																																																								
	Folha	Termo			Data de Emissão / /																																																								
ENDEREÇO (*)																																																													
Tipo Logradouro	Nome do Logradouro			Número	Complemento																																																								
Bairro	CEP	DDD	Telefone																																																										
DADOS CLÍNICOS DO PACIENTE																																																													
Pressão Arterial Sistólica (*)	Pressão Arterial Diastólica (*)	Cintura (cm)	Peso (kg) (*)																																																										
Altura (cm) (*)	Glicemia Capilar (mg/d)	☐ Em jejum	☐ Pós prandial																																																										
Fatores de risco e Doenças concomitantes		Não	Sim	Presença de Complicações																																																									
Antecedentes Familiares - cardiovasculares				Infarto Agudo Miocárdio																																																									
Diabetes Tipo 1				Outras coronariopatias																																																									
Diabetes Tipo 2				A/C																																																									
Tabagismo				Pé diabético																																																									
Sedentarismo				Amputação por diabetes																																																									
Sobrepeso/Obesidade				Doença Renal																																																									
Hipertensão Arterial																																																													
TRATAMENTO																																																													
Não Medicamentoso: ☐																																																													
<table border="1"> <tr> <th rowspan="2">Tipo</th> <th colspan="6">Comprimidos/dia</th> <th rowspan="2">Unidades/dia</th> </tr> <tr> <th>1/2</th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> <th>6</th> </tr> <tr> <td>Hidroclorotiazida 25mg</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td rowspan="5"> Insulina </td> </tr> <tr> <td>Propranolol 40mg</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Captopril 25mg</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Glibenclamida 5mg</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Metformina 850 mg</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>						Tipo	Comprimidos/dia						Unidades/dia	1/2	1	2	3	4	5	6	Hidroclorotiazida 25mg								Insulina 	Propranolol 40mg								Captopril 25mg								Glibenclamida 5mg								Metformina 850 mg							
Tipo	Comprimidos/dia						Unidades/dia																																																						
	1/2	1	2	3	4	5		6																																																					
Hidroclorotiazida 25mg								Insulina 																																																					
Propranolol 40mg																																																													
Captopril 25mg																																																													
Glibenclamida 5mg																																																													
Metformina 850 mg																																																													
Outros ☐ SIM ☐ NÃO																																																													
Data da Consulta (*)		Assinatura do Responsável pelo atendimento (*)																																																											

Legenda: (*) Campos obrigatórios, com exceção: nome pai; data naturalização e nº portaria, se nacionalidade brasileira (nascido no Brasil); complemento, DDD e telefone. (**) Pelo menos um dos documentos é obrigatório. TV = Tabela no verso do formulário.

ANEXO 2

Nº _____

Projeto: Avaliação da efetividade no controle da hipertensão arterial sistêmica e associação com fatores de risco comparando a atenção do programa de saúde da família e de unidades básicas de saúde de municípios do Nordeste do Brasil

QUESTIONÁRIO

A. IDENTIFICAÇÃO DO QUESTIONÁRIO

Município: () João Pessoa () Campina Grande () Natal Pesquisador: _____
 Data da coleta de dados: ____/____/____ Supervisor de campo: ____/____/____ Ass: _____
 Data da digitação 1: ____/____/____ Ass: _____ Data da digitação 2: ____/____/____ Ass: _____

B. IDENTIFICAÇÃO DO USUÁRIO

B.1 Nome do usuário: _____
 B.2 Endereço: _____
 B.3 Telefone: _____
 B.4 Contato: _____

