Maria Ligia Damato Capuani

**Mortalidade entre doadores de sangue soropositivos para
doença de Chagas (1996-2000) em São Paulo: um estudo de
relacionamento probabilístico de dados**

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo
para obtenção do título de Doutor em Ciências

Programa de Doenças Infecciosas e Parasitárias
Orientador: Profa. Dra. Ester Cerdeira Sabino

**São Paulo
2016**

Maria Ligia Damato Capuani

**Mortalidade entre doadores de sangue soropositivos para
doença de Chagas (1996-2000) em São Paulo: um estudo de
relacionamento probabilístico de dados**

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo
para obtenção do título de Doutor em Ciências

Programa de Doenças Infecciosas e Parasitárias
Orientador: Profa. Dra. Ester Cerdeira Sabino

**São Paulo
2016**

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Capuani, Maria Ligia Damato

Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo : um estudo de relacionamento probabilístico de dados / Maria Ligia Damato Capuani. -- São Paulo, 2016.

Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Doenças Infecciosas e Parasitárias.
Orientadora: Ester Cerdeira Sabino.

Descritores: 1.Mortalidade 2.Doença de Chagas 3.Registro médico coordenado
4.Banco de dados 5.Doadores de sangue 6.Doenças parasitárias

USP/FM/DBD-262/16

Dedicatória

Dedico este trabalho às minhas filhas Luisa Capuani Masini e Luciana Capuani Masini, que me apoiaram incondicionalmente durante os anos em que me dediquei a esta tese. Apesar de terem sido tempos bastante difíceis para nós três, por vários motivos, espero ter demonstrado com esta tese que com determinação e perseverança conseguimos atingir os objetivos aos quais nos propomos, e, apesar de cometermos erros ao longo de nossas vidas, há sempre tempo para corrigir e recomeçar.

Agradecimentos

Agradeço pelo apoio, incentivo e colaboração em todas as etapas deste trabalho a Dra Ester Cerdeira Sabino e a Dra Ana Luiza Bierrenbach, minhas orientadora e co-orientadora. Sem a colaboração delas esta tese não teria sido realizada.

Também agradeço a minha irmã Maria Lúcia Damato Capuani, pelo apoio durante os quatro anos que passei pelo programa de doutorado, nunca deixando que eu esquecesse meu objetivo final.

Por fim gostaria de agradecer aos meus pais, Mario Bruno Capuani e Dirce da Silva Damato Capuani, por sempre terem sido, ao longo de suas vidas, exemplos de pessoas comprometidas com a educação neste país.

Esta dissertação ou tese está de acordo com as seguintes normas, em vigor no momento desta publicação:

Referências: adaptado de International Committee of Medical Journals Editors (Vancouver).

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. Guia de apresentação de dissertações, teses e monografias. Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria F. Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3a ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

Abreviaturas dos títulos dos periódicos de acordo com List of Journals Indexed in Index Medicus.

Sumário

Lista de abreviaturas

Lista de tabelas

Lista de Figuras

Resumo

Summary / Abstract

1.	Introdução	6
1.1.	Epidemiologia	7
1.2.	Mortalidade	9
1.3.	Relacionamento probabilístico de registros	12
1.4.	Mortalidade em doença de Chagas	15
2.	Objetivo	19
2.1.	Objetivo Estudo Preliminar.....	19
2.2.	Objetivo Estudo Mortalidade	19
3.	Métodos	21
3.1.	Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico.....	21
3.2.	Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo	30
4.	Resultados	46

4.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico.....	46
4.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo	48
5. Discussão.....	62
5.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico	62
5.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo	67
6. Conclusão	73
6.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico.....	73
6.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo	73
7. Referências Bibliográficas.....	74

Apêndices

Lista de abreviaturas

CEP	Comitê de Ética em Pesquisa
CID-10	Classificação Internacional de Doenças - 10ª Revisão
CONEP	Comissão Nacional de Ética em Pesquisa
Datusus	Departamento de Informática do Sistema Único de Saúde
DO	Declaração de óbito
ECG	Eletrocardiograma
ELISA	Ensaio imunoenzimático
FPS	Fundação Pró-Sangue - Hemocentro de São Paulo
HC-FMUSP	Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo
HI	Hemaglutinação indireta
HIV	Vírus da imunodeficiência humana
HTLV I e II	Vírus Linfotrópico da célula Humana I e II
HR	Taxa de risco instantânea ("Hazard ratio")
IFI	Imunofluorescência indireta

OMS	Organização Mundial da Saúde
OPAS	Organização Panamericana da Saúde
PBLOCO	Código fonético do primeiro nome
RL	Record Linkage
RPD	Relacionamento probabilístico de dados
SES	Secretarias Estaduais de Saúde
SIF	Sífilis
SIM	Sistema de Informação sobre Mortalidade
SMS	Secretárias Municipais de Saúde
SVS	Secretaria de Vigilância em Saúde
<i>T.cruzi</i>	<i>Trypanosoma cruzi</i>
UBLOCO	código fonético do último nome
VHB	Virus da Hepatite B
VHC	Virus da Hepatite C
VPP	Valor Preditivo Positivo

Lista de tabelas

Tabela 1 -	Parametros demográficos e epidemiológicos da doença de Chagas na América Latina por país - 2010.....	7
Tabela 2 -	Distribuição de doadores soropositivos e soronegativos para doença de Chagas por semestre.....	33
Tabela 3 -	Co-infecções de doadores soropositivos para doença de Chagas da FPS entre 1996 a 2000.....	34
Tabela 4 -	Distribuição de registros de óbitos por ano.....	35
Tabela 5 -	Sensibilidade, especificidade e valor preditivo positivo de vários escores do processo de relacionamento probabilístico de registros.....	47
Tabela 6 -	Distribuição dos doadores de sangue da FPS selecionados para o processo de relacionamento probabilístico por sexo, faixa etária e ano de doação.....	49
Tabela 7 -	Distribuição de doadores soropositivos e soronegativos por escore.....	50
Tabela 8 -	Distribuição de pares verdadeiros por doadores soropositivos e soronegativos para doença de Chagas.....	51

Tabela 9 - Causas básicas de morte no SIM dos pares verdadeiros por Capítulos.....	52
Tabela 10 - Número de óbitos e Risco Relativo por sorologia, faixa etária e grupo.....	53
Tabela 11 - Causas básicas de morte do Grupo 3 - Doenças Cardíacas para doadores soropositivos para doença de Chagas.....	56
Tabela 12 - Frequencia de causas associadas por Grupos.....	57
Tabela 13 - Número de óbitos por ano de doação.....	57
Tabela14 - Números e taxas de mortalidade, brutas e ajustadas, para diferentes desfechos.....	60

Lista de Figuras

Figura 1 - Descrição da limpeza de dados dos arquivos referência e comparação utilizados no processo de relacionamento probabilístico.....	24
Figura 2 - Classificação dos códigos CID-10 nos grupos 1 a 6.....	37
Figura 3 - Probabilidade de sobrevivência segundo sorologia.....	58
Figura 4 - Probabilidade de sobrevivência segundo sorologia por faixa etária.....	59

Resumo

Capuani MLD. Mortalidade entre doadores de sangue soropositivos para a doença de Chagas (1996 – 2000) em São Paulo: um estudo de relacionamento probabilístico de dados.

INTRODUÇÃO: As taxas de mortalidade para doença de Chagas variam, consideravelmente de 0,2 a 19,2% anualmente dependendo principalmente do estágio da doença. Considera-se que pacientes na fase indeterminada tem taxas baixas, semelhante a população em geral. Este estudo compara a taxa de mortalidade e as causas de morte de doadores de sangue da Fundação Pró-Sangue - Hemocentro de São Paulo (FPS) soropositivos para a doença de Chagas e de doadores soronegativos para todos os testes de triagem realizados no processo de doação. Doadores de sangue constituem uma população, na qual os indivíduos soropositivos são em sua maioria assintomáticos. **MÉTODOS:** Este é um estudo de coorte retrospectiva de doadores de sangue de 1996 a 2000: 2842 doadores soropositivos para doença de Chagas e 5684 doadores soronegativos para todos os testes de triagem. O status vital foi designado realizando-se o relacionamento probabilístico de dados (RPD) utilizando-se o Sistema de Informação sobre Mortalidade do Brasil (SIM). Utilizamos um método RDP validado preliminarmente com sensibilidade de 94% (IC 95%, 90% - 97%), especificidade de 100% (IC 95%, 98% - 100%). Apenas os dados dos anos 2001 a 2009 do SIM foram disponibilizados para análise. Análise univariada foi conduzida considerando desfechos e fatores de riscos potenciais. Os testes Chi-quadrado e Exato de Fisher foram utilizados conforme apropriado. A

diferença de sobrevida dos dois grupos no tempo foi graficamente representada pela Curva de Kaplan-Meyer e testado pelo teste Logrank. Regressões múltiplas de Cox foram utilizadas para derivação de riscos relativos instantâneos (hazard ratios - HR) para os desfechos do estudo, ajustando-se por variáveis de confusão. RESULTADOS: RPD encontrou 159 óbitos entre os 2842 doadores soropositivos (5,6%) e 103 óbitos entre os 5684 doadores soronegativos (1,8%). A sorologia de doadores foi associada com óbitos relacionados a doença de Chagas (HR 2,3 - IC 95%: 1,8 - 3,0). As curvas de sobrevida foram diferentes quando comparando doadores soropositivos e soronegativos, jovens e velhos ($p < 0,001$). Das 159 mortes entre doadores soropositivos, 26 foram devidas a doença de Chagas (CID-10: B57.0 - B57.5), 47 foram devidas a outras doenças cardíacas (CID-10: I10 a I80.2), sem menção a doença de Chagas como uma causa associada, e apenas 1 com menção a doença de Chagas na Parte II da declaração de óbito. CONCLUSÕES: Doadores de sangue soropositivos seguidos retrospectivamente por 9 a 14 anos morrem ao menos 2,3 vezes mais que doadores soronegativos. As causas de morte no SIM podem estar mal classificadas para os doadores de sangue soropositivos assintomáticos.

DESCRITORES: mortalidade; doença de Chagas; registro médico coordenado, banco de dados, doadores de sangue, doenças parasitárias.

Summary / Abstract

Capuani MLD. Mortality among blood donors seropositive for Chaga's Disease (1996 - 2000) in São Paulo: a medical record linkage study.

INTRODUCTION: Mortality rates for Chagas' disease vary considerably between 0.2 and 19.2% annually depending mostly on the stage of the disease. Patients at the indeterminate phase are considered to have low rates, similar to the overall population. This study compares mortality rates and causes of death of blood donors seropositive for Chagas' disease and seronegative for all screening tests at Fundação Pró-Sangue - Hemocentro de São Paulo (FPS). Blood donors constitute a particular population in which seropositive individuals are mostly asymptomatic. **METHODS:** This is a retrospective cohort study of blood donors from 1996 to 2000: 2842 seropositive for Chagas disease and 5684 seronegative. Death status was ascertained by performing probabilistic record linkage (RL) with the national mortality information system (SIM). RL was validated in a preliminary study and sensitivity and specificity of the method were determined to be 94% (CI 95%, 90% - 97%) and 100% (CI 95%, 98% - 100%) respectively. Only SIM records for 2001 to 2009 were available for analysis. Univariate analysis was conducted considering outcomes and potential risks factors. Chi-square test or Fisher exact test were used as appropriate. The different survival of the two groups over time was graphically represented by the Kaplan-Meier curve and tested with the Logrank test. Multiple Cox regressions were used to derive Hazard ratios (HR) for the outcomes studied, adjusting for confounders. **RESULTS:** RL

identified 159 deaths among the 2842 seropositive blood donors (5.6%) and 103 deaths among the 5684 seronegative (1.8%). Serology of blood donors was associated with Chagas disease related death (HR 2.3 - 95% IC: 1.8 - 3.0). The survival curves differed when comparing seropositive and seronegative donors, young and old ($p < 0.001$). Out of the 159 deaths among seropositive donors, 26 were due to Chagas disease (ICD10 B57.0 - B57.5), 47 were due to other heart diseases (ICD10 - I10 to I80.2), with no mention of Chagas disease as an associated cause, and only 1 with Chagas disease mentioned in Part II of death certificate. **CONCLUSIONS:** Seropositive blood donors die at least 2.3 times more than seronegative donors. Causes of death in SIM may be misclassified for asymptomatic seropositive individuals.

DESCRIPTORS: mortality, Chagas disease, medical record linkage, database, blood donors, parasitic diseases.

1. Introdução

A doença de Chagas, também conhecida como tripanossomíases americana, foi descoberta em 1909 pelo cientista brasileiro Dr. Carlos Ribeiro Justiniano Chagas (1879 - 1934) e é causada pelo protozoário parasita *Trypanosoma cruzi* (*T.cruzi*) (OMS 2010). Classificada como uma das 17 doenças negligenciadas no mundo pela Organização Mundial da Saúde (OMS), sua transmissão ocorre pela picada de insetos da subfamília Triatomínea, chamados popularmente de barbeiros, e seu principal vetor no Brasil é a espécie *Triatoma infestans*. A transmissão também pode ocorrer por transfusão de sangue, transplante de órgãos, transmissão vertical, alimentos contaminados, e acidentes laboratoriais (OMS 2011).

Em seres humanos a doença de Chagas se manifesta em duas fases, a fase aguda e crônica. A fase aguda dura aproximadamente 2 meses e na maioria dos casos é assintomática. Nos casos em que os sintomas existem esses podem ser lesões de pele (chagomas), febre, dor de cabeça, dores musculares, dores abdominais, inchaço nas pernas e rosto, dificuldade em respirar entre outros (OMS 2011). Na fase crônica da doença, a maioria dos casos se apresenta sem sinais de sintomas. Indivíduos assintomáticos que tem resultados de eletrocardiograma (ECG) normais, raio x de torax, esôfago e cólon normais são considerados casos na forma indeterminada da doença (Comunicação 1985). Esta forma pode durar a vida toda de um indivíduo.

Com o tempo, cerca de 30 a 40% dos casos evoluem para a forma clínica, que se divide em forma cardíaca, forma gastrointestinal ou ambas (Rassi et al. 2010).

1.1. Epidemiologia

A estimativa do número de pessoas infectadas no mundo caiu de aproximadamente 20 milhões em 1981 para 10 milhões em 2009 (OMS 2011) e para 7 a 8 milhões em 2014 (Andrade et al. 2014). A maioria dos infectados vive em 21 países da América Latina: Argentina, Belize, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Equador, El Salvador, Guiana Francesa, Guatemala, Guiana, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, Suriname, Venezuela e Uruguai (OMS 2011). Dados de incidência e prevalência da doença de Chagas na população destes países estão descritos na Tabela 1 e são baseados em estimativas feitas pela OMS e Organização Panamericana de Saúde (OPAS) em 2010.

Tabela 1 - Parametros demográficos e epidemiológicos da doença de Chagas na América Latina por país, 2010

País	População	Casos Infectados	Taxa de Prevalência (%)	Taxa de Incidência por ano (%)	População Exposta em áreas Endêmicas
América do Sul					
Argentina	41 343 000	1 505 235	3,640	0,002	2 242 528
Bolívia	9 947 000	607 186	6,104	0,081	586 434
Brasil	190 755 799	1 156 821	0,030	0,008	25 474 365
Chile	17 095 000	119 660	0,699	0,000	
Colômbia	45 805 000	437 960	0,956	0,011	4 813 543
Equador	14 483 499	199 872	1,379	0,014	4 199 793
Guiana /Guiana Francesa / Suriname	1 501 962	12 600	0,838	0,018	377 258
Paraguai	8 668 000	184 669	2,130	0,003	1 703 659
Peru	28 948 000	127 282	0,439	0,007	1 290 415
Uruguai	3 301 000	7 852	0,237	0,000	
Venezuela	27 223 000	193 339	0,710	0,003	1 033 450
América Central					
Belize	315 000	1 040	0,330	0,003	70 252
Costa Rica	4 516 000	7 667	0,169	0,000	233 333
El Salvador	6 952 000	90 222	1,297	0,013	1 019 000
Guatemala	13 550 000	166 667	1,230	0,009	1 400 000
Honduras	7 989 000	73 333	0,917	0,011	1 171 133
Nicaragua	5 604 000	29 300	0,522	0,006	642 750
Panamá	3 557 687	18 337	0,515	0,004	466 667
América do Norte					
México	112 468 855	876 458	0,779	0,005	23 474 780
Total	544 023 802	5 815 500	1,055	0,005	70 199 360

FONTE: OMS - (OMS-2010)

Apesar da América Latina ser reconhecida como a única área endêmica da doença, nas últimas décadas, devido a migração e globalização, a doença de Chagas foi detectada em regiões não endêmicas: Canadá, Estados Unidos da América, Austrália, Japão, Bélgica, França, Itália, Espanha, Suíça, Reino Unido, Áustria, Croácia, Dinamarca, Alemanha, Luxemburgo, Países Baixos, Noruega, Portugal, Romênia e Suécia (OMS 2011). Dados de prevalência nas regiões consideradas não endêmicas são ainda escassos, mas alguns estudos tentam estimá-los com base na população de imigrantes dos países endêmicos e a prevalência também estimada em cada país de origem (Bern e Montgomery 2009; Schmunis e Yadon 2010; Angheben et al. 2011; Basile et al. 2011).

A doença de Chagas é comumente associada com pobreza em área rural onde o barbeiro é encontrado (Nunes et al. 2013). Porém desde 1970 / 1980 com a migração da população rural para as cidades o padrão epidemiológico da doença mudou para infecção urbana. Com a globalização da doença, o desafio de prevenir, controlar e tratá-la ficou mais complexo. Na América Latina iniciativas intergovernamentais nesse sentido existem desde 1991. São elas a Iniciativa do Cone Sul (1991); a Iniciativa dos Países Andinos e a Iniciativa de Países da América Central (1997); a Iniciativa de Países Amazônicos pela Vigilância e Controle da Doença de Chagas (2004) e a iniciativa do governo do México (2003). Em 2005 a doença de Chagas foi classificada pela OMS como doença negligenciada a fim de promover esforços para o controle e eliminação junto com as outras doenças assim classificadas. Em Julho de 2007 na reunião de Genebra,

promovida pela OMS e a OPAS, 28 países se propuseram a estabelecer a Iniciativa de Países Não Endêmicos para aumentar a rede de prevenção, controle e tratamento (OMS 2010).

No Brasil, no período entre 1980 e 2000, houve um grande esforço das autoridades no sentido de erradicar o inseto e interromper a transmissão transfusional. Na verdade, esse esforço teve início nos anos de 1940, mas as medidas de controle da transmissão transfusional só se espalharam pelo Brasil com o surgimento da AIDS e a generalização destas medidas pelo país (Dias e Coura 1997). As principais medidas de controle da doença implementadas foram a eliminação do *Triatoma infestans* peri e intradomiciliar por uso de inseticidas, melhoria habitacional, educação sanitária e a adoção de testes de triagem entre doadores de sangue, para detectar o parasita e evitar a transmissão (Coura e Dias 2009). As medidas adotadas no Brasil, que objetivavam a interrupção da transmissão pelo *Triatoma infestans*, o principal vetor da transmissão domiciliar no Brasil, foram bem sucedidas e reconhecidas pela a OPAS em 9 de Junho de 2006, quando a OPAS entregou ao Ministro da Saúde o Certificado Internacional da Eliminação da Transmissão da Doença de Chagas (Massad 2008).

1.2. Mortalidade

Segundo Lecorps e Paturet (1999)* citado por Fortes e Zoboli (2008) "Um dos objetivos da Saúde Pública é o estudo do processo saúde-doença, que visa promover, proteger e restaurar a saúde dos indivíduos e das populações." Para compreender a distribuição de doenças e de recursos de

*Lecorps P, Paturet J. Santé publique-Du biopouvoir à la démocratie. Paris:EHESP; 1999

saúde em uma população, dados como número de nascimentos, óbitos e internações são de extrema relevância. Dados de óbitos distribuídos por sexo, idade e causa de morte são indispensáveis para o entendimento do processo saúde-doença e para a determinação de políticas e saúde pública no mundo. Apenas em 1948 um modelo de declaração de óbito padronizado foi adotado internacionalmente. Antes disso, cada país registrava os óbitos ocorridos de formas diferentes (Brasil 2011), dificultando a comparação dos dados entre países.

No Brasil, o Ministério da Saúde adotou a Declaração de Óbito (DO) padronizada para todo o território nacional em 1976. O Sistema de Informação sobre Mortalidade (SIM), criado no ano anterior, consolida os dados das DOs gerados em nível municipal e tem como gestor a Secretaria de Vigilância em Saúde (SVS) (Brasil 2011).

O SIM vem sendo aperfeiçoado desde 1975. Hoje em dia o SIM contém informações sobre óbitos ocorridos em todo o Brasil, porém nos primeiros anos de sua operação os dados coletados abrangiam apenas as capitais, por não haver legislação que estabelecesse o fluxo e periodicidade de coleta e envio de dados (Jorge et al. 2007). Os dados de óbitos ocorridos no Brasil de 1996 em diante são disponibilizados, sem qualquer informação nominal dos falecidos, pelo Departamento de Informática do Sistema Único de Saúde (Datasus) e podem ser consultados pelo público na página da internet do Datasus (Saúde 2014). Desde 1996, o SIM contém mais de 1.000.000 de registros por ano. Registros com dados nominais ou seja, aqueles com variáveis que indicam a identidade dos indivíduos falecidos, estão

disponíveis apenas a partir de 2001 e somente pessoas devidamente autorizadas têm acesso a eles. As DOs são preenchidas exclusivamente por médicos. As informações das DOs sobre causa de morte estão separadas em 2 partes: a parte I destina-se às "doenças relacionadas com a cadeia de acontecimentos patológicos que levaram diretamente à morte", contendo um total de 4 linhas, e a Parte II destina-se às "afecções contribuintes, embora não relacionadas"(OMS 1993). As doenças listadas pelos médicos nas DOs são posteriormente classificadas pelo código de Classificação Internacional de Doenças - 10ª Revisão (CID-10), para posterior registro no SIM. Na parte I da declaração de óbito, a afecção registrada na última linha preenchida é geralmente considerada como a causa básica de morte, as outras afecções na Parte I do atestado são consideradas causas consequenciais e a primeira linha é a causa imediata de óbito. Já as afecções que constam na Parte II da declaração de óbito são consideradas causas contribuintes (Brasil 2001).

Para a pesquisa científica / epidemiológica um banco de dados que consolida as informações de morte de forma padronizada é de inestimável valor. O acesso de pesquisadores diretamente ligados às Secretarias Municipais de Saúde (SMS), às Secretarias Estaduais de Saúde (SES), às SVSs e ao Ministério da Saúde aos dados nominais do SIM são garantidos por lei e visa a continuação do aperfeiçoamento dos dados. Pesquisadores sem ligação direta com as Secretarias apenas têm acesso aos dados nominais mediante a aprovação do protocolo de pesquisa junto ao sistema Comitê de Ética em Pesquisa / Comissão Nacional de ética em Pesquisa (CEP/CONEP) e a devida autorização do Ministério da Saúde (Brasil. 2011; 2012).

1.3. Relacionamento probabilístico de registros

Estudos que relacionam os diversos bancos de dados dos sistemas de informação em saúde existentes no país tem contribuído tanto para melhorar a qualidade dos dados desses bancos quanto para o entendimento do processo saúde-doença. No entanto, a falta de um identificador único para cada indivíduo dificulta o relacionamento entre bancos. Newcombe (1959) foi um dos pioneiros a utilizar um método automático de relacionamento de registros. Seu estudo visava acompanhar grandes grupos de pessoas que haviam sido expostas a baixos níveis de radiação para conhecer as causas de morte, mutações naturais e fertilidade. Para tal, ele cruzou registros de nascimentos com registros de casamento da Província da Colúmbia Britânica no Canadá em 1955. Além de passar esses registros para um computador, ele comparou as datas de nascimento com as datas de casamento e os sobrenomes dos nascidos com os sobrenomes dos casais. Para otimizar a comparação de sobrenomes ele utilizou o método de indexação Soundex de Robert C. Russel (1922). Fellegi e Sunter (1969), Jaro (1989; 1995), Dean (1996) e Cook (2001) avançaram nos estudos e hoje o método de relacionamento de registros consiste basicamente em três etapas; padronização dos registros em ambas as bases, blocagem e pareamento dos registros de acordo com os blocos formados. Atualmente, existem ferramentas eletrônicas que permitem o relacionamento de registros de base de dados diferentes (Campbell et al. 2008; Jurczyk 2009). Camargo e Coeli (2000) desenvolveram uma ferramenta probabilística de código aberto que utiliza o Soundex adaptado para a fonética da língua portuguesa, visando associar registros considerando a fonética da

língua portuguesa. Esta ferramenta foi denominada de RECLINK (a versão III usada nesse estudo está disponível em http://www.iesc.ufrj.br/reclink/RecLink_arquivos/RecLinkdl.html) e vem sendo utilizada em trabalhos brasileiros para relacionar registros de vários bancos de dados na área da saúde (Coeli et al. 2003; Bierrenbach et al. 2010; Fonseca et al. 2010; Migowski et al. 2011).

Existem dois métodos de relacionamento de registros, o determinístico, no qual o relacionamento tem por base a concordância exata de uma ou várias variáveis presentes nas duas bases, e o probabilístico, que utiliza um escore cujo cálculo leva em consideração a similaridade entre as variáveis que estão sendo comparadas. Na lógica probabilística, pares com escores altos têm maior probabilidade de serem pares verdadeiros, isto é, pares em que ambos os registros pertencem a um mesmo indivíduo. Já pares com escores baixos tem menor probabilidade dos registros pertencerem ao mesmo indivíduo, sendo reconhecidos como pares falsos. O problema está em definir como par verdadeiro ou falso os pares com escores na zona intermediária, "zona cinza", os quais requerem normalmente uma revisão manual para serem corretamente classificados (OMS 2012). Dependendo de revisão manual frequentemente demanda um alto custo em termos de recursos humanos especializados para cumprir tal tarefa e compromete a privacidade dos dados. Para que a revisão manual seja excluída do processo de relacionamento probabilístico é necessária a determinação de uma nota de corte tal que todo o par formado com escore acima desta nota possa ser considerado par verdadeiro. A

determinação desta nota de corte tem que levar em consideração a variabilidade dos registros dos bancos que serão relacionados.

Em uma etapa preliminar, a validação do método de relacionamento probabilístico e determinação de uma nota de corte para a comparação dos campos nome, nome mãe e data de nascimento entre o SIM e o banco de dados formado com registros de doadores da Fundação Pró-Sangue - Hemocentro de São Paulo (FPS) e pacientes do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) foi realizado como parte essencial do trabalho dessa tese, com o objetivo de estudar a precisão de um processo de relacionamento probabilístico usado para rastrear doadores de sangue da FPS e pacientes do HC-FMUSP no banco de dados do SIM. Os métodos e resultados estão descritos abaixo (ver Métodos e Resultados Parciais).

Em estudos que visam avaliar o desempenho de um processo de relacionamento probabilístico, os resultados da revisão manual ou de um banco de dados com pares verdadeiros e falsos previamente conhecidos são muitas vezes considerados como o "padrão-ouro" contra o qual os escores de probabilidade calculados no relacionamento probabilístico são comparados. Entre as diversas medidas que têm sido utilizadas para avaliar a acurácia dos métodos de relacionamento probabilístico de registros estão a sensibilidade, o valor preditivo positivo (VPP) e a especificidade. A sensibilidade é a capacidade do processo de relacionamento reconhecer pares verdadeiro, a especificidade é a capacidade do processo de relacionamento não reconhecer pares falsos como pares verdadeiros. O VPP é a proporção de pares

verdadeiros entre o total de pares formados no processo de relacionamento. Sensibilidade e VPP são frequentemente chamados de "recall" e precisão (recall and precision) respectivamente no campo de relacionamento de registros (Christen 2007). Como nos estudos sobre testes diagnóstico, a alta precisão pode ser normalmente alcançada às custas de um valor de "recall" baixo e vice e versa. Neste trabalho preliminar, objetivou-se estudar a precisão de um processo de relacionamento probabilístico usado para rastrear doadores de sangue da FPS e pacientes do HC-FMUSP no banco de dados do SIM.

Utilizando o método de relacionamento probabilístico, em uma segunda etapa dessa tese, tivemos como meta relacionar os registros do SIM com os registros de doadores de sangue da FPS soronegativos e soropositivos para a doença de Chagas para conhecermos a taxa de mortalidade da doença no Brasil e suas causas básicas.

1.4. Mortalidade em doença de Chagas

A taxa de mortalidade em indivíduos na fase crônica da doença foi descrita por vários autores de diversos países e recentemente revistas por Rassi e cols.(2007). As taxas variam, consideravelmente de 0,2 a 19,2% anualmente. A revisão sistemática acima citada tinha por objetivo descrever os fatores preditores de mortalidade na fase crônica da doença, revisando estudos publicados entre janeiro de 1985 e fevereiro de 2006. Após a identificação dos artigos publicados neste período e a exclusão de 579 artigos que não satisfizeram os critérios de inclusão para a revisão, 12 artigos foram considerados. Destes, 7 artigos eram brasileiros, 3 venezuelanos, 1 argentino e

1 norte americano. Os tamanhos amostrais dos estudos variaram entre 25 e 987 indivíduos com idade média entre 43 a 59 anos. Alguns estudos incluíam indivíduos com e sem cardiopatia e outros apenas indivíduos com cardiopatia. Os estudos também diferiam quanto a descrição das causas de morte. Alguns apenas consideraram a causa básica de morte para a análise e outros consideraram as múltiplas causas de morte listadas nas declarações de óbito, ou seja, a causa básica de morte, as causas consequenciais e as causas contribuintes de morte. Assim como esperado, a taxa de mortalidade anual foi menor nos estudos que incluíam pacientes com e sem cardiopatias. A variação da taxa de mortalidade descrita nessa revisão provavelmente é devida às diferenças dos desenhos de estudo, incluindo critérios de inclusão de pacientes, de cada artigo considerado.

Muitos estudos foram feitos para avaliar a forma clínica da doença, porém pesquisadores também se esforçam para compreender a forma indeterminada da doença em particular (Lopes et al. 1981; Coura et al. 1985; Maguire et al. 1987; Dias 1989; Ribeiro e Rocha 1998; Manzullo e Chuit 1999; Dias et al. 2002; Ribeiro et al. 2014) e o prognóstico para indivíduos assintomáticos e nesta forma da doença em geral tem sido favorável (Acquatella et al. 1987; Maguire, Hoff, Sherlock et al. 1987; Ianni et al. 2001).

No Brasil estima-se que existam de 2 a 3 milhões de indivíduos infectados em 2013 (Barros Pereira Junior e Markman Filho 2014) dos quais 600.000 com complicações cardíacas ou gastrointestinais, causando aproximadamente 6.000 mortes por ano, ou seja, 0,6% de todas as mortes que ocorrem no país (Martins-Melo 2012). Essa estimativa tem como base as causas básicas de morte, consequenciais ou associadas e contribuintes do

CID-10 B57 e todas as subcategorias (B57.0 a B57.5) descritas no Sistema de Informação sobre Mortalidade (SIM) no Brasil entre 1997 e 2007 (Martins-Melo 2012; Martins-Melo et al. 2012).

Indivíduos que morreram por outras causas que não as classificadas no Capítulo 1 - Categoria B57 - Doença de Chagas, como por exemplo, CID I42.8 - Outras Cardiomiopatias ou I42.9 - Cardiomiopatia não especificada, e que não tiveram os códigos B57.0 a B57.5 mencionados no atestado de óbito como causa básica ou associada, ou como causa consequencial ou contribuinte, não foram incluídos nos cálculos dessa estimativa. Indivíduos assintomáticos que nunca foram diagnosticados com a doença também não foram incluídos. Muitos indivíduos assintomáticos com a doença de Chagas podem não saber que contraíram a doença e assim serem potenciais transmissores do *T. cruzi* por transfusão de sangue, transmissão vertical ou doação de órgãos (Bern et al. 2011), bem como podem morrer sem nunca terem sido diagnosticados com a doença.

Em recente revisão da literatura e meta análise sobre o aumento de mortalidade devido a doença de Chagas (Cucunuba et al. 2016) os resultados indicam que existe um aumento de risco tanto para indivíduos sintomáticos quanto para indivíduos assintomáticos.

A população de doadores de sangue difere das populações da maioria de outros estudos por conter indivíduos assintomáticos classificados como infectados nos testes de triagem na hora da doação (Sabino et al. 2013). Um estudo realizado entre doadores de primeira vez da FPS no período de 2007 a 2008 mostrou que a prevalência de soropositividade para Chagas entre

doadores de primeira vez na FPS estava ao redor de 26/100.000 doações. (Sabino et al. 2010). Nesse trabalho, Sabino e col. consideraram para o cálculo da prevalência não só os casos confirmados positivos para a doença, como também os casos classificados como prováveis positivos para a doença pelos testes sorológicos. Doadores de sangue soropositivo para doença de Chagas não podem ser classificados como indivíduos com a forma indeterminada da doença por não serem submetidos a exames eletrocardiográficos e de raio X torácico durante a triagem. Contudo, como estes candidatos a doação tem que passar por critérios de elegibilidade para doação, que excluem doenças cardíacas sintomáticas, podemos considerar que eles também não podem ser classificados como indivíduos com a forma cardíaca da doença (Sabino, Ribeiro, Salemi et al. 2013).

2. Objetivo

Este trabalho se dividiu em duas etapas: o estudo preliminar, que objetivou validar o método de relacionamento probabilístico entre os bancos de dados da FPS/HC-FMUSP e o SIM; e o estudo sobre a mortalidade de doadores soropositivos para a Doença de Chagas de sangue da FPS entre 1996 e 2009.

2.1. Objetivo Estudo Preliminar

Esse estudo preliminar visou validar a metodologia de relacionamento probabilístico entre os bancos de dados da FPS/HC-FMUSP e o SIM e estabelecer uma nota de corte para essa metodologia. Essa nota de corte objetivava maximizar a sensibilidade e concomitantemente minimizar a necessidade de revisão manual em estudos futuros que usassem a metodologia de relacionamento probabilístico de banco de dados para descrever as taxas de mortalidade de doenças transmissíveis por transfusão de sangue em brasileiros.

2.2. Objetivo Estudo Mortalidade

Esse estudo visa comparar a taxa de mortalidade e as causas de óbito de candidatos a doadores de sangue da FPS soropositivos para doença de Chagas e a doadores soronegativos da FPS num período de 9 - 14 anos.

Objetivos:

Identificar se existe diferença entre as taxas de mortalidade por todas as causas de morte entre doadores de sangue soropositivos da FPS entre 1996 e 2000 para a doença de Chagas e doadores soronegativos por um período de 9 a 14 anos.

Descrever as causas dos óbitos ocorridos de doadores de sangue soropositivos e soronegativos para a doença de Chagas seguidos por um período de 9 a 14 anos.

3. Métodos

3.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico

Desenho do Estudo

Para validar o método desenvolvido para utilização da ferramenta de Relacionamento Probabilístico de Registros entre o banco de dados SIM, e o banco de dados composto por registros de doadores de sangue da FPS e pacientes do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP), calculamos a sensibilidade e especificidade do método e definimos uma nota de corte.

Os bancos de dados utilizados neste estudo foram:

- 3.1.1. *Arquivo Comparação*: SIM versão 19/04/2007
- 3.1.2. *Arquivo Referência*: banco formado por indivíduos vivos (doadores de sangue da FPS em 2005) e mortos (pacientes do HC-FMUSP). (ver Sujeitos do Estudo)

O método consistiu em limpar os campos dos dois bancos, relacioná-los probabilisticamente, e gerar um arquivo de pares formados com escores e informações provenientes dos dois bancos. A partir dos escores dos pares formados definiu-se a nota de corte entre pares verdadeiros e falsos.

Tamanho da Amostra e Poder estatístico do estudo

Para um estudo descritivo no qual se esperava uma sensibilidade e especificidade de 90%, uma amostra de 196 (em cada grupo- vivos e mortos) permitiu definir estes parâmetros com uma precisão 85.8%-94.2% com nível de confiança de 95% (Hulley 2008).

Sujeitos do Estudo

O *arquivo referência* foi criado somando 200 registros aleatoriamente selecionados do banco de dados da FPS a 196 registros de pacientes (todos) do banco de dados da Clínica de Hematologia do HC-FMUSP que morreram durante a internação. No primeiro banco de dados, os 200 registros aleatoriamente selecionados entre registros de indivíduos de 18 anos de idade ou mais velhos e que doaram sangue em 2007 foram considerados como registros do arquivo de referência de indivíduos VIVOS no período de 2001 a 2005. No segundo banco de dados utilizado para compor o *arquivo referência*, os 196 registros de todos indivíduos de 18 anos de idade ou mais velhos que morreram durante a internação entre de 2001 e 2005 na Clínica de Hematologia do HC-FMUSP foram considerados indivíduos MORTOS. Considerando a razão de 200 para 196 indivíduos, nós artificialmente criamos um cenário no qual a taxa de mortalidade era próxima de 50%. O *arquivo comparação* foi o banco de dados do SIM (versão 19/04/2007). Para que o SVS disponibilizasse o SIM para este estudo, um Termo de Responsabilidade (apêndice 2) foi assinado pelo diretor técnico da FPS pela guarda e sigilo dos dados e declarando para quais finalidades os dados seriam utilizados. Esta versão do SIM abrange o período de 2001 a 2006, contendo registros de todos os brasileiros que morreram com a idade de 18 anos ou mais. Por motivo de

atrasos nas notificações de óbito, os dados de 2006 ainda estavam incompletos em abril de 2007.

Pré-Processamento dos dados.