C. INFORMAÇÕES GERAIS (FONTES SECUNDÁRIAS)

C.1	Cód. SIA/SUS :		
C.2	Nº Prontuário:	Data nascimento: ____/____/____	
C.3	Nome da Unidade de Saúde (US)		
	Distrito (US):		
C.5	Endereço da US:		
C.6	Tipo de Unidade	UBS	1 ()
		UBS/PACS	2 ()
		UBSF/PACS	3 ()
C.A	Pressão arterial e dados antropométricos no momento do cadastro Hipertensão		
C.7	Data do cadastro 2006/2007: ____/____/____		
C.8	Pressão Arterial Sistólica	mmHg	
C.9	Pressão Arterial Diastólica	mmHg	
C.10	Peso	kg	
C.11	Altura	cm	
C.12	Cintura	cm	
C.B	Dados antropométricos (1ª medida de 2008)		
C.13	Peso	kg	
C.14	Altura	cm	
C.15	Cintura	cm	
C.C	Pressão arterial e antropometria (momento da entrevista)		
C.16	Pressão Arterial Sistólica	mmHg	
C.17	Pressão Arterial Diastólica	mmHg	
C.18	Peso 1 _____	Peso 2 _____	kg
C.19	Altura 1 _____	Altura 2 _____	cm
C.20	Cintura 1 _____	Cintura 2 _____	cm
C.21	Quadril 1 _____	Quadril 2 _____	cm
C.D	NÃO ACOMPANHADOS EM 2008 ()	PA Sistólica	PA Diastólica
	Nº de Medicamentos da HAS		
	Data 1ª consulta 2008: ____/____/____		
	Data 2ª consulta 2008: ____/____/____		
	Data 3ª consulta 2008: ____/____/____		
	Data 4ª consulta 2008: ____/____/____		
	Data 5ª consulta 2008: ____/____/____		
	Data 6ª consulta 2008: ____/____/____		
	Data 7ª consulta 2008: ____/____/____		
	Data 8ª consulta 2008: ____/____/____		

Data 9ª consulta 2008: ____/____/____				
Data 10ª consult 2008: ____/____/____				
Data 11ª consult 2008: ____/____/____				
Data 12ª consult 2008: ____/____/____				
D. INFORMAÇÕES SÓCIO-DEMOGRÁFICAS				
			hipertensão	usuário
D.1	Sexo	Feminino	0a ()	0b ()
		Masculino	1a ()	1b ()
D.2	Idade		____anos	____anos
D.3	Situação familiar/ conjugal	Convive c/ companheiro(a) e filho (a)	1a ()	1b ()
		Convive c/ companheiro(a) c/ laços conjugais e s/ filho (a)	2a ()	2b ()
		Convive c/ familiares s/ companheiro(a)	3a ()	3b ()
		Convive c/ companheiro(a) c/ filho (a) e outros	4a ()	4b ()
		Convive c/ outras pessoas s/ laços consanguíneos e/ou laços conjugais	5a ()	5b ()
		Vive só	6a ()	6b ()
D.4	Escolaridade	Não sabe ler/escrever	1a ()	1b ()
		Alfabetizado	2a ()	2b ()
		Ensino fundamental incompleto (1º grau incompleto)	3a ()	3b ()
		Ensino fundamental completo (1º grau completo)	4a ()	4b ()
		Ensino médio completo (2º grau incompleto)	5a ()	5b ()
		Ensino médio completo (2º grau completo)	6a ()	6b ()
		Ens. superior (incompleto)	7a ()	7b ()
		Ens. superior (completo)	8a ()	8b ()
		Especialização/Residência	8a ()	8b ()
		Mestrado	9a ()	9b ()
		Doutorado	10a ()	10b ()
D.5	Raça/Cor	Branca	1a ()	1b ()
		Preta	2a ()	2b ()
		Amarela	3a ()	3b ()

		Parda	4a ()	4b ()			
		Indígena	5a ()	5b ()			
Para as questões D.6 e D.7 responda: S – sim; N – não;							
D.6	O(a) Sr(a) trabalhava antes de saber que tinha HAS?			1.() S 2.() N 9.() NS/NR			
D.7	O(a) Sr(a) teve que mudar ou parar de trabalhar por causa da HAS?			1.() S 2.() N 9.() NS/NR			
D.8	Ocupação		Antes HAS	Depois HAS			
		desempregado	0a ()	0b ()			
		Do lar: contribuinte	1a ()	1b ()			
		Do lar: não-contribuinte	2a ()	2b ()			
		Empregado contribuinte	3a ()	3b ()			
		Empregado não contribuinte	4a ()	4b ()			
		Autônomo Contribuinte	5a ()	5b ()			
		Autônomo não contribuinte	6a ()	6b ()			
		Aposentado	7a ()	7b ()			
Outro	8a ()	8b ()					
D.9 COMPOSIÇÃO DA FAMÍLIA							
NOME		GRAU PARENTES CO	IDADE	RENDA NO ÚLTIMO MÊS (VALOR R\$)			
SUB TOTAL							
D.10 OUTRAS RENDAS							
1	Bolsa Família						
2	Bolsa Escola						
3	Aluguel						
4	Outras: _____						
TOTAL RENDA FAMILIAR							
E. FATORES DE RISCO E DOENÇAS CONCOMITANTES							
		hiperdia			usuário		
E.1	Antecedente familiar Cardiovascular	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.2	Diabetes T1	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.3	Diabetes T2	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.4	Tabagismo	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.5	Sedentarismo	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.6	Sobrepeso/ Obesidade	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.7	Infarto Agudo Miocárdio	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.8	Outra Coronariopatia	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.9	AVC	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.10	Pé diabético	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.11	Amputação por diabetes	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.12	Doença Renal	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E. A	Informações do prontuário				Informações do usuário		
E.13	Etilismo	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.14	Menopausa	() S1	() N1	() N/S1	() S2	() N2	() N/S2

E.15	Uso de anti-contraceptivo hormonal	() S1	() N1	() N/S1	() S2	() N2	() N/S2
E.16	Nº de medicação da HAS:						
E.17	Nº de vezes que retornou em 2008 (no prontuário)						
F. SAÚDE DO CASO CONFIRMADO DE HAS							
Para as questões F.1, F.2 e F.3 responda: 1-Sempre; 2– Quase Sempre; 3–Às vezes; 4–Quase Nunca; 5 – Nunca; 0-Não se aplica; 99-NS/NR							
F.1	Alguma vez o(a) Sr.(a) deixa de realizar qualquer atividade habitual (trabalhar, estudar, lazer), por conta da HAS?						
F.2	Com que frequência o(a) Sr(a). deixa de tomar a medicação para HAS?						
F.3	Alguma vez o(a) Sr.(a) achou que o medicamento lhe trouxe algum efeito desagradável?						
Para as questões F.4 e F.6 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR							
F.4	O(a) Sr(a) pratica atividade física, com que frequência?						
F.5	Alguma vez o(a) Sr(a) foi orientado(a) a fazer atividade física, por algum profissional de saúde da sua unidade?						
F.6	Com que frequência, o(a) Sr(a) faz dieta orientada por algum profissional de saúde da sua unidade?						
G. ACESSO AO DIAGNÓSTICO							
Para a questão G.1 responda: 1-UBSF/PACS; 2-UBS; 3-UBS/PACS; 4-Hospital Público; 5-Hospital Privado; 6-Consultório Particular; 8-Outros(____); 0-Não se aplica; 99-NS/NR							
G.1	Qual foi o serviço de saúde que descobriu (diagnosticou) que o(a) Sr(a) estava doente de HAS?						
Para a questão G.2 responda: 1 – 5 ou mais vezes; 2 – 4 vezes; 3 – 3 vezes; 4 – 2 vezes; 5 – 1 vez; 0 – Não se aplica; 99-NS/NR							
G.2	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura), quantas vezes precisou procurar a unidade de saúde para descobrir que era hipertenso?						
Para a questão G.3 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR							
G.3	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura), procurou o serviço de saúde mais próximo da sua casa?						
Para as questões G.4 a G.8 responda: 1-Sempre; 2– Quase Sempre; 3–Às vezes; 4–Quase Nunca; 5 – Nunca; 0-Não se aplica; 99-NS/NR							
G.4	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura) e procurou o serviço de saúde para consultar, demorou mais de 60 minutos para ser atendido?						
G.5	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura) teve dificuldade para se deslocar até o serviço de saúde?						
G.6	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura) precisou utilizar algum tipo de transporte motorizado para ir até o serviço de saúde?						
G.7	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura) gastou dinheiro com transporte para ir até o serviço de saúde?						
G.8	Quando o(a) Sr.(a) começou a ter os sintomas da HAS (dor de cabeça; tontura) perdeu o turno de trabalho ou compromisso para consultar no serviço de saúde?						
H. ACESSO AO TRATAMENTO							
Para a questão H.1 e H.2 responda: 1-UBSF/PACS; 2-UBS; 3-UBS/PACS; 4-Hospital Público; 5-Hospital Privado; 6-Consultório Particular; 8-Outros(____); 0-Não se aplica; 99-NS/NR							
H.1	Qual o serviço de saúde que o(a) Sr(a) faz as consultas para o tratamento da HAS?						
H.2	Qual o serviço de saúde que o(a) Sr(a) recebe a medicação para o tratamento da HAS?						
Para a questão H.3 a H.5 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR							
H.3	Se o(a) Sr(a) passar mal por causa da medicação ou da HAS, consegue uma consulta no prazo de 24hs na unidade de saúde que faz tratamento?						
H.4	Os profissionais da unidade de saúde que acompanham seu tratamento de HAS costumam visitá-lo em sua moradia?						