Tanto o arquivo referência quanto o arquivo comparação continham as variáveis número identificador, nome, nome da mãe, sexo e data de nascimento entre outras, sendo que o número identificador não era o mesmo nos dois arquivos. Estas variáveis passaram por uma extensa etapa de pré-processamento que teve como meta limpar e padronizar os dados que foram utilizados no processo de relacionamento probabilístico. Durante o pré-processamento dos dados utilizamos a ferramenta SQL-Analysis Server (Microsoft SQL Server 2008), e realizamos as seguintes transformações: registros nos quais a idade era menor que 18 anos o valor da variável idade foi transformado para 9999, e logo após foram excluídos registros com idade = 9999 (496.056 registros excluídos no SIM e 9 no arquivo referência); todos os registros com nome de mãe em branco foram preenchidos com "***" (222.044 registros no SIM e 0 no arquivo referência); o valor 9 foi definido para toda a informação inválida ou em branco da variável sexo, os dados de data de nascimento nulos foram transformados em "18000101"(47.533 registros do SIM e 0 do arquivo referência). Importante ressaltar que registros com nomes em branco ou inválidos foram excluídos (477.174 registros) dos passos de relacionamento subsequentes. Os registros excluídos durante o pré-processamento representaram 16,9% do total dos registros do SIM nesta versão. Foram considerados registros inválidos, os registros que para o campo

nome estavam preenchidos com "IGNORADO", "IGNORADA", "FALECIDA", "FALECIDO", "IGNORADO PELO DECLARANTE", "IGNORADA NO TERMO CASAMENTO", "IG", "IGM", "IGN", "*****", "\", "~IG", "] IG", "NAO SABE", "***", "EM BRANCO" entre outros. Todos os valores originais dos campos foram mantidos, porém estes campos foram renomeados para "nomedocampo_orig". O SqlL permite que novos valores possam ser adicionados a lista de valores inválidos sempre que um deles é encontrado por revisão manual. Isso permite que a limpeza dos dados seja contínua e acumulativa. Dos dados do SIM de 2001, 27,9% dos registros foram excluídos por nomes inválidos ou em branco, porém dados do SIM de 2002, 2003, 2004, 2005 e 2006 apenas 4% dos registros foram excluídos pelo mesmo motivo. Após os passos de limpeza o arquivo comparação continha 4.775.164 registros e o arquivo referência continha 396 registros (Figura 1).

Figura 1 - Descrição da limpeza de dados dos arquivos referência e comparação utilizados no processo de relacionamento probabilístico

FPS: Fundação Pró-Sangue Hemocentro de São Paulo; HC-FMUSP: Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo; SIM: Sistema de Informação sobre Mortalidade versão 19/04/2007

Estratégia do processo de relacionamento probabilístico de registros

Utilizando o RECLINK-III, mais uma etapa de padronização dos campos dos dois arquivos, referência e comparação, foi realizada. Todos os caracteres dos campos nome, nome da mãe, sexo foram transformados em caracteres com letras maiúsculas. Também foram removidos acentos ortográficos, preposições e espaços duplos (exemplo: "Maria Conceição dos Santos" → "MARIA CONCEICAO SANTOS"). Os formatos do campo data de nascimento foram padronizados para o formato AAAAMMDD, e finalmente todas as vírgulas e outros tipos de pontuação foram eliminados. A partir da variável nome, quatro outras variáveis secundárias foram geradas: primeiro nome, último nome e os códigos fonéticos tanto do primeiro quanto do último nome. Três passos de blocagem utilizando combinações de códigos fonéticos de primeiro nome, último nome e sexo foram realizados (Jaro 1989). A comparação de registros apenas entre blocos faz com que o esforço

computacional seja bem reduzido, pois reduz o número de pares que serão comparados durante o processo de relacionamento. No primeiro passo de blocagem selecionamos o código fonético do primeiro nome (PBLOCO), o código fonético do último nome (UBLOCO) e sexo como chave de blocagem, ou seja, neste passo apenas registros do mesmo sexo, códigos fonéticos do último e código fonético do primeiro nomes foram comparados. No segundo passo selecionamos para a chave e blocagem apenas o código fonético do primeiro nome e sexo. E finalmente no terceiro passo selecionamos para a chave de blocagem o código fonético do último nome e sexo. As variáveis de pareamento utilizadas nos três passos de blocagem foram o nome, nome da mãe e data de nascimento. RECLINK-III estima escores para cada par formado sendo que o escore mais alto representa a probabilidade mais alta do par ser um par verdadeiro. Os escores são a proporção do logaritmo de base 2 da razão de verossimilhança entre as probabilidades m_i e μ_i , que derivam de uma abordagem de relacionamento probabilístico padrão. Nesta abordagem m é a probabilidade que a sequência de caracteres de uma variável de pareamento em particular concorde para pares que são verdadeiros (m é análogo à sensibilidade), e μ a probabilidade que a sequência de caracteres de uma variável de pareamento em particular concorde para pares que são falsos (μ é análogo a 1 - especificidade) por acaso. Quando o conteúdo das variáveis concorda entre os registros comparados, o escore é proporcional ao logaritmo (m / μ), e quando discordam o escore é proporcional ao logaritmo $((1 - m) / (1 - \mu))$. Utilizamos o algoritmo de distância de Levenshtein para comparar os caracteres das variáveis nome e nome da mãe, o qual retornou valores entre 1 (concordância total) e 0 (discordância total), com 92% de sensibilidade (m) e

1% de especificidade (μ), com a proporção de concordância mínima de 85%. Para a variável data de nascimento utilizamos o algoritmo de caractere. Este algoritmo retorna valores entre 1 (concordância total) e 0 (discordância total), com 90% de sensibilidade (m) e 5% de especificidade (μ), com a proporção de concordância mínima de 65%. Os escores para cada variável de pareamento foram somados para determinar o escore do par formado. A concordância total entre as três variáveis de pareamento de um par de registros resultou em um escore máximo $\log_2 (0.92/0.01) + \log_2 (0.92/0.01) + \log_2 (0.9/0.05) = 17.22$, enquanto a discordância total das três variáveis resultou num escore mínimo de $\log_2 (0.08/0.99) + \log_2 (0.08/0.99) + \log_2 (0.1/0.95) = -10.51$. (Camargo e Coeli 2000)

Pós Processamento

A revisão manual dos pares formados em cada passo de blocagem foi realizada no módulo "Combinar" do aplicativo RECLINK-III. Neste módulo, o usuário pode visualizar os pares formados, seus escores, as variáveis de pareamento utilizadas e outras variáveis pertencentes aos registros que não foram utilizadas no processo de relacionamento. Todos os pares formados passaram por revisão manual, exceto os pares que obtiveram escores -10.51 (97,5% de 709.550 pares do primeiro passo de blocagem, 99,9% de 1.448.313 no segundo passo de blocagem) que foram considerados pares falsos sem revisão manual.

Após a classificação dos pares em pares verdadeiros ou falsos, RECLINK-III automaticamente gerou 3 arquivos: um arquivo com os pares verdadeiros, um arquivo com os registros do arquivo referência que não

formaram um par verdadeiro e um arquivo com os registros do arquivo comparação que não formaram um par verdadeiro. Os registros dos dois últimos arquivos foram relacionados no próximo passo de blocagem. Após a realização dos três passos de blocagem os três arquivos gerados com pares verdadeiros foram associados em um só arquivo para análise. Todos os passos do processo de relacionamento probabilístico foram realizados de forma que o status vital dos indivíduos do arquivo referência era desconhecido para o pesquisador.

Análise dos Dados

Para testar a performance da nossa estratégia de relacionamento probabilístico de registros, calculamos a sensibilidade e especificidade e seus intervalos de confiança binomial de 95% assumindo nota de corte para cada escore. Sensibilidade foi definida como a proporção de pares verdadeiros entre os 196 indivíduos sabidamente MORTOS (que tinham registros correspondentes no SIM), e especificidade foi definida como a proporção de pares falsos entre os 200 indivíduos sabidamente VIVOS (e que não deveriam ter registro correspondente no SIM). Para cada nota de corte escolhida, o VPP também foi calculado como a proporção de pares verdadeiros entre a soma do total de pares formados (pares verdadeiros + pares falsos) num intervalo de confiança binominal de 95%. Como esperado, para cada registro do arquivo referência muitos pares foram gerados, mas apenas os pares com o escore mais alto de cada par gerado foi considerado. Quando dois pares do mesmo registro do arquivo de referência obtinham o mesmo escore, apenas o par considerado por revisão manual como o par verdadeiro mais provável foi

considerado. A escolha da nota de corte levou em conta a necessidade de maximizar a especificidade do processo do relacionamento e minimizar a possibilidade de gerar pares falsos, porém aceitando se necessário algum nível de fracasso em encontrar pares verdadeiros.

Considerações Éticas:

Este estudo utilizou registros de dois bancos de dados já existentes. As variáveis: nome e nome da mãe foram utilizados somente com o propósito de associação destes registros. Tão logo os pares foram formados e classificados como pares verdadeiros ou pares falsos, estes campos foram apagados do arquivo associado e identificados pelo número de registro do doador ou pelo número do registro de óbito.

A Informação sobre os sujeitos desta pesquisa tem caráter confidencial e o computador que armazena os bancos de dados utilizados no estudo tem acesso restrito (usuário/senha) e não disponibiliza acesso a internet.

Estas medidas foram tomadas pelos pesquisadores como um esforço para garantir a confidencialidade dos dados, antes, durante e depois do estudo. A aprovação do Comitê de Ética do HC-FMUSP e CONEP para este estudo foi obtida (CAAE: 0543.0.015.000-08 / Parecer nº 1006/2008 de 16/12/2008). O estudo de validação do método de relacionamento probabilístico foi aprovado primeiramente para o cruzamento do SIM com o banco de todos os receptores de sangue do FPS/HC-FMUSP em 2004.

Riscos:

O risco potencial para os sujeitos da pesquisa foi mínimo, já que toda informação foi obtida em bancos de dados já existentes. O único risco seria o de perda accidental da confidencialidade referente ao status de vivo/morto do sujeito e sua causa mortis. Como mencionado anteriormente, os pesquisadores se esforçaram para garantir esta confidencialidade.

Benefícios:

A validação deste método torna possível a identificação de óbitos entre sujeitos de outros estudos que envolvam doadores de sangue da FPS e/ou pacientes do HC-FMUSP.

Termo de Consentimento Livre e Esclarecido (TCLE):

Por consenso, a CONEP e o CEP do HC-FMUSP consideraram desnecessário o termo de consentimento livre e esclarecido para estudos que revêem prontuários, dada a dificuldade da obtenção do mesmo e a importância deste tipo de estudo. Assim sendo, não houve necessidade de consentimento do sujeito neste estudo. A permissão da utilização do SIM (Sistema de Informação de Mortalidade) de 2001 – 2006 com nomes disponíveis foi concedida pelo Ministério da Saúde do Brasil. A permissão de uso do banco de dados de doadores da Fundação Pró-Sangue também foi concedida. Ambas as permissões estão disponíveis no apêndice 2.

3.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo

Desenho do Estudo

Foi conduzido um estudo de coorte retrospectivo de doadores de sangue da FPS de 1996 a 2000.

População do estudo

A coorte retrospectiva foi composta por todos os doadores da FPS do período de 1996 a 2000 diagnosticados como soropositivos para a doença de Chagas (n=2842, denominados "com fator de exposição" ou soropositivos) e por doadores soronegativos (n=5684, denominados "sem fator de exposição" ou soronegativos) para todos os testes de triagem de rotina. Para cada doador de sangue soropositivo para Doença de Chagas, dois doadores soronegativos foram aleatoriamente selecionados entre os doadores de sangue soronegativos para Chagas que também fossem soronegativos para as demais doenças dos testes de triagem e que doaram sangue no mesmo semestre dos doadores soropositivos.

Para que uma pessoa seja aceita como doador de sangue na FPS, ela deve passar por um processo de triagem que obedece as normas nacionais e internacionais (MS 2011) e que envolve as seguintes etapas: Ao chegar ao hemocentro os sinais vitais do candidato à doação são medidos: Peso, temperatura, pressão arterial, batimentos cardíacos, e anemia (hematócritos e hemoglobina). Só após os sinais vitais serem considerados normais pela equipe do hemocentro o candidato será encaminhado para a segunda etapa de triagem. Nesta etapa o candidato passa por uma entrevista com uma enfermeira em local privado. A entrevista tem por objetivo avaliar se a doação pode trazer riscos ao doador ou ao receptor da bolsa doada. Neste momento o candidato à doação é questionado sobre doenças anteriores, comportamentos

de risco quanto a algumas doenças infecciosas e transmitidas por transfusão. Suas respostas são armazenadas no banco de dados do hemocentro. Nesta etapa, se o candidato estiver ciente que tem a Doença de Chagas e responder honestamente às perguntas, será recusado para doação e não continuará o processo de triagem. No entanto, se o candidato não souber do seu status em relação à doença, ou mentir, será conduzido à próxima etapa. O candidato à doação é então conduzido à frente de um computador e em sigilo absoluto responde se se considera parte do grupo de risco para doenças transmissíveis pelo sangue. Se a resposta do candidato é positiva, o candidato à doação é conduzido para a próxima etapa, a coleta de sangue e exames sorológicos de triagem e sua bolsa é posteriormente descartada pelo hemocentro. Se a resposta é negativa, o sangue é coletado, e na última etapa de triagem uma amostra de seu sangue é testado para várias doenças transmissíveis por transfusão sanguínea. Na FPS as seguintes doenças são testadas: Sífilis (SIF), Vírus da imunodeficiência humana (HIV), Vírus Linfotrópico da célula Humana I e II (HTLV I e II), Vírus da Hepatite C (VHC), Vírus da Hepatite B (VHB), e Doença de Chagas. A bolsa de sangue só é liberada para transfusão se todos os testes de triagem forem negativos. Quando qualquer um deles for positivo, a bolsa é descartada e o candidato à doação é convocado para voltar ao hemocentro e colher uma segunda amostra para testes confirmatórios. Baseada nos resultados destes testes a equipe do hemocentro aconselha e encaminha o candidato a tratamento em unidade de serviço público adequado. A Tabela 2 descreve a distribuição dos sujeitos selecionados por semestre.

Tabela 2: Distribuição de doadores soropositivos e soronegativos para doença de Chagas por semestre.

Semestre	TOTAL DOADORES	SOROPOSITIVOS Doadores Soropositivos para Doença de Chagas	Total de Doadores com Sorologia Realizada	SORONEGATIVOS DISPONÍVEIS Total de Doadores Soronegativos para todos os testes de triagem	SORONEGATIVOS Nº Necessários para o estudo	SORONEGATIVOS Selecionados Aleatoriamente	TOTAL SUJEITOS
1º sem 1996	135.988	343	106.138	86.799	686	684	1.027
2º sem 1996	131.915	428	105.035	86.756	856	855	1.283
1º sem 1997	138.342	366	107.582	90.880	732	730	1.096
2º sem 1997	147.286	351	116.345	99.153	702	706	1.057
1º sem 1998	140.645	303	110.860	94.458	606	600	903
2º sem 1998	148.136	275	120.650	100.923	550	557	832
1º sem 1999	134.038	210	106.791	92.680	420	418	628
2º sem 1999	128.633	196	108.180	96.720	392	391	587
1º sem 2000	124.161	203	103.557	92.541	406	407	610
2º sem 2000	125.425	167	104.553	94.024	334	336	503
TOTAL	1.354.569	2.842	1.089.691	934.934	5.684	5.684	8.526

Definição da exposição

Para definir se os doadores eram soropositivos e soronegativos para Chagas e para as demais doenças testadas na rotina do serviço da FPS, foram realizados os testes de triagem acima mencionados. Para a doença de Chagas três testes comerciais foram utilizados: imunofluorescência indireta (IFI), hemaglutinação indireta (HI) e ensaio imunoenzimático (ELISA). No período de 1996 a 2000 os testes IFI e HI utilizados pela FPS foram ambos produzidos pelo Biolab, Rio de Janeiro Brasil. Nos anos de 1996 a 1998 o ELISA utilizado foi o produzido pela Hemobio Chagas, Embrabio São Paulo, Brazil, em 1999 o ELISA utilizado foi o Chagasnostika, produzido pela Organon Tecknica, Rio de Janeiro, Brazil e finalmente em 2000 o teste ELISA utilizado pela FPS na triagem de doadores foi o Chagatest, produzido pela Wiener, Rosario, Argentina. O doador foi considerado soropositivo para a doença de Chagas se sua amostra foi reativa em todos os testes realizados (Sabino et al. 2003). Para as outras doenças testadas na triagem de doadores o ELISA foi realizado.

Verificamos entre os doadores soropositivos para a doença de Chagas para quais outras doenças estes também testaram positivos. (Tabela 3).

Tabela 3 - Co-infecções de doadores soropositivos para doença de Chagas da FPS entre 1996 a 2000.

Ano Doação / Sorologia Positiva	1996	(%)	1997	(%)	1998	(%)	1999	(%)	2000	(%)	Total	(%)
CHAGAS	736	(95,5)	682	(95,1)	561	(97,1)	389	(95,8)	346	(93,5)	2 714	(95,50)
CHAGAS E SIF	14	(1,8)	9	(1,3)	7	(1,2)	7	(1,7)	10	(2,7)	47	(1,65)
CHAGAS E VHB	12	(1,6)	9	(1,3)	2	(0,3)	1	(0,2)	8	(2,2)	32	(1,13)
CHAGAS E VHC	2	(0,3)	6	(0,8)	2	(0,3)	3	(0,7)	6	(1,6)	19	(0,67)
CHAGAS E HTLV	4	(0,5)	6	(0,8)	4	(0,7)	3	(0,7)	0		17	(0,60)
CHAGAS E VHC E	0		2	(0,3)	0		1	(0,2)	0		3	(0,11)
CHAGAS E VHC E	0		1	(0,1)	1	(0,2)	0		0		2	(0,07)
CHAGAS E VHC E	1	(0,1)	0		0		1	(0,2)	0		2	(0,07)
CHAGAS E VHB E	0		1	(0,1)	0		1	(0,2)	0		2	(0,07)
CHAGAS E HIV	1	(0,1)	0		0		0		0		1	(0,04)
CHAGAS E VHC E	1	(0,1)	0		0		0		0		1	(0,04)
CHAGAS E HIV E H	0		0		1	(0,2)	0		0		1	(0,04)
CHAGAS E HTLV E	0		1	(0,1)	0		0		0		1	(0,04)
TOTAL	771	100	717	100	578	100	406	100	370	100	2 842	100

Definição do desfecho

Para identificar os óbitos entre os 8526 doadores de sangue da FPS, realizamos o relacionamento probabilístico de registros do banco dos doadores com o banco do Sistema de Informação de Mortalidade (SIM), utilizando o método previamente validado (ver estudo de acurácia da metodologia de relacionamento probabilístico - apêndice 1). A única diferença é que apesar de considerarmos a nota de corte sugerida no estudo da validação, tivemos a preocupação de também realizar a revisão manual de todos os pares que obtiveram escore acima de 6,0 ou igual a 6,0.

O banco de mortalidade contém todos os registros nacionais de óbitos ocorridos nos anos de 2001 a 2009. Esse banco foi extraído em novembro de 2010. As informações de óbito no Brasil para o ano de 2009 ainda não são consideradas completas nessa versão, pois nem todos os óbitos ocorridos no

Brasil em 2009 haviam sido registrados no SIM. Esta versão do SIM, quando comparada com os dados da página da internet do Datasus difere em número total de registros em 12.933 registros a menos, sendo a distribuição desta diferença por ano de 1.677 em 2001, 169 em 2002, -483 em 2004, 3.205 em 2005 e 8.365 em 2009. O total de registros no SIM na versão entregue a nós continha 9.224.216 registros e os dados da página da internet do Datasus visitado em 27 de Setembro de 2014 para o mesmo período continha 9.237.149 registros.

Os dados do SIM de 2001 a 2009 versão Novembro de 2010 foram enviados pelo MS em 9 arquivos, ou seja, por ano. Utilizamos a ferramenta SQL-Analysis Server (Microsoft SQL Server 2008) para limpar e transformar os campos que foram utilizados no processo de relacionamento probabilístico utilizando o mesmo script utilizado no estudo preliminar (ver pg 16), porém apenas excluimos registros nos quais o nome era inválido (ver pg 19) ou nulo e registros de indivíduos cuja idade era menor ou igual a 1 ano, apesar de esperarmos apenas encontrar óbitos de indivíduos maiores de 18 anos de idade. Optamos por trabalhar com todos os registros do SIM a não ser os registros de indivíduos de 0 a 1 ano, pois para esta idade o campo nome contém muitos nomes inválidos. A Tabela 4 descreve os passos de limpeza realizados no SIM para este relacionamento de registros e o número de registros do SIM que foram excluídos antes do processo de relacionamento probabilístico.

Tabela 4: Distribuição de registros de óbitos por ano

Ano	Arquivo Original	Nomes Nulos	Menores que 1 ano	Arquivo Comparação	Datasus
1996	0	0	0	0	908 883
1997	0	0	0	0	903 516
1998	0	0	0	0	931 895
1999	0	0	0	0	938 658
2000	0	0	0	0	946 686
2001	959 815	239 383	53 714	666 718	961 492
2002	982 638	3 198	61 990	917 450	982 807
2003	1 002 340	2 488	61 100	938 752	1 002 340
2004	1 024 556	2 095	57 623	964 838	1 024 073
2005	1 003 622	1 880	54 257	947 485	1 006 827
2006	1 031 691	0	52 295	979 396	1 031 691
2007	1 047 824	0	48 949	998 875	1 047 824
2008	1 077 007	0	47 493	1 029 514	1 077 007
2009	1 094 723	0	45 722	1 049 001	1 103 088
2010	0	0	0	0	1 136 947
2011	0	0	0	0	1 170 498
2012	0	0	0	0	1 181 166
Total	9 224 216	249 044	483 143	8 492 029	17 355 398

Fonte: <http://tabnet.datasus.gov.br/cgi/defthtm.exe?sim/cnv/obt10uf.def>

- 27/09/2014 ■ Dados do SIM não nominais

O arquivo de doadores de sangue que foi submetido ao relacionamento probabilístico não continha uma variável que mostrasse o status do resultado dos testes de triagem, para que o pesquisador responsável por realizar o processo de relacionamento não fosse influenciado por esse conhecimento.

Os sujeitos do estudo foram pareados por semestre de doação visto que o número de doadores soropositivos para a doença de Chagas em 1996 foi bastante superior ao número de doadores soropositivos em 2000 (1996→771 soropositivos, 2000→370 soropositivos).

Os campos do SIM linhas A (1-4), B (1-4), C (1-4), D (1-4), se referem às informações da Parte I da declaração de óbito, ou seja causas consecuenciais de morte; os campos linha II (1-4) se referem às informações da Parte II da declaração, causa contribuinte de morte; e o campo Causa Básica é a causa selecionada como causa básica de morte seguindo os

procedimentos descritos no manual de instruções do CID-10 Utilizamos definições de causa de óbito diferentes para a estatística descritiva da coorte e para a estatística analítica descritas a seguir.

Para a estatística descritiva, os óbitos foram classificados em seis grupos: (1) doença de Chagas, (2) neoplasias, (3) doenças cardíacas, (4) outras doenças, (5) traumatismos e (6) mal definidas e inespecíficas baseado na descrição de causa básica de morte. O número do DO, a data do óbito, a causa básica de morte e as linhas A (1-4), B (1-4), C (1-4), D (1-4), e II (1-4) dos registros de óbito encontrados no SIM foram associados ao arquivo gerado com os registros dos pares verdadeiros. Classificamos as categorias dos códigos do CID-10 em 6 grupos para analisarmos nossos dados: Grupo 1 - doenças de Chagas, Grupo 2 - neoplasias, Grupo 3 - doenças cardíacas, Grupo 4 - outras doenças, Grupo 5 - traumatismos, Grupo 6 - mal definidas e inespecíficas. A figura 2 descreve os capítulos e códigos do CID 10 considerados em cada grupo. Os códigos CID-10 U049-U99, do Capítulo XXII - Códigos para propósitos especiais, que são utilizados para novas doenças de etiologia incerta; e os códigos Z000-Z999, do Capítulo XXI - Fatores que influenciam o estado de saúde e o contato com os serviços de saúde, não foram classificados nestes grupos.

Figura 2 - Classificação dos códigos CID-10 nos grupos 1 a 6

Utilizamos ainda três definições de causas de óbitos para melhor compreender a distribuição de óbitos no SIM por CID 10.

A. Óbitos confirmados por Chagas

Somente foram considerados óbitos confirmados de Chagas os doadores cujos registros pareados de óbito indicaram a presença dos seguintes códigos do CID-10 listados como causa básica: B57.0 - Forma aguda da doença de Chagas, com comprometimento cardíaco; B57.1 - Forma aguda da doença de Chagas, sem comprometimento cardíaco; B57.2 - Doença de Chagas (crônica) com comprometimento cardíaco; B57.3 - Doença de Chagas (crônica) com comprometimento do aparelho digestivo; B57.4 - Doença de Chagas (crônica) com comprometimento do sistema nervoso; B57.5 - Doença de Chagas (crônica) com comprometimento de outros órgãos.

B. Óbitos confirmados e possíveis por Chagas

Nessa definição, além dos códigos do CID-10 listados na definição anterior, foram somados códigos de doenças cardíacas que indicam uma possível causa chagásica, mesmo que não haja menção específica a essa doença. Esses códigos foram selecionados pelo Dr. Antonio Luiz Ribeiro, professor da Universidade Federal de Minas Gerais, pesquisador cardiologista e considerado um "expert" na Doença de Chagas por seus colegas. O Dr. Luiz Antonio Ribeiro não teve acesso à lista das causas de óbitos ocorridos entre os doadores soropositivos e soronegativos para doença de Chagas para fazer a seleção. Esses códigos são os seguintes: I42.0 - Cardiomiopatia dilatada, I42.2 - Bloqueio atrioventricular total, I47.0 - Arritmia ventricular por reentrada, I47.2 - Taquicardia ventricular, I49.0 - Flutter e fibrilação ventricular, I50.0 -

Insuficiência cardíaca congestiva, I50.1 - Insuficiência ventricular esquerda, I51.7 - Cardiomegalia e I50.9 - Insuficiência cardíaca não especificada.

C. Óbitos por toda as causas: não importando quais eram as causas básicas de óbito.

É importante mencionar que não foi possível obter os dados nominais do SIM para o período de 1996 a 2000, apesar de termos solicitado ao SVS o SIM com registros entre 1996 e 2009, conforme Termo de Responsabilidade que consta no apêndice 2 desta tese. Os registros de óbitos com dados nominais para o período entre 1996 a 2000 ainda não estavam disponíveis nesta versão do SIM. Ou seja, para os doadores de 1996, não temos informação sobre eventuais óbitos que tenham ocorrido nos 5 primeiros anos de seguimento. Já para os doadores de 1997, não temos informação sobre eventuais óbitos que tenham ocorrido nos 4 primeiros anos de seguimento, e assim por diante. O tratamento usado para circunscrever essa limitação será detalhado a seguir.

Tamanho da Amostra e Poder estatístico do estudo

Como nesse estudo utilizamos os dados de todos os doadores de sangue soropositivos para doença de Chagas no período de 1996 a 2000 (n=2842), o tamanho da população já estava definido. Estipulando o poder estatístico em 80%, com este tamanho de amostra pré-definido, restou calcular qual a magnitude de efeito que o estudo poderia detectar para identificar a diferença entre as taxas de mortalidade de doadores soropositivos e doadores soronegativos. Em um estudo de coorte retrospectiva com uma taxa de mortalidade da população em geral de 6,38 mortes / 1000 habitantes

(Indexmundi 2011), uma amostra de 2842 doadores de sangue soropositivos para doença de Chagas e de 5684 doadores soronegativos tem poder de 80% para detectar uma diferença relativa de 2,00 com um nível de significância de 0,05. Ou seja, se considerarmos que o número de doadores soronegativos que morreram no período for na mesma proporção que o número de pessoas que morrem na população do país neste mesmo período, poderemos detectar uma diferença absoluta de 6,4 mortes / 1000 habitantes entre o número de doadores soropositivos e soronegativos que morreram neste mesmo período, com um poder estatístico de 80% e nível de significância de 0,05.

Análise dos Dados:

Idade e sexo eram as duas únicas variáveis que representavam potenciais fatores de risco e / ou fatores de confusão da associação entre os resultados e a principal variável explicativa sob investigação - status sorológico - disponível para análise no banco de dados FPS. A idade foi analisada tanto como uma variável contínua (anos como a unidade de análise) ou como uma variável categórica (1 = 18-29 anos, 2 = 30-39 anos, 3 = 40-49 anos e 4 => 50 anos).

A análise univariada foi realizada considerando-se todos os resultados e potenciais factores de risco. Os testes qui-quadrado ou teste exato de Fisher foram realizados conforme apropriado para a comparação entre as diversas variáveis e calculamos os riscos relativos e seus intervalos de confiança.

Na análise multivariada modelos de riscos proporcionais de Cox, ou Modelos de regressão múltipla de Cox, foram utilizados para obter taxas de

risco instantâneas (HR) para os desfechos estudados (usando alternadamente as três definições de desfecho acima mencionadas) entre os doadores de sangue soropositivos e soronegativos para doença de Chagas, ajustando os modelos para variáveis confundidoras e/ou modificadores de efeito relevantes. O procedimento de seleção "backwards" foi usado para remover uma a uma dos modelos as variáveis que representavam fatores de riscos potenciais. As variáveis removidas foram aquelas com valores de $p \geq 0,2$ no teste de Wald. As demais variáveis foram então sucessivamente removidas com base no seu efeito de confusão e a sua contribuição para os modelos. Variáveis cuja a retirada do modelo provocou mudanças substanciais ($> 10\%$) nos HRs foram mantidas, assim como as variáveis cuja a retirada obtiveram resultados nos testes de variação da razão de verossimilhança significativos (valores de $p < 0,05$).

Para cada desfecho e fator de risco em potencial, a suposição de riscos proporcionais foi testada estatisticamente utilizando-se o método de Resíduos de Schoenfeld (comando `stphtest`) e examinando a natureza paralela das curvas nos gráficos log-log (comando `stphplot`). As curvas foram consideradas como paralelas e os testes não foram estatisticamente significantes, indicando que não houve evidência de violação das suposições. Além disso, para avaliar o risco proporcional incluímos um termo de interação dependente do tempo na análise multivariada de regressão de Cox e testamos a significância estatística desta interação usando o teste de razão de verossimilhança. Consistentemente os resultados indicaram que não houve evidência de violação das suposições.

Utilizamos as Curvas de Kaplan-Meier para analisar as diferenças de sobrevida relacionadas a variável explanatório principal. Estas curvas foram comparadas por meio do teste do logrank . Todas as conclusões adotaram um nível de significância de 5% ($p < 0,05$).

Em relação ao fato de não termos obtido os óbitos da coorte no período de 1996 a 2000, só no período de 2001 a 2009, tínhamos ao menos três alternativas para avaliar e decidir o caminho da análise que seria realizada :

1) Não levar esse fato em consideração seria o mesmo que assumir que não houve diferença de mortalidade entre os dois grupos nesses anos precoces de seguimento (5 anos para doadores de 1996, 4 para os de 1997, 3 anos para os de 1998, 2 anos para os de 1999 e 1 ano para os de 2000);

2) Para podermos levar esse fato em consideração de modo a propormos uma alternativa de correção, poderíamos estudar o comportamento dos óbitos dos dois grupos entre os doadores do ano 2000, para os quais sabemos quantos morreram em cada um dos 5 primeiros anos de seguimento, e assumirmos que o mesmo padrão aconteceu para os doadores de 1996 a 1999.

3) Considerar que ninguém morreu nos anos de 1996 a 2000. Isso faria com que a taxa de mortalidade calculada nos mostrasse o mínimo de óbitos esperados e não assumimos nenhuma morte não conhecida de fato.

Optamos pela última alternativa, pois as duas primeiras fariam com que assumíssemos que houve óbitos nos dois grupos. A terceira alternativa nos

possibilitou não imputar dados faltantes e analisar de fato as mortes nos anos para os quais temos informação.

Considerações Éticas:

Este estudo utilizou registro de duas bases de dados existentes. As variáveis, nome, nome de mãe foram utilizadas apenas para o relacionamento probabilístico. Assim que os pares foram formados e classificados como verdadeiros ou falsos, estes campos foram apagados dos arquivos de associação e os indivíduos foram identificados pelo ID de doador ou pelo Documento de óbito (DO).

Toda a informação dos sujeitos foi mantida em sigilo em caráter confidencial e o computador com as bases de dados teve acesso restrito (Usuário / Senha) e não teve acesso a internet. Os pesquisadores envolvidos neste estudo tomaram todas as providências para garantir a confidencialidade das informações dos sujeitos da pesquisa.

Riscos:

O risco potencial para os sujeitos foi mínimo, uma vez que as informações são provenientes de base de dados já existentes. O único risco teria sido a perda acidental de confidencialidade referente ao status de óbito e causa de morte, e como mencionado anteriormente, tentamos manter a confidencialidade.

Benefícios:

Conhecer a taxa de mortalidade e causas de morte entre os doadores de sangue soropositivos para a doença de Chagas permitirá o aprimoramento de políticas de saúde pública e o melhor aconselhamento dos doadores.

Termo de Consentimento Livre e Esclarecido (TCLE):

Por consenso, a CONEP e o CEP da FMUSP consideraram desnecessário o termo de consentimento livre e esclarecido para estudos que revêem prontuários, dada a dificuldade da obtenção do mesmo e a importância deste tipo de estudo. Assim sendo, não houve necessidade de consentimento do sujeito neste estudo. Este estudo utilizou bancos de dados presente na FPS e SIM. Os identificadores foram retirados dos bancos após o relacionamento probabilístico.

A permissão da utilização do SIM de 2001 – 2009 com nomes disponíveis foi concedida pelo Ministério da Saúde do Brasil. A permissão de uso do banco de dados de doadores da Fundação Pró-Sangue também foi concedida. Ambas as permissões estão disponíveis no apêndice 2.

4. Resultados

4.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico

No arquivo referência, que foi composto da extração de 200 registros do banco de dados da FPS e 196 registros do banco de dados do HC-FMUSP, 100% das variáveis de pareamento, ou seja, nome, nome da mãe e data de nascimento tinham seus campos completos e sem abreviações. O arquivo comparação, SIM 2001 - 2006, continha 4.775.164 registros de indivíduos que morreram aos 18 anos de idade ou mais e, 100% dos campos Nome estavam completos, 93,35% dos campos nome da mãe estavam completos e 99% dos campos data de nascimento estavam completos. No primeiro passo de blocagem 179 pares verdadeiros foram formados e no segundo passo de blocagem este número cresceu para 185 de um total de 196 pares verdadeiros possíveis. Como nenhum novo par verdadeiro foi encontrado no terceiro passo de blocagem, reportamos resultados dos dois primeiros passos apenas. A Tabela 5 mostra a totalidade de pares formados entre os arquivos comparação e referência, dadas as limitações das variáveis de blocagem. Pares com escores altos foram mais encontrados no primeiro passo de blocagem. Se verificarmos todos os escores de pares formados no primeiro passo de blocagem, veremos que apesar de 5 pares terem sido formados com escore 9,8 apenas um deles foi formado pela primeira vez e seu escore mais alto foi 9,8 nos outros quatro pares os registros do arquivo referência já haviam formado par com um registro do arquivo comparação e seus escores foram maiores que 9,8.

Tabela 5 - Sensibilidade, especificidade e valor preditivo positivo de vários escores do processo de relacionamento probabilístico de registros

Escore	Passo1		Passo2		Passos 1 + 2	Passos 1 + 2	Passos 1 + 2	Passos 1 + 2	Passos 1 + 2
	Todos os pares	*Pares pela 1ª vez	Todos os pares	*Pares pela 1ª vez	Pares Verdadeiros	Pares Falsos	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Valor Preditivo Positivo (IC 95%)
17,22	147	147	0	0	147	0	75% (68-80)	100% (98-100)	100% (97-100)
17,01	2	2	0	0	2	0	76% (69-82)	100% (98-100)	100% (98-100)
16,99	1	1	0	0	1	0	76,5% (70-82)	100% (98-100)	100% (98-100)
16,97	2	2	0	0	2	0	77,6% (71-83)	100% (98-100)	100% (98-100)
16,95	1	1	0	0	1	0	78,1% (72-84)	100% (98-100)	100% (98-100)
16,93	2	2	0	0	2	0	79,1% (73-85)	100% (98-100)	100% (98-100)
16,91	1	1	0	0	1	0	79,6% (73-85)	100% (98-100)	100% (98-100)
16,89	1	1	0	0	1	0	80,1% (74-85)	100% (98-100)	100% (98-100)
16,87	3	3	1	1	4	0	82,1% (76-87)	100% (98-100)	100% (98-100)
16,78	1	1	0	0	1	0	82,7% (77-88)	100% (98-100)	100% (98-100)
16,7	4	4	0	0	4	0	84,7% (79-89)	100% (98-100)	100% (98-100)
16,6	2	2	0	0	2	0	85,7% (80-90)	100% (98-100)	100% (98-100)
9,8	5	1	0	0	1	0	86,2% (81-91)	100% (98-100)	100% (98-100)
9,5	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
9,03	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
8,98	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
8,83	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
8,55	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
8,26	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
7,98	1	0	0	0	0	0	86,2% (81-91)	100% (98-100)	100% (98-100)
7,06	12	11	5	5	16	0	94,4% (90-97)	100% (98-100)	100% (98-100)
6,54	2	0	0	0	0	0	94,4% (90-97)	100% (98-100)	100% (98-100)
6,23	2	0	0	0	0	0	94,4% (90-97)	100% (98-100)	100% (98-100)
6,02	34	6	0	0	0	6	94,4% (90-97)	97% (94-99)	97% (93-99)
5,86	1	0	0	0	0	0	94,4% (90-97)	97% (94-99)	97% (93-99)
5,78	1	0	0	0	0	0	94,4% (90-97)	97% (94-99)	97% (93-99)
5,74	1	0	0	0	0	0	94,4% (90-97)	97% (94-99)	97% (93-99)
5,71	14	1	0	0	0	1	94,4% (90-97)	96,5% (93-99)	96% (93-98)
5,7	1	1	0	0	0	1	94,4% (90-97)	96% (92-98)	96% (92-98)
5,65	1	1	0	0	0	1	94,4% (90-97)	95,5% (92-98)	95% (91-98)
5,64 to -10,51	709 284	106	1 448 307	85	0	191	94,4% (90-97)	95,5 to 0%	48 to 0%
TOTAL	709 532	294	1 448 313	91	185	200			

A maioria dos pares verdadeiros obtiveram o escore máximo 17,22. Nos escores entre 17,01 a 9,8 foram encontrados no mínimo um e no máximo quatro pares verdadeiros em cada escore. Os primeiros seis pares falsos foram encontrados no escore 6,02. No escore 9,8 a sensibilidade foi de 86,2% (IC 95%: 80,5- 90,7) e no escore 7,06 foi de 94,4% (IC 95%: 90,2- 97,2), o valor máximo obtido. A especificidade foi de 100% (IC 95%: 98,2-100,0) até o escore 6,23. Estes resultados apontam para uma escolha natural de uma nota

de corte capaz de discriminar pares falsos de pares verdadeiros, pois os ultimos pares verdadeiros foram encontrados no escores 7,06 e os primeiros pares falsos foram encontrados no escore 6,02. Utilizando uma nota de corte dentro deste intervalo, o VPP foi de 100%(IC 95%: 98,2-100,0). A revisão manual revelou porque 16 pares verdadeiros foram encontrados pela primeira vez no escore 7,06. O problema estava na variável nome. Em um dos registros destes pares, um elemento do nome estava faltando, ex: em um dos registros havia o nome de solteira (Maria Antonia Lima) e no outro o nome de casada (Maria Antonia Lima Castro). Estes nomes são fictícios.