H.5	O(a) Sr.(a) faz o tratamento da HAS no serviço de saúde mais próximo da sua casa?		
Para as questões H.6 a H.11 responda: 1-Sempre; 2– Quase Sempre; 3– Às vezes; 4– Quase Nunca; 5 – Nunca; 0-Não se aplica; 99-NS/NR			
H.6	Quando o(a) Sr.(a) vai a unidade de saúde para consultar seu problema de HAS, perde seu turno de trabalho ou compromisso?		
H.7	O(a) Sr.(a) tem dificuldade para se deslocar até a unidade de saúde para ser consultado?		
H.8	Quando o(a) Sr.(a) vai a unidade de saúde para consultar seu problema de HAS precisa utilizar algum tipo de transporte motorizado?		
H.9	Quando o(a) Sr.(a) vai a unidade de saúde para consultar, paga pelo transporte?		
H.10	Desde quando iniciou o seu tratamento para HAS faltou medicamento?		
H.11	Quando o(a) Sr.(a) vai a unidade de saúde para consultar, demora mais de 60 minutos para ser atendido?		
Para as questões H.12 a H.17 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
H.12	O(A) Sr.(a) consegue realizar exames solicitados pelo profissional de saúde que acompanha seu tratamento?		
H.13	O(A) Sr.(a) consegue no período de 10 dias receber os resultados dos exames solicitados pelo profissional de saúde que acompanha seu tratamento?		
H.14	O(A) Sr(a). foi questionado pelos profissionais da unidade se aceitava fazer o tratamento medicamentoso?		
H.15	Em algum momento durante a consulta o Sr(a). é questionado pelos profissionais da unidade se está tomando a medicação?		
H.16	Em algum momento durante a consulta o Sr(a). é questionado pelos profissionais da unidade se esta fazendo a dieta diariamente?		
H.17	Em algum momento durante a consulta o Sr(a). é questionado pelos profissionais da unidade quanto ao horário que esta tomando a medicação?		
I. ADESÃO/VÍNCULO			
Para as questões I.1 a I.10 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
I.1 O(A) Sr(a) recebe apoio por parte:	I.1a	Família?	
	I.1b	Colegas trabalho?	
	I.1c	Amigos?	
	I.1d	Profissionais de saúde da unidade?	
	I.1e	Outros: _____	
I.2	Quando o(a) Sr.(a) vai à unidade de saúde para a consulta de HAS, é atendido pelo mesmo profissional?		
I.3	Se o(a) Sr.(a) tem alguma dúvida sobre o seu tratamento, consegue falar com o mesmo profissional da unidade de saúde que o atende?		
I.4	Quando o(a) Sr.(a) faz alguma pergunta ao profissional da unidade de saúde sente que é compreendido?		
I.5	O profissional da unidade de saúde que o(a) Sr.(a) está fazendo o tratamento responde às suas perguntas de maneira clara?		
I.6	O profissional da unidade de saúde dá tempo suficiente para que o(a) Sr.(a) fale suas dúvidas ou preocupações?		
I.7	Quando o(a) Sr.(a) consulta na unidade de saúde, o profissional conversa sobre outros problemas de saúde?		
I.8	O profissional da unidade de saúde explica sobre a importância de seguir o tratamento indicado para o controle da HAS?		
I.9	O profissional da unidade de saúde pergunta sobre todos os medicamentos que o(a) Sr.(a) está utilizando?		
I.10	Os profissionais da unidade de saúde relacionam-se bem com as pessoas da comunidade?		
Para a questão I.11 responda: 1 - Sempre; 2 – Quase Sempre; 3 – Às vezes; 4 – Quase Nunca; 5 – Nunca; 0 – Não se aplica; 99-NS/NR			
I.11	O(A) Sr.(a) já pensou em mudar da unidade de saúde por causa dos profissionais?		
I.12 Numa escala de 1 a 5, considerando 1 (pior situação) e 5 (melhor situação), qual seria o seu grau de satisfação em relação aos	I.12a	Médico	
	I.12b	Enfermeiro	
	I.12c	Auxiliar Enfermagem	
	I.12d	Agente Comunitário de Saúde/ACS	
	I.12e	Outros Profissionais	