A acurácia deste método de relacionamento probabilístico foi bastante alta, com sensibilidade de 94% (IC 95%, 90% - 97%), especificidade de do método de 100% (IC 95%, 98% - 100%) e a nota e corte entre 7,06 e 6,23.

4.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo

O relacionamento probabilístico de registros foi realizado entre os registros do arquivo referência contendo informações extraídas do banco de dados da FPS de todos os doadores soropositivos para a doença de Chagas entre 1996 e 2000 e doadores soronegativos para todos os testes de triagem e o arquivo comparação, SIM versão 01/11/2010, contendo dados dos brasileiros que vieram a óbito entre 2001 e 2009.

Análise dos Dados:

As características demográficas dos doadores de sangue da FPS estão descritas na Tabela 6 abaixo. Dos 8526 doadores selecionados para o

relacionamento 6022 (70,6%) eram do sexo masculino, e representam 1976 (69,5%) dos doadores soropositivos. Não verificamos diferença estatística significativa de sexo entre doadores soropositivos e doadores soronegativos neste estudo. Como consequência do desenho do estudo, o arquivo referência tem duas vezes mais doadores soronegativo que soropositivos e sua distribuição quanto a ano de doação é equivalente, como mostra a Tabela 6 abaixo.

Optamos por dividir a população do estudo em quatro faixas etárias, 1 = 18-29 anos, 2 = 30-39 anos, 3 = 40-49 anos e 4 => 50 anos. Por se tratar de uma população de doadores, verificamos que nossa população é formada por 90,0% de doadores com idade até 49 anos. Devido à diminuição do número de casos da doença de Chagas no Brasil, e devido também a diminuição de doadores de primeira vez na FPS entre 1995 e 2001 (Gonzalez et al. 2003), os casos soropositivos e, por consequência os soronegativos, são em menor número nos anos de doação de 1999 (14,3%) e 2000 (13,1%) se comparados com 1996 (27,1%), 1997 (25,3%), e 1998 (20,3%).

Tabela 6 - Distribuição dos doadores de sangue da FPS selecionados para o processo de relacionamento probabilístico por sexo, faixa etária e ano de doação.

	Soropositivos		Soronegativos		TOTAL	
		%		%	No	%
Todos Sujeitos		2 842		5 684	8 526	
Sexo						
Masculino	1 976	69,5	4 046	71,2	6 022	70,6
Feminino	866	30,5	1 638	28,8	2 504	29,4
Faixa Etária (anos)						
18 - 29	646	22,7	2 886	50,8	3 532	41,4
30 - 39	795	28,0	1 629	28,7	2 424	28,4
40 -49	893	31,4	833	14,7	1 726	20,2
>50	508	17,9	336	5,9	844	9,9
Ano de doação						
1996	771	27,1	1 539	27,1	2 310	27,1
1997	717	25,2	1 436	25,3	2 153	25,3
1998	578	20,3	1 157	20,4	1 735	20,3
1999	406	14,3	809	14,2	1 215	14,3
2000	370	13,0	743	13,1	1 113	13,1

Soropositivos- Doadores Soropositivos para Doença de Chagas nos testes de triagem da
Soronegativos - Doadores soronegativos para todos os testes de triagem da FPS

Desfecho

O processo de relacionamento utilizado para identificar os óbitos no período estudado encontrou 250 pares verdadeiros no primeiro passo de bloqueio, 10 pares verdadeiros no segundo passo de bloqueio e 2 pares verdadeiros no terceiro passo de bloqueio, totalizando 262 pares verdadeiros, ou seja, 262 óbitos. Pares formados com escores abaixo de 6,0 foram desprezados e não passaram por revisão manual. Destes 262 pares verdadeiros, 159 pares eram soropositivos para doença de Chagas e 103 eram soronegativos. A maioria dos pares verdadeiros, 172 (65,6%), obtiveram o escore máximo 17,22 e foram encontrados no primeiro passo de bloqueio, ou seja, houve concordância total dos campos nome, nome da mãe e data de nascimento entre os registros do arquivo referência e comparação. A Tabela 7 mostra a distribuição dos pares verdadeiros e seus escores por doadores soropositivos e soronegativos para a doença de Chagas

Tabela 7 - Distribuição de doadores soropositivos e soronegativos por escore

Escore	Bloqueio												TOTAL				
	Passo 1				Passo 2				Passo 3				Soro+		Soro-		Total
	Soro+	%	Soro-	%	Soro+	%	Soro-	%	Soro+	%	Soro-	%	Soro+	%	Soro-	%	
17,22	104	65,4	68	66,0	0	0	0	0	0	0	0	0	104	65,4	68	66,0	172
17,00 - 16,17	29	18,2	14	13,6	1	0,6	0	0	1	0,6	1	1,0	31	19,5	15	14,6	46
15,86 - 15,24	4	2,5	3	2,9	0	0	0	0	0	0	0	0	4	2,5	3	2,9	7
10,69 - 9,11	9	5,7	7	6,8	0	0	2	1,9	0	0	0	0	9	5,7	9	8,7	18
7,06 - 6,52	8	5,0	4	3,9	3	1,9	4	3,9	0	0	0	0	11	6,9	8	7,8	19
TOTAL	154	96,9	96	93,2	4	2,5	6	5,8	1	0,6	1	1,0	159	100	103	100	262

Soro + - Doadores soropositivos para doença de Chagas nos testes de triagem da FPS, Soro - - Doadores soronegativos para todos os testes

Bloqueio - **Passo 1** - PBLOCO + UBLOCO + SEXO, **Passo 2** - PBLOCO + SEXO, **Passo 3** - UBLOCO + SEXO

Com base no resultado do relacionamento probabilístico, encontramos 262 doadores que morreram no período de 2001 a 2009, 159 (60,7%) eram doadores soropositivos e 103 (39,3%) eram soronegativos. 214 (81,7%) dos doadores que faleceram eram do sexo masculino, dos quais 131 (61,2%) eram soropositivos e 83 (38,8%) eram soronegativos. Entre os doadores soropositivos ocorreram mais óbitos na faixa etária de 40 a 49 anos de idade e

50 anos ou mais velhos, 102 óbitos representando 38,9% do total de óbitos encontrados. Entre os doadores soronegativos os óbitos ocorreram com maior frequência nas faixas etárias de 18 a 29 anos e 40 a 49 anos, 68 óbitos representando 26% do total de óbitos encontrados (ver Tabela 10). Como optamos por selecionar 2 doadores soronegativos para cada doador soropositivo, a Tabela 8 descreve a proporção de óbitos em relação ao número de indivíduos em cada categoria, ou seja, dos 2842 doadores soropositivos para doença de Chagas 5,6% faleceram e entre o total de doadores soronegativos para a doença de Chagas (1,8%).

Tabela 8 - Distribuicao de pares verdadeiros por doadores soropositivos e soronegativos para doença de Chagas

	Todos os sujeitos N= 8 526 n (%)	Soropositivos N=2 842 n (%)	Soronegativos N=5 684 n (%)	valor - p
Sexo				0,114
Masculino	6 022 (70,6)	1 976 (69,5)	4 046 (71,2)	
Feminino	2 504 (29,4)	866 (30,5)	1 638 (28,8)	
Faixa Etária (anos)				<0,001
18-29	3 523 (41,4)	646 (22,7)	2 886 (50,8)	
30-39	2 424 (28,4)	795 (28,0)	1 629 (28,7)	
40-49	1 726 (20,3)	893 (31,4)	833 (14,6)	
>50	844 (9,9)	508 (17,9)	336 (5,9)	
Ano de Doação				1.000
1996	2 310 (27,1)	771 (27,1)	1 539 (27,1)	
1997	2 153 (25,3)	717 (25,2)	1 436 (25,3)	
1998	1 735 (20,3)	578 (20,4)	1 157 (20,3)	
1999	1 215 (14,2)	406 (14,3)	809 (14,2)	
2000	1 113 (13,1)	370 (13,0)	743 (13,1)	
Mortalidade	262 (3,1)	159 (5,6)	103 (1,8)	<0,001
Soropositivos - Doadores Soropositivos para Doença de Chagas nos testes de triagem da FPS, Soronegativos - Doadores soronegativos para todos os testes de triagem da FPS				

Classificação de óbitos conforme grupos definidos

Classificamos os 262 registros dos doadores que encontramos no SIM nos grupos definidos para a análise descritiva, mencionados anteriormente (ver

Figura 1). Inicialmente, apenas levando em consideração o que consta no SIM como causa básica de morte, a distribuição dos óbitos encontrados nos diversos capítulos do CID-10 está descrita na Tabela 9. Os óbitos distribuídos por subcategorias, ou seja, códigos de quatro caracteres, estão descritos no apêndice 3, aonde a Tabela 9 se encontra mais detalhada.

Não encontramos nenhum registro para os quais os códigos U049-U99 ou Z000 a Z999, dos Capítulos XXII e XXI do CID-10 respectivamente, foram designados como causa básica de morte.

Tabela 9 - Causas básicas de morte no SIM dos pares verdadeiros por Capítulos

Descrição Causa Básica	Soro +	Soro -	TOTAL
Capítulo I - Algumas doenças infecciosas e parasitárias	32	3	35
Capítulo II - Neoplasias [tumores]	22	17	39
Capítulo IV - Doenças endócrinas, nutricionais e metabólicas	0	2	2
Capítulo V - Transtornos mentais e comportamentais	1	0	1
Capítulo VI - Doenças do sistema nervoso	0	1	1
Capítulo IX - Doenças do aparelho circulatório	58	20	78
Capítulo X - Doenças do aparelho respiratório	9	2	11
Capítulo XI - Doenças do aparelho digestivo	8	8	16
Capítulo XIV - Doenças do aparelho geniturinário	2	0	2
Capítulo XVII - Malformações congênitas, deformidades e anomalias cromossômicas	0	1	1
Capítulo XVIII - Sint., sinais e ach anormais de ex clínicos e de lab, não classif em outra parte	9	8	17
Capítulo XX - Causas externas de morbidade e de mortalidade	18	41	59
	159	103	262

Soro+: Doadores Soropositivos para Doença de Chagas nos testes de triagem da FPS,
Soro -: Doadores Soronegativos para Doença de Chagas nos testes de triagem da FPS

Entre os 159 óbitos de doadores soropositivos para doença de Chagas 26 (16,4%) foram classificados no Grupo 1 - doença de Chagas, pois a causa básica de óbito foi codificada como B57.0 - B57.5 no SIM. Entre os 103 óbitos

de doadores soronegativos para todos os testes de triagem da FPS, o código B57.0 - B57.5 não foi encontrado em nenhuma das linhas do SIM.

Apesar da população do estudo ser composta na sua maioria por homens, não encontramos diferença estatisticamente significativa para a variável sexo comparada a sorologia do doador ($p=0,114$). No entanto verificamos diferença significativa para a variável faixa etária quando comparada a sorologia do doador ($p<0,001$).

Em uma análise univariada, os riscos relativos e seus intervalos de confiança foram calculados para os óbitos em cada um dos grupos por faixa etária e estão apresentados na Tabela 10 abaixo.

Tabela 10 - Número de óbitos e Risco Relativo por sorologia, faixa etária e grupo.

Causa Básica de Morte		Soro +	%	Soro -	%	Valor de P	Risco Relativo (IC - 95%)
Óbitos Total	18 - 29	22	0,8	35	0,6	<0,001	2,8 (1,7 - 4,6)
	30 - 39	35	1,2	20	0,4	<0,001	3,6 (2,1 - 6,2)
	40 - 49	52	1,8	33	0,6	<0,074	1,5 (1,0 - 2,3)
	>=50	50	1,8	15	0,3	<0,004	2,2 (1,3 - 3,9)
	Todos	159	5,6	103	1,8	<0,001	3,1 (2,4 - 3,9)
Grupo 1 Doença de Chagas	18 - 29	5	0,2	0	0	<0,001	-
	30 - 39	10	0,4	0	0	<0,001	-
	40 - 49	7	0,2	0	0	<0,016	-
	>=50	4	0,1	0	0	<0,155	-
	Todos	26	0,9	0	0	<0,001	-
Grupo 2 Neoplasias	18 - 29	0	0	1	0	1,000	0 -
	30 - 39	0	0	3	0,1	<0,555	0 -
	40 - 49	8	0,3	7	0,1	1,000	1,1 (0,4 - 2,9)
	>=50	14	0,5	6	0,1	<0,489	1,5 (0,6 - 4,0)
	Todos	22	0,8	17	0,3	<0,003	2,6 (1,4 - 4,9)
Grupo 3 Doenças cardíacas	18 - 29	5	0,2	2	0	<0,003	11,2 (2,2 - 57,4)
	30 - 39	6	0,2	4	0,1	<0,089	3,1 (0,9 - 10,9)
	40 - 49	18	0,6	5	0,1	<0,011	3,4 (1,3 - 9,0)
	>=50	18	0,6	3	0,1	<0,022	4,0 (1,2 - 13,4)
	Todos	47	1,7	14	0,2	<0,001	6,7 (3,7 - 12,2)
Grupo 4 Outras doenças	18 - 29	3	0,1	0	0	<0,006	-
	30 - 39	12	0,4	6	0,1	<0,004	4,1 (1,5 - 10,9)
	40 - 49	12	0,4	12	0,2	1,000	0,9 (0,4 - 2,1)
	>=50	10	0,4	5	0,1	<0,792	1,3 (0,5 - 3,8)
	Todos	37	1,3	23	0,4	<0,001	3,2 (1,9 - 5,4)
Grupo 5 Traumatismos	18 - 29	8	0,3	27	0,5	<0,482	1,3 (0,6 - 2,9)
	30 - 39	5	0,2	4	0,1	<0,164	2,6 (0,7 - 9,5)
	40 - 49	2	0,1	8	0,1	<0,057	0,2 (0,5 - 1,1)
	>=50	1	0	0	0	1,000	-
	Todos	16	0,6	39	0,7	<0,503	0,8 (0,5 - 1,5)
Grupo 6 Mal definidas ou inespecíficas	18 - 29	1	0	5	0,1	1,000	0,9 (0,1 - 7,6)
	30 - 39	2	0,1	3	0,1	<0,665	1,4 (0,2 - 8,2)
	40 - 49	5	0,2	1	0	<0,220	4,7 (0,5 - 39,8)
	>=50	3	0,1	1	0	1,000	2,0 (0,2 - 19,0)
	Todos	11	0,4	10	0,2	<0,064	2,2 (0,9 - 5,2)

Soro+ = Soropositivo para doença de Chagas nos testes de triagem da FPS

Soro- = Soronegativo para doença de Chagas nos testes de triagem da FPS

Para o total de óbitos podemos observar que existe uma diferença estatisticamente significativa entre doadores soropositivos e soronegativos para a doença de Chagas, com Risco Relativo (RR) 3,1 (IC 95%= 2,4 - 3,9). Nas faixas etárias de 18 a 29 anos, 30 a 39 anos e maiores de 50 anos, com Risco Relativo (RR) de 2,8 (IC 95% = 1,7 - 4,6), 3,6 (IC 95% = 2,1 - 6,2) e 2,2 (IC 95%= 1,3 - 3,9) respectivamente. Classificando os óbitos pelos 6 grupos

previamente definidos, observamos que para os óbitos classificados no Grupo 3 - Doenças Cardíacas o RR entre soropositivos e soronegativos é de 6,7 (IC 95%=3,7-12,2). A diferença de óbitos entre doadores soropositivos e soronegativos nas faixas etárias de 18 a 29 anos, 40 a 49 anos e maiores que 50 anos de idade neste grupo é estatisticamente significativa com RR de 11,2 (IC 95%= 2,2 - 57,4), 3,4 (IC 95%= 1,3 - 9,0) e 4,0 (IC 95%= 1,2 - 13,4) respectivamente. No Grupo 4 - Outras doenças, o RR entre doadores soropositivos e soronegativos é de 3,2 (IC 95%= 1,9 - 5,4), no entanto observamos diferença significativa apenas para a faixa etária de 30 a 39 anos, RR 4,1 (IC 95%= 1,5 - 10,9). Não observamos diferenças estatisticamente significantes entre os óbitos de soronegativos e soropositivos nos grupos 5 - Traumatismos e 6 Causas mal definidas ou inespecíficas. Verificamos uma diferença significativa no total de óbitos de soropositivos e soronegativos no grupo 2 - Neoplasias, RR 2,6 (IC 95% =1,4 - 4,9), que quando testada por faixa etária não se confirmou.

Em todos os grupos, exceto no grupo 5 - Traumatismos, o número total de óbitos entre soropositivos foi maior que em soronegativos.

As causas associadas mais frequentes encontradas no Grupo 1 - Doença de Chagas, entre os doadores soropositivos foram I42 - Cardiomiopatias, para 10 (38,5%) dos óbitos, R57 - Choque não classificado em outra parte, para 9 (34,6%) dos óbitos e I50 - insuficiência cardíaca, para 6 (24,6%) dos óbitos entre soropositivos. Entre os soropositivos que foram classificados no Grupo 3 - Doenças Cardíacas, apenas um doador soropositivo tem menção a doença de Chagas como comorbidade, ou seja, o CID 10 B57.2

foi mencionado na parte II do atestado. Todos os outros 46 doadores soropositivos deste grupo não tem menção alguma a doença de Chagas no SIM. As causas básicas de morte e causas associadas a possíveis óbitos por doença de Chagas para óbitos de doadores soropositivos classificados no Grupo 3 - Doenças Cardíacas estão descritas na Tabela 11 a seguir:

Tabela 11 - Causas básicas de morte do Grupo 3 - Doenças Cardíacas - para doadores soropositivos para doença de Chagas.

Grupo 3 - Doenças Cardíacas - Soropositivos				Causas Associadas - Possíveis Óbitos por Doença de Chagas								
Causa Básica de Morte	Descrição	N	%	I42.0	I42.2	I47.0	I47.2	I49.0	I50.0	I50.1	I50.9	I51.7
I21.9	Infarto agudo do miocárdio não especificado	23	48,94	2	0	0	0	0	0	0	0	1
I42.0	Cardiomiopatia dilatada	5	10,64	5	0	0	0	0	0	0	0	0
I11.9	Doença cardíaca hipertensiva sem insuficiência cardíaca (congestiva)	3	6,38	0	0	0	0	0	0	0	0	0
I25.9	Doença isquêmica crônica do coração não especificada	3	6,38	1	0	0	0	0	0	0	1	1
I51.7	Cardiomegalia	3	6,38	0	0	0	0	0	0	0	1	2
I25.1	Doença aterosclerótica do coração	2	4,26	0	0	0	0	0	0	0	1	0
I10	Hipertensão essencial (primária)	1	2,13	0	0	0	0	0	0	0	0	0
I11.0	Doença cardíaca hipertensiva com insuficiência cardíaca (congestiva)	1	2,13	0	0	0	0	0	1	0	0	0
I21.0	Infarto agudo transmural da parede anterior do miocárdio	1	2,13	0	0	0	0	0	0	0	0	0
I24.8	Outras formas de doença isquêmica aguda do coração	1	2,13	1	0	0	0	0	0	0	0	0
I42.9	Cardiomiopatia não especificada	1	2,13	0	0	0	0	0	0	0	0	0
I44.3	Outras formas de bloqueio atrioventricular e as não especificadas	1	2,13	0	0	0	0	0	0	0	0	0
I50.0	Insuficiência cardíaca congestiva	1	2,13	0	0	0	0	0	1	0	0	0
I51.4	Miocardite não especificada	1	2,13	0	0	0	0	0	0	0	1	0
Total		47	100	9	0	0	0	0	2	0	4	4

Dos 47 óbitos de doadores soropositivos para doença de Chagas classificados no Grupo 3, observamos que 31 (66,0%) não tinham como causa associada de óbito ao menos uma causa que consideramos como causa de possível óbito por doença de Chagas. No entanto, 16 (34%) tinham ao menos uma causa de possível óbito por doença de Chagas entre as causas associadas, um deles com 3 causas associadas e outro com 2 causas associadas.

A Tabela 12 descreve as causas associadas mais frequentes para os Grupos 1, 3, 5 e 6.

Tabela 12 - Frequência de causas associadas por Grupos

CID 10	Descrição	Óbitos
Soropositivos Grupo 1 - Doença de Chagas		26
I42	Cardiomiopatias	10
R57	Choque não classificado em outra parte	9
I50	Insuficiência cardíaca	6
Soronegativos Grupo 1 - Doença de Chagas		0
-	-	-
-	-	-
-	-	-
Soropositivos Grupo 3 - Doenças Cardíacas		47
I21	Infarto agudo do miocárdio	25
I25	Doença isquêmica crônica do coração	19
J81	Edema pulmonar, não especificado de outra forma	13
Soronegativos Grupo 3 - Doenças Cardíacas		14
I21	Infarto agudo do miocárdio	11
R57	Choque não classificado em outra parte	3
I50	Insuficiência cardíaca	4
Soropositivos Grupo 5 - Traumatismos		16
S06	Traumatismo intracraniano	6
T79	Algumas complicações precoces dos traumatismos não classificadas em outra parte	5
X95	Agressão por meio de disparo de outra arma de fogo ou de arma não especificada	3
Soronegativos Grupo 5 - Traumatismos		39
S06	Traumatismo intracraniano	15
T79	Algumas complicações precoces dos traumatismos não classificadas em outra parte	15
X95	Agressão por meio de disparo de outra arma de fogo ou de arma não especificada	12
Soropositivos Grupo 6 - Mal definidas ou inespecíficas		11
R99	Outras causas mal definidas e as não especificadas de mortalidade	6
R98	Morte sem assistência	3
Y34	Fatos ou eventos não especificados e intenção não determinada	2
Soronegativos Grupo 6 - Mal definidas ou inespecíficas		10
R99	Outras causas mal definidas e as não especificadas de mortalidade	6
Y34	Fatos ou eventos não especificados e intenção não determinada	2
R98	Morte sem assistência	1

A Tabela 13 contém a distribuição de óbitos por ano de doação.

Tabela 13 - Número de óbitos por ano de doação

Tabela 13a: Número de óbitos por ano de Doação para doadores Soropositivos						
Ano de Doação / anos de seguimento	1996	1997	1998	1999	2000	Total
1	-	-	-	-	0	0
2	-	-	-	0	0	0
3	-	-	0	0	0	0
4	-	0	1	0	1	2
5	2	1	3	6	2	14
6	7	3	7	4	3	24
7	4	4	4	1	4	17
8	3	3	4	2	2	14
9	5	5	5	3	2	20
10	8	10	6	1	-	25
11	9	7	5	-	-	21
12	11	4	-	-	-	15
13	7	-	-	-	-	7
Total Óbitos	56	37	35	17	14	159
Total Doadores Soro+	771	717	578	406	370	2 842

Tabela 13b: Número de óbitos por ano de Doação para doadores Soronegativos

Ano de Doação /anos de seguimento	1996	1997	1998	1999	2000	Total
1	-	-	-	-	0	0
2	-	-	-	0	1	1
3	-	-	0	0	1	1
4	-	0	1	3	1	5
5	1	2	1	1	1	6
6	3	5	3	2	1	14
7	4	2	2	2	1	11
8	2	3	2	1	1	9
9	5	1	2	2	4	14
10	5	1	5	3	-	14
11	5	4	3	-	-	12
12	4	7	-	-	-	11
13	5	-	-	-	-	5
Total Óbitos	34	25	19	14	11	103
Total Doadores Soro-	1 539	1 436	1 157	809	743	5 684

As probabilidades de sobrevivência dos doadores segundo sorologia e faixa etária estão representadas nas Figuras 3, 4, 5 e 6 abaixo. As curvas de sobrevida entre doadores soropositivos e soronegativos quando comparadas são diferentes e estatisticamente significantes ($p < 0,001$).

Figura 3 - Probabilidade de sobrevivência segundo sorologia

Figura 4 - Probabilidade de sobrevivência segundo sorologia por faixa etária

Por causa da diferença em nossa população quanto a idade, ajustamos também as curvas para que a sobrevivência dos doadores fosse melhor representada. Na faixa etária de doadores entre 18 a 29 anos existe diferença estatisticamente significativa entre soropositivos e soronegativos ($p < 0,001$). Na faixa etária de doadores entre 30 a 39 anos também existe diferença estatisticamente significativa entre soropositivos e soronegativos ($p < 0,001$). Entre 40 e 49 anos não houve diferença estatística ($p = 0,0967$). E finalmente na faixa etária de doadores maiores que 50 anos de idade a diferença é estatisticamente significativa ($p < 0,0044$). Quando comparamos apenas as curvas dos doadores soropositivos nas quatro faixas etárias também encontramos diferença estatisticamente significativa ($p < 0,001$).

As taxas de risco instantâneas (HR) para os desfechos A) Óbitos confirmados por Chagas, B) Óbitos confirmados e possíveis por Chagas C) Óbitos por todas as causas, entre os doadores de sangue soropositivos e soronegativos para doença de Chagas, foram calculadas utilizando-se os modelos de riscos proporcionais de Cox, ou Modelos de regressão múltipla de Cox. Doadores soropositivos tem HR 2,3 (IC 95%= 1,8 - 3,0 valor p <0,001) vezes maior que soronegativos quando classificados no grupo C) Óbitos por todas as causas e HR 17,9 (IC 95%= 6,3 - 50,8 valor p <0,001) vezes maior que soronegativos quando classificados no grupo B) Óbitos confirmados e possíveis por Chagas. As taxas brutas diminuíram quando ajustadas por sexo e faixa etária. Estes resultados estão descritos na Tabela 14 abaixo.

Tabela 14 - Números e taxas de mortalidade, brutas e ajustadas, para diferentes desfechos.

Definição dos Grupos (Modelos)	Número de óbitos		Taxa de mortalidade 100.000 PA		HR Bruta (IC - 95%)	HR Ajustada*** (IC - 95%)
	Soro+* N=2842	Soro-* N=5684	Soro+ PA=33295,042**	Soro- PA=67172,69**	valor - p	valor - p
A. Óbitos confirmados por Chagas Códigos CID10 B57 (Causas básicas ou associadas de óbito)	26	0	78,1	0	-	-
B. Óbitos confirmados e possíveis por Chagas Códigos CID10 B57 I470 I472 I490 I500 I501 I517 I509 (Causas básicas ou associadas de óbito)	45	4	135,2	6	22,8 (8,2-63,4) <0,001	17,9 (6,3-50,8) <0,001
C. Óbitos por todas as causas	159	103	477,4	153,4	3,1 (2,4-4,0) <0,001	2,3 (1,8-3,0) <0,001

* Soro+: Doadores soropositivos para a Doença de Chagas, Soro-: Doadores soronegativos para todos os testes de triagem da FPS

** PA - pessoas-ano

*** Ajustes:

Modelo B foi ajustado por faixa etária e sexo : Ajuste HR para faixa etária = cada incremento na faixa etária foi associada com um HR de 1,4 (95%IC 1,1-1,9), p=0,016

Modelo C foi ajustado por faixa etária e sexo : Ajuste HR para sexo = sexo masculino foi associado com um HR de 2,0 (95%IC 1,5-2,8), p=0,001

Ajuste HR para faixa etária = cada incremento na faixa etária foi associada com um HR de 1,6 (95%IC 1,4-1,8), p=0,001

Realizamos também a análise multivariada sem considerar os 128 doadores soropositivos para a doença de Chagas co-infectados com outras doenças (ver Tabela 3). Doadores soropositivos não co-infectados tem HR 2,2 (IC 95%= 1,7 - 2,9, valor de p<0,001) vezes mais que doadores soronegativos quando classificados no grupo C) Óbitos por todas as causas e HR 16,1 (IC

95%= 5,6 - 45,9 valor $p < 0,001$) vezes maior que soronegativos quando classificados no grupo B) Óbitos confirmados e possíveis por Chagas.

5. Discussão

5.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico

O método de relacionamento probabilístico de registros utilizado neste estudo visou auxiliar a identificação de doadores de sangue do banco de dados da FPS no SIM. Para determinar a acurácia deste método, utilizamos um arquivo que continha registros para os quais o status vital era conhecido. Assim sendo, fomos capazes de saber quais dos pares formados eram realmente pares verdadeiros. Esta estratégia de utilizar um status vital conhecido para determinar a acurácia de um método de relacionamento probabilístico já foi usado em outros estudos, incluindo estudos brasileiros (Pacheco et al. 2008; Fonseca, Coeli, de Fatima de Araujo Lucena et al. 2010).

A acurácia de nosso método foi bastante alta, com sensibilidade de 94% e especificidade de 100%, utilizando uma nota de corte entre 7,06 e 6,23 como um ponto de corte. Nossos resultados são levemente melhores dos que os de um estudo brasileiro que utilizou o RECLINK-III e as mesmas variáveis de pareamento para relacionar registros do banco de dados de vigilância da AIDS com o SIM, onde encontraram uma sensibilidade de 87,6% e especificidade de 99,6% (Fonseca, Coeli, de Fatima de Araujo Lucena et al. 2010). No entanto, apesar das diferenças dos nossos arquivos de referência e do banco de dados de vigilância que utilizaram, temos que levar em conta que enquanto nosso arquivo de referência não continha nenhum valor incompleto

ou faltante nas variáveis de pareamento, o arquivo deles era bem maior e muito provavelmente continha muitos registros com este tipo de erro. A utilização de grandes bancos de dados obviamente faz com que a revisão manual seja um desafio real. Nossos resultados são comparáveis àqueles de outro estudo brasileiro que utilizou um processo de relacionamento determinístico para relacionar registros de duas coortes de estudos de pacientes infectados por HIV com o SIM, onde encontraram uma sensibilidade de 96,5% e especificidade de 99,6% (Pacheco, Saraceni, Tuboi et al. 2008).

Apesar da sensibilidade do método ser alta, o relacionamento probabilístico não encontrou 11 indivíduos do arquivo de referência, sabidamente mortos, no SIM. Existem ao menos quatro explicações possíveis para este fato: 1) o óbito destes indivíduos não constarem no SIM, 2) o óbito destes indivíduos foram registrados no SIM, mas os registros foram descartados antes do relacionamento probabilístico por terem dados faltantes nas variáveis nome, ou terem sido erroneamente registrados no SIM como indivíduos com idade menor que 18 anos, ou ainda por terem dados faltantes na identificação do óbito (DO), 3) o óbito destes indivíduos terem sido registrados no SIM, mas as informações disponíveis nas três variáveis de pareamento conterem diferenças a tal ponto que o par dos registros formados pelo arquivo referência e o arquivo comparação ter sido classificado com a última nota possível (-10,52), e neste caso não ter sido realizada a revisão manual, 4) os pares entre os registros do arquivo referência e o arquivo comparação não terem sido formados, porque os registros foram bloqueados

utilizando-se a variável sexo, presente nos três passos de blocagem que realizamos.

A primeira explicação sugere que exista sub notificação no SIM. SIM já passou por aperfeiçoamentos significantes na cobertura e qualidade de dados na última década (Jorge et al. 2010), mas na maioria estas melhorias foram necessárias em áreas rurais das regiões Norte e Nordeste do país, e para indivíduos que morreram fora de um hospital e em idades extremas (Franca et al. 2008). Este não é o caso dos 196 indivíduos do nosso arquivo de referência que sabemos que morreram durante sua internação no maior hospital universitário público da cidade de São Paulo, aonde certamente houve uma declaração de óbito. Outra possibilidade é que estes óbitos ainda não tivessem sido registrados no SIM por atraso nas notificações. Também achamos improvável, pois os óbitos ocorreram entre 2002 e 2005 e o banco de dados do SIM foi extraído em abril de 2007.

A segunda e a terceira explicação se refere a questões de qualidade de dados no SIM, o que pode nos impedir de alcançar uma sensibilidade mais alta. A sensibilidade de um método de relacionamento probabilístico depende muito da qualidade dos dados dos bancos utilizados. Acreditamos que os 11 indivíduos que não encontramos estavam entre os 16,7% dos registros do SIM que foram descartados antes do relacionamento probabilístico devido a falta de dados na variável nome, ou que tenham sido erroneamente classificados como indivíduos com idade inferior a 18 anos. Por outro lado, considerando-se a natureza do relacionamento probabilístico de dados e o fato que a última nota representa pares com total discordância nas três variáveis de pareamento, é

muito improvável que algum par verdadeiro destes 11 indivíduos fosse encontrado na nota -10,52, para a qual os pares não foram revistos manualmente.

Sexo foi escolhida como uma variável de blocagem e foi utilizada conjuntamente com as outras, pois acreditamos que poucos registros estariam mal classificados quanto a sexo ou não teriam nenhum valor neste campo. Entretanto estes erros podem ter ocorrido, mas mesmo assim, é pouco provável que esta possibilidade por si só explique não termos encontrado os registros dos 11 indivíduos. De qualquer forma, em relacionamentos probabilísticos futuros, evitaremos ter a mesma variável participando de todos nossos passos de blocagem.

Com apenas 396 indivíduos provenientes de dois bancos de dados (banco de dados da FPS e do HC-FMUSP), nosso arquivo referência era relativamente pequeno e portanto não muito heterogeneo. Se repetíssemos nosso método utilizando a banco de dados inteiro da FPS ou uma amostra maior do banco de dados da FPS, o maior número de registros iria obviamente aumentar a variabilidade dos valores presentes nas variáveis de pareamento, o que aumentaria o número de pares falso positivos, e possivelmente aumentaria as notas destes pares (Christen 2007; Fonseca, Coeli, de Fatima de Araujo Lucena et al. 2010). Como isto poderia diminuir a especificidade observada para um dado valor de nota, talvez a escolha natural de uma nota de corte, como a que temos no estudo presente, não teria sido apresentada. A consequência imediata seria a necessidade de averiguar se a mesma nota de corte utilizada neste estudo poderia ser utilizada em outro. Isto se daria

revisando manualmente os pares na tal chamada "zona cinza". É importante notar que quando estamos relacionando dados de um ou mais bancos com o banco de mortalidade, estamos lidando com uma situação na qual esperamos encontrar apenas um par verdadeiro por indivíduo, a menos que duplicações existam em ambos os bancos. Hoje o RECLINK remove de ambos os bancos de dados os registros que formam um par considerado verdadeiro em um passo de blocagem e no próximo passo de blocagem estes registros não formarão pares. Para o aperfeiçoamento e tornar mais rápida a revisão manual em cada passo de blocagem, sugerimos que o RECLINK removesse automaticamente da lista de pares formados que serão submetidos a revisão manual, todos os pares com notas mais baixas que contenham um dos dois registros que tenham atingido uma nota mais alta. Esta estratégia diminuiria significativamente o trabalho da revisão manual.

Outra consequência se realizássemos o relacionamento probabilístico de dados utilizando o banco de dados inteiro da FPS ou uma amostra mais representativa deste banco, que provavelmente teria uma razão de óbitos diferente, seria naturalmente a alteração do valor preditivo para um ponto de corte selecionado (Hulley 2008). Mais importante, tais mudanças na acurácia também impactariam na forma como os estudos longitudinais que dependem dos resultados de relacionamento probabilístico de dados seriam interpretados. Como mencionado em outros estudos semelhantes (Fonseca, Coeli, de Fatima de Araujo Lucena et al. 2010), é melhor usar um ponto de corte mais estrito para maximizar a especificidade do processo de pareamento, evitando pares falsos mesmo aceitando algum nível de fracasso em encontrar pares

verdadeiros. O motivo é que se compararmos os fatores de risco de morte entre doadores ou receptores de sangue, pares falsos, obtidos como um erro não diferencial (ex: a mesma taxa de erro entre os níveis de fatores de risco) distorceria tanto da diferença de risco e a razão de risco em direção a hipótese nula, no entanto o fracasso em encontrar alguns pares verdadeiros somente distorceria a diferença de risco, mas não a razão de risco (Camargo e Coeli 2000).

5.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo

Neste estudo nós pudemos comparar a taxa de mortalidade de doadores de sangue soropositivos para doença de Chagas em relação a controles negativos. Nossos resultados indicam que os doadores soropositivos para Chagas tem 2,3 vezes mais chance de morrerem que os indivíduos soronegativos. Verificamos também que após classificar os óbitos ocorridos no período por Grupos de acordo com o CID 10 da causa de morte, o número de óbitos entre os doadores soropositivo foi maior no Grupo 3 - Doenças cardíacas, 47 de 159 (30%), enquanto que entre os soronegativos a frequência de óbitos foi maior no Grupo 5 - Traumatismos, 39 dos 103 óbitos (38%). Quando consideradas mortes devido à doença de Chagas ou mortes que tinham como causas anormalidades cardíacas que poderiam indicar Chagas, o risco instantâneo de morte entre indivíduos soropositivos é 17,9 vezes maior que entre soronegativos.

Como os doadores soropositivos co-infectados com outras doenças representam apenas 4,5% do total de doadores soropositivos de nosso estudo,

a exclusão destes da análise multivariada diminui porém não alterou nossos resultados.

A maioria dos estudos sobre mortalidade da doença de Chagas focaram nos pacientes já diagnosticados com a forma cardíaca (RASSI Jr A. 2007; Nunes et al. 2011). No entanto, existem estudos sobre mortalidade de indivíduos na fase crônica da doença com forma indeterminada (Maguire, Hoff, Sherlock et al. 1987; Dias 1989; Ribeiro e Rocha 1998; Ianni, Arteaga, Frimm et al. 2001), porém, como a progressão da doença é lenta, tais estudos necessitam de tamanho amostrais grandes e um longo período de seguimento para reunir o número de desfechos necessários. Estudos populacionais são caros, e em áreas com baixa prevalência como a cidade de São Paulo, com taxas menores que 1% (Sabino, Goncalvez, Salles et al. 2003), é necessário a triagem de um grande número de indivíduos para chegar a um número razoável de pacientes soropositivos.