seguintes profissionais que o atende?			
J. ELENCO DE SERVIÇOS			
Para as questões J.1 a J.13 responda seguindo a escala de 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
Com que frequência as ações a seguir são oferecidas pela equipe que acompanha o seu problema da HAS.			
J.1	Informação sobre a HAS e seu controle?		
J.2	Informação sobre os medicamentos e seus efeitos?		
J.3	Educação em saúde (informação sobre outros temas de saúde)?		
J.4	Verificação de PA todas as vezes que vai para consulta?		
J.5	Visitas domiciliares?		
J.6	O(A) Sr.(a) encontra um profissional na unidade de saúde para atendê-lo em todos os dias úteis da semana?		
J.7	Participação em grupos de doentes de HAS na unidade de saúde?		
J.8	Os Horários de funcionamento do serviço de saúde para atendimento são sempre respeitados?		
J.9	O(A) Sr(a) recebe todos os medicamentos necessários para o tratamento da HAS?		
J.10	O serviço oferece/garante exames solicitados pelo profissional da saúde?		
J.11	O(A) Sr(a) recebe resultados exames em até 10 dias?		
J.12	O profissional da unidade de saúde explica os resultados dos exames?		
J.13	É agendada a sua consulta de retorno?		
L. COORDENAÇÃO			
Para as questões L.1 a L.7 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
L.1	Os profissionais da unidade de saúde pegam seu prontuário/ficha do(a) Sr(a) durante a consulta?		
L.2	Durante seu atendimento o profissional da unidade de saúde anota as suas queixas no seu prontuário?		
L.3	O(A) Sr(a) é avisado(a) sobre o agendamento da sua consulta de retorno na unidade de saúde?		
L.4	Quando o(a) Sr(a) tem algum problema de saúde, recebe encaminhamento por escrito para o outro serviço de saúde pelo profissional que acompanha seu tratamento?		
L.5	Quando o(a) Sr(a) tem algum problema de saúde e é encaminhado para outro serviço de saúde tem o atendimento garantido no serviço referendado?		
L.6	O(A) Sr(a) retorna à unidade de saúde com informações escritas sobre os resultados da consulta realizada no outro serviço?		
L.7	Os profissionais da unidade de saúde discutem com o(a) Sr(a) sobre os resultados da consulta realizada no outro serviço?		
M. ENFOQUE NA FAMÍLIA			
Para as questões M.1 a M.3 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
M.1	Os profissionais da unidade de saúde procuram conhecer as pessoas que moram com o(a) Sr.(a)?		
M.2	Os profissionais da unidade de saúde conversam com as pessoas que moram com o(a) Sr.(a) sobre a HAS, estilo de vida, o seu tratamento e outros problemas de saúde?		
M.3	Os profissionais de saúde conversam sobre a importância do envolvimento da sua família no seu tratamento?		
N. ORIENTAÇÃO PARA A COMUNIDADE			
Para as questões N.1 a N.3 responda: 1 – Nunca; 2 – Quase nunca; 3 – Às vezes; 4 – Quase sempre; 5 – Sempre; 0 – Não se aplica; 99-NS/NR			
N.1	Os profissionais da unidade de saúde conversam sobre a importância da sua participação e da sua família em instituições da comunidade (igrejas, associação de bairro, etc.) como apoio para resolver seus problemas de saúde?		
N.2	Com que frequência os serviços de saúde desenvolvem ações sobre HAS com as Igrejas, Associações de Bairro, escolas, etc.?		
N.3	Os profissionais da unidade de saúde conversam sobre a influência da família/amigos/colegas no seu tratamento?		