Em uma revisão e meta análise recente sobre mortalidade por doença de Chagas (Cucunuba, Okuwoga, Basanez et al. 2016), 25 estudos foram incluídos na meta análise e o número de pessoas-ano destes estudos variou entre 4569 expostos / 7621 não expostos, estudo com maior número pessoas-ano, e 23 expostos / 61 não expostos, estudo com menor número de pessoas-ano. Em nosso estudo analisamos 33.295,04 soropositivos / 67.172,69 soronegativos pessoas-ano. Este grande número de pessoas-ano nos permite detectar diferenças que outros estudos não conseguiram detectar.

Em 2008, Guiltman e col nos Estados Unidos publicaram um estudo usando doadores de sangue para avaliar a mortalidade de hepatite C (Guiltinan et al. 2008). Os doadores de sangue são um grupo excepcional para este tipo

de estudo, pois a triagem sorológica é feita de forma rotineira e os controles negativos são fáceis de serem obtidos. Nós aproveitamos a idéia para definir da mesma maneira a taxa de mortalidade para doença de Chagas.

Diferente do que acontece nos EUA, não existe um identificador único para relacionar os indivíduos no banco de mortalidade que estejam presentes em outros bancos de dados e portanto a busca tem que ser nominal. No Brasil, o SIM tem cobertura a nível nacional e a qualidade dos dados vem melhorando substancialmente nas últimas duas décadas (Jorge, Laurenti e Gotlieb 2007; Franca, de Abreu, Rao et al. 2008; Jorge, Laurenti e Di Nubila 2010; Schmidt et al. 2015). Mesmo que possamos ter perdido algumas mortes no RPD, que como já mencionado anteriormente tem sensibilidade de 94% (IC 95%, 90% - 97%), e 100% (IC 95%, 98% - 100%) respectivamente, estamos seguros que as mortes encontradas representam pares verdadeiros.

A maior limitação de nosso estudo foi o fato que dados nominais não foram disponibilizados para os anos de 1996 a 2000, e portanto, não foi possível verificar se os doadores de sangue de nossa coorte morreram nos primeiros anos de seguimento. Teoricamente, poderíamos ter descartado os doadores deste período e analisado apenas aqueles que doaram sangue a partir de 2001, porém observamos que a frequência de doadores soropositivos em nosso banco de dados foi progressivamente menor durante os anos do estudo. Esta redução se deve provavelmente a dois fatores: 1) o controle da transmissão vetorial no país reduziu o número de indivíduos soropositivos na população em geral (Sabino, Goncalvez, Salles et al. 2003), e 2) Para aumentar o número de doações, de 1995 a 2001, a FPS promoveu uma campanha com o intuito de estimular o retorno de doadores soronegativos para todos os testes

de triagem ao hemocentro para nova doação, fazendo com que doadores de primeira vez fossem cada vez menos frequentes (Goncalez, Sabino e Charmone 2003). Se tivéssemos utilizado doadores de 2001 em diante, o número de doadores soropositivos teria sido substancialmente menor e o período de seguimento teria sido mais curto, o que poderia ter interferido em na análise desta doença de progressão lenta.

No entanto, as mortes nos primeiros anos após a doação não foram frequentes, como observado para aqueles que doaram nos anos de 2000 e 1999 (ver Tabela 13), para os quais pudemos verificar se morreram após um ou dois anos da doação, respectivamente. Além disso, mesmo que tivéssemos dados de mortes de todo o período do estudo, não temos nenhum motivo para supor que haveriam mais mortes entre os doadores soronegativos nos primeiros anos de seguimento que neutralizassem o excesso de mortes nos últimos anos de seguimento dos doadores soropositivos. Em resumo, nós acreditamos que esta limitação não teve grande influência no nosso resultado, principalmente em relação a comparação com os controles, pois tivemos o cuidado de parear os controles por ano de doação.

Outra limitação do nosso estudo foi o fato de não termos disponíveis os eletrocardiogramas e exames de raio x dos doadores no momento da triagem (Ribeiro et al. 2013). Sem estes exames não podemos classificar os sujeitos deste estudo como indivíduos na fase crônica e na forma indeterminada da doença de chagas. Porém acreditamos que por serem candidatos a doação e terem passado pela entrevista durante a triagem estes sejam indivíduos assintomáticos e sem cardiomiopatias significantes quando entraram no estudo (Sabino, Ribeiro, Salemi et al. 2013), pois entre as perguntas de triagem, os

candidatos à doação respondem perguntas como: "Você tem ou teve problema cardíaco grave?", e " Tem arritmia e usa medicação?". Porém não podemos afirmar que nenhum dos doadores tem ECG ou raio-X normais.

Nossos dados mostram que o risco instantâneo de morte de doadores soropositivos para a doença de Chagas foi de 2,3 vezes maior que os controles soronegativos, quando todas as causas de morte foram consideradas e 17,9 vezes maior quando as mortes consideradas foram devidas a doença de Chagas ou a causas básicas de morte devido a anormalidades cardíacas que poderiam indicar doença de Chagas.

Hoje sabemos que o tratamento etiológico não melhora os indivíduos com cardiomiopatia chagásica (Morillo et al. 2015), no entanto a discussão para os indivíduos na forma indeterminada ainda é uma questão aberta. Acreditamos que estes dados são relevantes, pois não é incomum encontrarmos artigos na literatura que consideram benigna e favorável a forma indeterminada da doença (Acquatella, Catalioti, Gomez-Mancebo et al. 1987; Maguire, Hoff, Sherlock et al. 1987; Ianni, Arteaga, Frimm et al. 2001). No entanto, esta conclusão é em geral baseada em seguimento de poucas pessoas e muitas vezes sem controle negativo apropriado (Cucunuba, Okuwoga, Basanez et al. 2016). Quando médicos lêem estes artigos, sentem que não tratar estes pacientes com as poucas opções terapêuticas disponíveis, nenhuma delas muito eficaz e com efeitos colaterais graves, é justificável (Ianni, Arteaga, Frimm et al. 2001; Mady et al. 2008). Nossos dados mostram que este conceito não está correto e que a mortalidade mesmo em populações mais jovens soropositivas está associada a uma taxa maior de mortalidade. Estudos sobre Chagas, esta doença negligenciada, são urgentemente

necessários para que novas drogas com menos efeitos colaterais sejam testadas e para encontrar correlações da progressão da doença.

Outro fato importante do nosso estudo foi mostrar que causas cardíacas, sem menção a doença de Chagas, foram significativamente mais comuns entre os soropositivos. Isso indica que o impacto da doença de Chagas na mortalidade brasileira está subestimada quando apenas os dados do SIM que mencionam a doença de Chagas são levados em consideração (Martins-Melo 2012). Quando levamos apenas os dados do SIM que tem como causa básica ou associada os códigos B57, a taxa de risco instantânea entre soropositivos e soronegativos é de 2,3 (IC 95%= 1,8 - 3,0). Porém quando acrescentamos nesta análise os códigos CID10 que definimos como possíveis causas de óbito por doença de Chagas descritos no SIM como causas básicas ou associadas nesta análise a taxa de risco instantânea sobe para 17,9 (IC 95%= 6,3 - 50,8).

6. Conclusão

6.1. Estudo Preliminar: Acurácia da metodologia de relacionamento probabilístico

Acreditamos que o método de relacionamento probabilístico possa ser utilizado em outros estudos que visem encontrar doadores de sangue da FPS e, se necessário, de outros bancos de sangue que colem dados de forma semelhante, no SIM, se um cuidado especial for tomado na escolha da nota de corte, para que a alta especificidade seja mantida.

6.2. Mortalidade entre doadores de sangue soropositivos para doença de Chagas (1996-2000) em São Paulo

Nossos dados mostram que doadores de sangue soropositivos para a doença de Chagas, que podemos considerar como indivíduos assintomáticos, seguidos retrospectivamente por 9 a 14 anos, morrem ao menos 2,3 vezes mais que doadores soronegativos. As causas de morte no SIM podem estar mal classificadas para indivíduos assintomáticos na fase crônica da doença.

7. Referências Bibliográficas

- Acquatella, H., F. Cataliotti, et al. (1987). "Long-term control of Chagas disease in Venezuela: effects on serologic findings, electrocardiographic abnormalities, and clinical outcome." *Circulation* **76**(3): 556-562.
- Andrade, D. V., K. J. Gollob, et al. (2014). "Acute Chagas Disease: New Global Challenges for an Old Neglected Disease." *PLoS Negl Trop Dis* **8**(7): e3010.
- Angheben, A., M. Anselmi, et al. (2011). "Chagas disease in Italy: breaking an epidemiological silence." *Euro Surveill* **16**(37).
- Barros Pereira Junior, C. e B. Markman Filho (2014). "Clinical and echocardiographic predictors of mortality in chagasic cardiomyopathy--systematic review." *Arq Bras Cardiol* **102**(6): 602-610.
- Basile, L., J. M. Jansa, et al. (2011). "Chagas disease in European countries: the challenge of a surveillance system." *Euro Surveill* **16**(37).
- Bern, C., S. Kjos, et al. (2011). "*Trypanosoma cruzi* and Chagas' Disease in the United States." *Clin Microbiol Rev* **24**(4): 655-681.
- Bern, C. e S. P. Montgomery (2009). "An estimate of the burden of Chagas disease in the United States." *Clin Infect Dis* **49**(5): e52-54.
- Bierrenbach, A. L., G. P. de Oliveira, et al. (2010). "Duplicates and misclassification of tuberculosis notification records in Brazil, 2001-2007." *Int J Tuberc Lung Dis* **14**(5): 593-599.
- Brasil (2001). Manual de procedimento do sistema de informações sobre mortalidade. FUNASA. Brasília, Assessoria de Comunicação e Educação em Saúde - Ascom: 36.
- Brasil (2011). Ministerio da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde Manual de Instruções para o Preenchimento da Declaração de Óbito. Brasília/DF, Ministério da Saúde.
- Brasil. (2011). Lei nº 12.527/2011.
- Brasil. (2012). DECRETO Nº 7.724, DE 16 DE MAIO DE 2012.
- Camargo, K. R., Jr. e C. M. Coeli (2000). "[Reclink: an application for database linkage implementing the probabilistic record linkage method]." *Cad Saude Publica* **16**(2): 439-447.
- Campbell, K. M., D. Deck, et al. (2008). "Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a 'basic' deterministic algorithm." *Health Informatics J* **14**(1): 5-15.
- Christen, P., Goiser, C. (2007). Quality and Complexity Measures for Data Linkage and Deduplication. *Quality Measures in Data Mining*. H. J. H. Fabrice J. Guillet. Berlin, Springer Berlin Heidelberg: 127- 151.
- Coeli, C. M., R. Blais, et al. (2003). "Probabilistic linkage in household survey on hospital care usage." *Rev Saude Publica* **37**(1): 91-99.
- Comunicação (1985). Validade do conceito de forma indeterminada de doença de Chagas. *Revista da Sociedade Brasileira de Medicina Tropical*, I Reunião Anual de Pesquisa Aplicada em Doença de Chagas. **18**: 46.
- Cook, L. J., L. M. Olson, et al. (2001). "Probabilistic record linkage: relationships between file sizes, identifiers and match weights." *Methods Inf Med* **40**(3): 196-203.
- Coura, J. R., L. L. de Abreu, et al. (1985). "Morbidity in Chagas' disease. IV. Longitudinal study of 10 years in Pains and Iguatama, Minas Gerais, Brazil." *Mem Inst Oswaldo Cruz* **80**(1): 73-80.

- Coura, J. R. e J. C. Dias (2009). "Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery." Mem Inst Oswaldo Cruz **104 Suppl 1**: 31-40.
- Cucunuba, Z. M., O. Okuwoga, et al. (2016). "Increased mortality attributed to Chagas disease: a systematic review and meta-analysis." Parasit Vectors **9**(1): 42.
- Dean, J. (1996). "Probabilistic Linkage." from <http://www.nedarc.org/statisticalHelp/advancedStatisticalTopics/probabilisticLinkage.html>.
- Dias, J. C. (1989). "The indeterminate form of human chronic Chagas' disease A clinical epidemiological review." Rev Soc Bras Med Trop **22**(3): 147-156.
- Dias, J. C., E. M. Machado, et al. (2002). "[Chagas' disease in Lassance, Minas Gerais State: Clinical-epidemiological re-evaluation ninety years after the discovery by Carlos Chagas]." Rev Soc Bras Med Trop **35**(2): 167-176.
- Dias, J. C. P. e J. R. Coura (1997). Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral [online]. Rio de Janeiro, FIOCRUZ.
- Fellegi, I. P. e A. B. Sunter (1969). "A Theory for Record Linkage." Journal of the American Statistical Association **64**(328): 1183-&.
- Fonseca, M. G., C. M. Coeli, et al. (2010). "Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database." Cad Saude Publica **26**(7): 1431-1438.
- Fortes, P. e E. Zoboli (2008). Ética na Saúde Pública. Saúde Pública: bases conceituais. São Paulo, Atheneu.
- Franca, E., D. X. de Abreu, et al. (2008). "Evaluation of cause-of-death statistics for Brazil, 2002-2004." Int J Epidemiol **37**(4): 891-901.
- Gonzalez, T., E. C. Sabino, et al. (2003). "Trends in the profile of blood donors at a large blood center in the city of São Paulo, Brazil." Rev Panam Salud Publica **13**(2-3): 144-148.
- Guiltinan, A. M., Z. Kaidarova, et al. (2008). "Increased all-cause, liver, and cardiac mortality among hepatitis C virus-seropositive blood donors." Am J Epidemiol **167**(6): 743-750.
- Hulley, S. B., Cummings, Steven R., Browner, Warren S., Grady, Deborah G., Newman, Thomas B. (2008). Delineando a Pesquisa Clínica: uma abordagem epidemiológica. Porto Alegre, Artmed.
- Ianni, B. M., E. Arteaga, et al. (2001). "Chagas' heart disease: evolutive evaluation of electrocardiographic and echocardiographic parameters in patients with the indeterminate form." Arq Bras Cardiol **77**(1): 59-62.
- Indexmundi, C. W. F.-. (2011). "Brasil Taxa de mortalidade." Retrieved 19/05/2012, from http://www.indexmundi.com/pt/brasil/taxa_de_mortalidade.html.
- Jaro, M. A. (1989). "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa,Florida." Journal of the American Statistical Association **84**(406): 414-420.
- Jaro, M. A. (1995). "Probabilistic linkage of large public health data files." Stat Med **14**(5-7): 491-498.
- Jorge, M. H., R. Laurenti, et al. (2010). "[Death and its epidemiological investigation: considerations about some relevant aspects]." Rev Bras Epidemiol **13**(4): 561-576.
- Jorge, M. H., R. Laurenti, et al. (2007). "[Quality analysis of Brazilian vital statistics: the experience of implementing the SIM and SINASC systems]." Cien Saude Colet **12**(3): 643-654.
- Jurczyk, P. (2009). Fine-grained Record Integration and Linkage Tool. Atlanta, Emory University Math & CC Department.
- Lopes, E. R., E. Chapadeiro, et al. (1981). "[Pathological anatomy of hearts from asymptomatic Chagas disease patients dying in a violent manner]." Mem Inst Oswaldo Cruz **76**(2): 189-197.

- Mady, C., B. M. Ianni, et al. (2008). "Benznidazole and Chagas disease: can an old drug be the answer to an old problem?" Expert Opin Investig Drugs **17**(10): 1427-1433.
- Maguire, J. H., R. Hoff, et al. (1987). "Cardiac morbidity and mortality due to Chagas' disease: prospective electrocardiographic study of a Brazilian community." Circulation **75**(6): 1140-1145.
- Manzullo, E. C. e R. Chuit (1999). "Risk of death due to chronic chagasic cardiopathy." Mem Inst Oswaldo Cruz **94 Suppl 1**: 317-320.
- Martins-Melo, F., Alencar CH, Ramos Jr AN, Heukelbach J (2012). "Epidemiology of Mortality Related to Chaga's Disease in Brazil, 1997 - 2007." PloS Negl Trop Dis **6**(2): e1508.
- Martins-Melo, F. R., A. N. Ramos Junior, et al. (2012). "Multiple causes of death related to Chagas' disease in Brazil, 1999 to 2007." Rev Soc Bras Med Trop **45**(5): 591-596.
- Massad, E. (2008). "The elimination of Chagas' disease from Brazil." Epidemiol Infect **136**(9): 1153-1164.
- Migowski, A., R. B. Chaves, et al. (2011). "Accuracy of probabilistic record linkage in the assessment of high-complexity cardiology procedures." Rev Saude Publica **45**(2): 269-275.
- Morillo, C. A., J. A. Marin-Neto, et al. (2015). "Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy." N Engl J Med **373**(14): 1295-1306.
- MS (2011). PORTARIA MS 1.353/2011.
- Newcombe, H. B., J. M. Kennedy, et al. (1959). "Automatic linkage of vital records." Science **130**(3381): 954-959.
- Nunes, M. C., W. Dones, et al. (2013). "Chagas disease: an overview of clinical and epidemiological aspects." J Am Coll Cardiol **62**(9): 767-776.
- Nunes, M. C., R. C. Reis, et al. (2011). "Risk estimation approach in Chagas disease is still needed." Int J Cardiol **147**(2): 294-296.
- OMS (1993). CID-10. São Paulo, Edusp - Editora da Universidade de São Paulo
- OMS (2010). Chagas disease: control and elimination.
- OMS (2010). Chagas disease in Latin America: an epidemiological updated based on 2010 estimates. 6: 33-44.
- OMS (2011). "Working to overcome the global impact of neglected tropical diseases - Summary." Wkly Epidemiol Rec **86**(13): 113-120.
- OMS (2012). Chapter 4 - Record Linkage. Assessing tuberculosis under-reporting through inventory studies, Available at http://www.who.int/tb/publications/inventory_studies/en/index.html.
- Pacheco, A. G., V. Saraceni, et al. (2008). "Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil." Am J Epidemiol **168**(11): 1326-1332.
- Rassi, A., Jr., A. Rassi, et al. (2010). "Chagas disease." Lancet **375**(9723): 1388-1402.
- RASSI Jr A., R. A., Rassi SG (2007). "Predictors of Mortality in Chronic Chagas Disease: A Systematic Review of Observational Studies." Journal of the American Heart Association **115**: 1101 - 1108.
- Ribeiro, A. e Rocha (1998). "Forma indeterminada da doença de Chagas: considerações acerca do diagnóstico e do prognóstico." Revista da Sociedade Brasileira de Medicina Tropical **31**(3): 301-314.
- Ribeiro, A. L., M. S. Marcolino, et al. (2014). "Electrocardiographic abnormalities in elderly Chagas disease patients: 10-year follow-up of the Bambui Cohort Study of Aging." J Am Heart Assoc **3**(1): e000632.
- Ribeiro, A. L., E. C. Sabino, et al. (2013). "Electrocardiographic abnormalities in Trypanosoma cruzi seropositive and seronegative former blood donors." PLoS Negl Trop Dis **7**(2): e2078.