SUAS SUGESTÕES, COMENTÁRIOS E/OU DÚVIDAS

ANEXO 3



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS DA SAÚDE
COMITÊ DE ÉTICA EM PESQUISA

CERTIDÃO

Certifico que o Comitê de Ética em Pesquisa, do Centro de Ciências da Saúde da Universidade Federal da Paraíba – CEP/CCS aprovou por unanimidade na 3ª Reunião Ordinária, realizada no dia 29-04-09, o projeto de pesquisa do interessado Professor Neir Antunes Paes, intitulada “AVALIAÇÃO DA EFETIVIDADE NO CONTROLE DA HIPERTENSÃO ARTERIAL SISTÊMICA E ASSOCIAÇÃO COM FATORES DE RISCO COMPARANDO A ATENÇÃO DO PROGRAMA DE SAÚDE DA FAMÍLIA E DE UNIDADES BÁSICAS DE SAÚDE DE MUNICÍPIOS DO NORDESTE DO BRASIL”. Protocolo nº. 0101.

Outrossim, informo que a autorização para posterior publicação fica condicionada à apresentação do resumo do estudo proposto à apresentação do Comitê.


Eliane Marques D. de Souza
Coordenadora - CEP-CCS-UFPB

ANEXO 4



Prefeitura Municipal de João Pessoa
Secretaria de Saúde

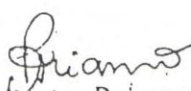
João Pessoa, 04 de junho de 2009.

ENCAMINHAMENTO

Cumprimentando-os cordialmente, encaminhamos o (a) pesquisador (a) NEIR ANTUNES PAES, para realização de coleta de dados da pesquisa intitulada "AVALIAÇÃO DA EFETIVIDADE NO CONTROLE DA HIPERTENSÃO ARTERIAL SISTÊMICA E ASSOCIAÇÃO COM FATORES DE RISCO COMPARANDO A ATENÇÃO DO PROGRAMA DE SAÚDE DA FAMÍLIA E DAS UNIDADES BÁSICAS DE SAÚDE DE MUNICÍPIOS DO NORDESTE DO BRASIL", a ser realizada nas Unidades de Saúde da Família dos Distritos Sanitários I, II, III, IV e V, relacionadas em anexo na cópia do processo.

Sem mais, e visando o bom andamento das pesquisas na Rede SUS de João Pessoa, subscrevo-me,

Atenciosamente,


Vania Priamo
Gerente de Educação em Saúde
Mat.: 51.972-3

Vania Priamo
Gerente de Educação na Saúde