- Russel, C. R. (1922). R.C, Russel Index. USA. **US1435663 (A)** — **1922-11-14**.
- Sabino, E. C., T. T. Goncalvez, et al. (2003). "Trends in the prevalence of Chagas' disease among first-time blood donors in Sao Paulo, Brazil." Transfusion **43**(7): 853-856.
- Sabino, E. C., A. L. Ribeiro, et al. (2013). "Ten-year incidence of Chagas cardiomyopathy among asymptomatic Trypanosoma cruzi-seropositive former blood donors." Circulation **127**(10): 1105-1115.
- Sabino, E. C., N. A. Salles, et al. (2010). "Enhanced classification of Chagas serologic results and epidemiologic characteristics of seropositive donors at three large blood centers in Brazil." Transfusion **50**(12): 2628-2637.
- Saúde, M. d. (2014). "Datasus - Mortalidade Brasil." Retrieved 31/08/2014, from <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def>.
- Schmidt, M. I., B. B. Duncan, et al. (2015). "Trends in mortality due to diabetes in Brazil, 1996-2011." Diabetol Metab Syndr **7**: 109.
- Schmunis, G. A. e Z. E. Yadon (2010). "Chagas disease: a Latin American health problem becoming a world health problem." Acta Trop **115**(1-2): 14-21.

Apêndices

**Apêndice 1: Accuracy of a probabilistic record-linkage methodology
used to track blood donors in the Mortality Information System database**

Accuracy of a probabilistic record-linkage methodology used to track blood donors in the Mortality Information System database

Acurácia da metodologia de relacionamento probabilístico de registros na identificação de doadores de sangue no banco de dados do Sistema de Informações sobre Mortalidade

La exactitud del método probabilístico de relación para identificar a donantes de sangre en la base de datos del Sistema de Información sobre Mortalidad

Ligia Capuani ¹
Ana Luiza Bierrenbach ²
Fatima Abreu ¹
Pedro Losco Takecian ³
João Eduardo Ferreira ³
Ester Cerdeira Sabino ¹

Abstract

The probabilistic record linkage (PRL) is based on a likelihood score that measures the degree of similarity of several matching variables. Screening test results for different diseases are available for the blood donor population. In this paper, we describe the accuracy of a PRL process used to track blood donors from the Fundação Pró-Sangue (FPS) in the Mortality Information System (SIM), in order that future studies might determine the blood donor's cause of death. The databases used for linkage were SIM and the database made up of individuals that were living (200 blood donors in 2007) and dead (196 from the Hospital das Clínicas de São Paulo that died in 2001-2005). The method consists of cleaning and linking the databases using three blocking steps comparing the variables "Name/Mother's Name/ Date of Birth" to determine a cut-off score. For a cut-off score of 7.06, the sensitivity and specificity of the method is 94.4% (95%CI: 90.0-97.0) and 100% (95%CI: 98.0-100.0), respectively. This method can be used in studies that aim to track blood donors from the FPS database in SIM.

Medical Record Linkage; Blood Donors; Mortality Rate

Resumo

O relacionamento probabilístico se baseia em um escore que é calculado levando em consideração a similaridade do pareamento de diversas variáveis. Dados de resultados de testes de triagem para diferentes doenças estão disponíveis para a população de doadores de sangue. Neste artigo descrevemos a acurácia de um processo de relacionamento probabilístico para identificar doadores de sangue da Fundação Pró-Sangue (FPS) no Sistema de Informações sobre Mortalidade (SIM). Os bancos utilizados para o relacionamento foram o SIM e o banco formado por indivíduos vivos (200 doadores de sangue em 2007) e mortos (196 pacientes do Hospital das Clínicas de São Paulo que morreram entre 2001-2005). O método consistiu em limpar e relacionar probabilisticamente os bancos em três passos de bloqueio comparando as variáveis "Nome/Nome Mãe /Data de Nascimento" para determinar um escore de corte. Para um escore de corte de 7,06 a sensibilidade e especificidade do método é de 94,4% (IC95%: 90-97) e 100% (IC95%: 98-100), respectivamente. Este método pode ser utilizado em estudos para identificar pacientes da FPS no SIM.

Registro Médico Coordenado; Doadores de Sangue; Coeficiente de Mortalidade

¹ Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil.
² Instituto de Ensino e Pesquisa, Hospital Sírio Libanês, São Paulo, Brasil.
³ Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil.

Correspondence
L. Capuani
Departamento de Moléstias Infecciosas e Parasitárias, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo 455, São Paulo, SP 01246-903, Brasil.
lcapuani@gmail.com

Introduction

Record linkage is a process that aims to accurately identify whether two or more records relate to the same individual. In Brazil, it has been used increasingly for studies involving the health information system databases managed by the Brazilian Ministry of Health, including the Mortality Information System (SIM). The process is used not only to detect and remove duplicate records relating to the same individual from within a database¹ but also to integrate records across health information system databases and in between them and other databases^{2,3,4,5,6,7}. By doing so, these studies help to improve data quality and integrity, allow for the reuse of existing data that can be used for answering a broader range of study questions, reduce costs and efforts related to data acquisition and also help to improve record linkage methodology and its assessment⁸.

Our team has been involved with studies of blood banking and transfusion medicine for many years. Some of our studies have made exclusive use of the Fundação Pró-Sangue Hemocentro de São Paulo (FPS) blood bank donors' and recipients' routine secondary data, while others have made use of such data in conjunction with primary data collected within the context of the international collaboration Retrovirus Epidemiology Donor Study (REDS II)^{9,10} and The Recipient Epidemiology and Donor Evaluation Study-III (REDS-III). We are now particularly interested in performing longitudinal studies in which we need to ascertain some long term outcomes (e.g. vital status) of donors and transfused recipients included in our primary and secondary databases. This will allow us to investigate their disease progression and case fatality rates, therefore contributing to a better characterization of the natural history of many transfusion-related diseases, as well as their history under certain controlled circumstances.

Such long-term data is not contained in our databases, so we depend on tracking our patients from other secondary sources that are known to include the relevant information, such as the SIM. There are two major complications in doing so. First, the number of records in these databases is huge. The SIM database, for example, has over 1 million records per year. Secondly, these databases do not share a unique personal identifier that would enable us to simply join them (like the US social security number or the long awaited "cartão do Sistema Único de Saúde (SUS)" number) and they do have many inexorable errors, abbreviations and missing or incorrectly formatted values in the personal identification variables (like name, mother's name and date of

birth). In such a scenario, a more sophisticated record linkage process is needed. As our intention is to repeat the linkage process several times using the same and also other similar data sources over time, the process needs to be robust and reproducible, and should preferably not be overly dependent on human intervention. In transfusion medicine, just a few longitudinal studies have been done which were made possible by the consolidation of different data sources through record linkage methods¹¹, but many such studies have greatly contributed in other topics^{12,13,14,15}.

There are two main types of record linkage: the deterministic record linkage in which the linkage of records is based on exact agreement of one or several matching variables, and the probabilistic linkage, which is based on a likelihood score that is calculated taking into consideration how similar each of several matching variables are. Pairs with high scores have higher probabilities of being true matches, and pairs with low scores have lower probabilities. The problem lies in pairs in the intermediate "grey" zone, which usually require manual review in order to be satisfactorily classified¹⁶. Having to depend on manual review often involves a high cost for probabilistic linkage exercises in terms of the engagement of skilled human resources in order to perform such a task, and it also raises privacy concerns since these individuals have access to nominal information. Alternatively, the need for manual review can be removed by explicitly selecting a single cut-off value above which all pairs are declared true matches.

In studies that aim to assess the performance of a probabilistic record linkage process, such as this one, the results of the manual review or of a database with known true matches and non-matches are often considered the "gold-standard" against which different values of the likelihood score are compared. Among the different measures that have been used for assessing the accuracy of record linkage processes are sensitivity, positive predictive value and specificity. Sensitivity is the capacity of the linkage process to recognize the true matches, while specificity is the capacity of the linkage process to recognize non-matches. Positive predictive is the proportion of matches found by the linkage process that are true. Sensitivity and positive predictive value are often respectively called recall and precision in the linkage field¹⁷. As in studies of diagnostic tests, there is often a trade-off between recall and precision, in which high precision can normally only be attained at the cost of low recall, and vice-versa.

In this paper, we aimed to study the accuracy of a probabilistic linkage process used to track

our blood bank patients in the SIM databases. Our goal was to define cut-off scores that would maximize sensitivity and minimize as much as possible the need to perform manual review of future studies that aimed to describe mortality rates of blood transfusion transmitted diseases in Brazilian individuals.

Materials and methods

Data sources

The reference file was created by appending 200 randomly selected records of the FPS database with 196 randomly selected hospital discharge records of the Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo (HCFMUSP) database. In the former database, 200 records were randomly chosen among the records of individuals aged 18 and above that had donated blood in 2007 and were, therefore, known to be alive during the period from 2001 to 2005. In the latter case, 196 records were randomly chosen among records of individuals aged 18 and above who were known to have died during their hospitalization in the period from 2001 to 2005, as reported in the database's outcome variable. Given the ratio of 200 to 196 individuals, we were artificially creating a scenario in which the death ratio was close to 50%.

The comparison file was the SIM database (version 19/Apr/2007) for the period of 2001 to 2006, containing records of all Brazilian individuals who had died at the age of 18 and above. Due to notification delays, as the database extraction had been done on 19 April 2007, the 2006 data were still incomplete.

The number of 196 deceased individuals to be included in our sample size was estimated taking into account an expected sensitivity of 85% (P), a 95% confidence interval ($Z_{\alpha} = 1.96$) with a total width of 10% (W) and using the equation $N = 4 Z_{\alpha}^2 P(1-P) \div W^2$ ¹⁸.

Pre-processing

Both the reference and the comparison files contained the variables identity number, name, mother's name, gender, date of birth, among others. They underwent an extensive pre-processing step aiming to clean and standardize the data to be used in the linkage process. The pre-processing was done using SQL-Analysis Server (Microsoft SQL Server 2008), and involved the following data transformations: the value 9999 was defined for all non-numeric values for age and for records in which age was less than 18, and were

excluded (496,056 records); all records with missing mother's name were kept but "***" was filled in (222,044); the value 9 was defined for all missing or invalid information for gender; date of birth null values were transformed into "18000101" (47,533). Importantly, in the SIM database, records with missing names or with invalid names (477,174) or with missing death certificate IDs (0) were excluded from subsequent linkage steps. The excluded records represented 16.9% of the total number of records. It is important to note that for this version of SIM, 27.9% of 2001 records were excluded because of missing names or invalid names, whereas for the years of 2002, 2003, 2004, 2005 and 2006 only 4% of the records were excluded for the same reason. After the cleaning steps the SIM database used for the linkage had 4,775,164 records as described in Figure 1.

Linkage strategy

The record linkage was performed using a software called Record Linkage III – version 3.0.4 4005 (RECLINK-III; http://www.iesc.ufrj.br/reclink/RecLink_arquivos/RecLinkdl.html), an open-source probabilistic linkage software specially developed to associate records taking into consideration the Portuguese language phonetics ¹⁹. Standardization of common fields was performed where correction of upper/lower case variations was done, removal of accents and prepositions inside names (e.g. "Maria Conceição dos Santos" → "Maria Conceicao Santos"), standardization of date formats was performed and finally, removal of commas and other punctuation marks. All missing values were correctly classified as such. Using the variable name, four other secondary variables were generated: first name, last name and first and last names phonetic codes. Three blocking steps strategy were used from a combination of the phonetic codes of the variables first name, last name and gender ²⁰. By allowing comparisons of record pairs only within blocks, the computational task is much reduced. The first blocking step used first name phonetic code, last name phonetic code and gender. The second blocking step used only first name phonetic code and gender and the third blocking step used only last name phonetic code and gender.

The matching variables used for pairing were name, mother's name and date of birth. RECLINK-III estimates scores for each pair of records, with a higher score representing a higher likelihood of the pair being a true match. Scores are proportional to log-likelihood ratios, which are derived in a standard probabilistic linkage approach. In this approach, m is the probability that the characters sequence of a particular matching

Figure 1

Cleaning description of the reference and comparison files used on the linkage process.

FPS: Fundação Pró-Sangue Hemocentro de São Paulo; HC-FMUSP: Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo; SIM: Mortality Information System.

variable agree for records that are true matches (m is analogous to sensitivity), and u is the probability that the characters sequence of a particular matching variable agree for records that are false matches (u is analogous to 1-specificity), i.e. by chance alone. When the contents of the

variable do agree in between the records being compared, the score is proportional to logarithm (m/u), and when they disagree to logarithm $((1-m)/(1-u))$. Name and mother's name had their characters' sequence compared using the Levenshtein distance, which returned values between

1 (total concordance) and 0 (total discordance) with 92% sensitivity (m) and 1% 1-specificity (u), with a minimum proportion of concordance of 85%. Date of birth had its sequences of digits, ignoring separators, compared using the character algorithm which returned values between 1 (total concordance) and 0 (total discordance), with 90% sensitivity (m) and 5% 1-specificity (u), with a minimum concordance of 65%. The scores for each matching variable are added to determine a composite score¹⁶. Total concordance of all three matching variables resulted in a maximum score of $\log(0.92/0.01) + \log(0.92/0.01) + \log(0.9/0.05) = 17.22$, while total discordance of all three matching variables resulted in a minimum score of $\log(0.08/0.99) + \log(0.08/0.99) + \log(0.1/0.95) = -10.51$, using base 2 logarithms.

Post-processing

The manual review of the pairs formed during each blocking step was performed by one of the authors (L.C.) in the module "Combine" of the RECLINK-III software. In this module, the user is able to see the score assigned to each pair and to visually compare variables from both records, including matching variables as well as additional ones that did not participate in the linking process. All pairs were manually reviewed except those in the last score level (-10.51), which comprised the majority of pairs (97.5% out of 709,550 for the first blocking step, 99.9% out of 1,448,313 for the second blocking step) and were directly considered non-matches. After classifying pairs into matches and non-matches, RECLINK-III automatically created three files: a file with matched pairs of records from both the reference and the comparison files, a file with all non-matched records from the reference file and another with all non-matched records from the comparison file. These two latter files were then used in the next blocking steps. The three files with the matched pairs of records found at each of the blocking steps were then appended for analysis. All steps in the linkage process were done blind to the vital status of individuals in our reference file.

Data analysis

In order to test the performance of our record linkage strategy, sensitivity and specificity and their 95% binomial confidence intervals were calculated assuming different cut-off points for score. Sensitivity was defined as the proportion of true matches among the 196 individuals known to be dead (which should have corresponding records in SIM), and specificity was defined as the proportion of non-matches among

the 200 individuals known to be alive (which should not have corresponding records in SIM). For the chosen cut-off point, the positive predictive value was also calculated as the proportion of true matches among the sum of true and false matches, along with its 95% binomial confidence interval. Sensitivity and positive predictive value are often respectively called recall and precision in the linkage field¹⁷.

As expected, for every record in the reference file many matches were generated, but only the pair with the highest score was considered in the sensitivity and specificity calculations. In other words, we only considered the highest score in which a record from our reference file appeared as a pair. When two or more pairs were formed with the same score, which included the same record from the reference file, only the most comparable one, as classified by the "gold standard" manual review, was considered to be a true match (for the 196 individuals known to be dead) or a false match (for the 200 individuals known to be alive).

The choice of the cut-off point took into consideration the need to maximize the specificity of the matching strategy, i.e. to minimize the possibility of making false matches while if necessary accepting some level of failures to find true matches.

Ethical approval for this study was obtained from the Ethics Research Committee of HC-FMUSP (CAAE: 0543.0.015.000-08).

Results

The reference file, which was composed of an extraction of 200 records from the FPS database and 196 records from the HC-FMUSP database, had 100% completion for the matching variables name, mother's name and data of birth, while the fields name and mother's name had no abbreviations. The comparison file, which was the SIM database for the period of 2001 to 2006 containing 4,775,164 records of individuals who had died at aged 18 and above, had 100% completion for name (as this was a prerequisite for SIM records to be included in the comparison file), 95.35% completion for mother's name and 99% completion for date of birth.

In the first blocking step, 179 true matches were formed and in the second blocking step, this number increased to 185 out of the total 196 possible true matches. As no new true match was found in the third blocking step, we will only report results from the first two.

Table 1 shows the totality of pairs formed in between the reference and the comparison

Table 1

Sensitivity, specificity and positive predictive value of several scores of the probabilistic linkage process.

Score	Step 1		Step 2		Steps 1 + 2		Steps 1 + 2	Steps 1 + 2	Steps 1 + 2
	All pairs	First time pairs *	All pairs	First time pairs *	True matches	False matches			
							Sensitivity [% (95%CI)]	Specificity [% (95%CI)]	Positive predictive value [% (95%CI)]
17.22	147	147	0	0	147	0	75.0 (68.0-80.0)	100.0 (98.0-100.0)	100.0 (97.0-100.0)
17.01	2	2	0	0	2	0	76.0 (69.0-82.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.99	1	1	0	0	1	0	76.5 (70.0-82.0.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.97	2	2	0	0	2	0	77.6 (71.0-83.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.95	1	1	0	0	1	0	78.1 (72.0-84.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.93	2	2	0	0	2	0	79.1 (73.0-85.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.91	1	1	0	0	1	0	79.6 (73.0-85.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.89	1	1	0	0	1	0	80.1 (74.0-85.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.87	3	3	1	1	4	0	82.1 (76.0-87.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.78	1	1	0	0	1	0	82.7 (77.0-88.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.70	4	4	0	0	4	0	84.7 (79.0-89.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
16.60	2	2	0	0	2	0	85.7 (80.0-90.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
9.80	5	1	0	0	1	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
9.50	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
9.03	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
8.98	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
8.83	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
8.55	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
8.26	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
7.98	1	0	0	0	0	0	86.2 (81.0-91.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
7.06	12	11	5	5	16	0	94.4 (90.0-97.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
6.54	2	0	0	0	0	0	94.4 (90.0-97.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
6.23	2	0	0	0	0	0	94.4 (90.0-97.0)	100.0 (98.0-100.0)	100.0 (98.0-100.0)
6.02	34	6	0	0	0	6	94.4 (90.0-97.0)	97.0 (94.0-99.0)	97.0 (93.0-99.0)
5.86	1	0	0	0	0	0	94.4 (90.0-97.0)	97.0 (94.0-99.0)	97.0 (93.0-99.0)
5.78	1	0	0	0	0	0	94.4 (90.0-97.0)	97.0 (94.0-99.0)	97.0 (93.0-99.0)
5.74	1	0	0	0	0	0	94.4 (90.0-97.0)	97.0 (94.0-99.0)	97.0 (93.0-99.0)
5.71	14	1	0	0	0	1	94.4 (90.0-97.0)	96.5.0 (93.0-99.0)	96.0 (93.0-98.0)
5.70	1	1	0	0	0	1	94.4 (90.0-97.0)	96.0 (92.0-98.0)	96.0 (92.0-98.0)
5.65	1	1	0	0	0	1	94.4 (90.0-97.0)	95.5.0 (92.0-98.0)	95.0 (91.0-98.0)
5.64 to -10.51	709284	106	1448307	85	0	191	94.4 (90.0-97.0)	95.5 to 0.0	48 to 0.0
Total	709532	294	1448313	91	185	200			

95%CI: 95% confidence interval.

* "First time pairs" refers to the first time a record from the reference file was found in a pair, searching from the highest to the lowest score.

databases, given the limitations imposed by the blocking variables. Pairs with high scores were mostly found in step 1. When considering the number of pairs formed in step 1 from the highest to the lowest scores, we can see that at score 9.8, even though five pairs were found, only one of them referred to the first time a specific record from the reference file was found in a pair. The other four referred to records from the reference file that had already formed pairs with the same or higher scores. The same happened at many lower scores, increasing as the score decreased. Most true matches were found at score 17.22. Only one to four true matches were found at scores ranging from 17.01 to 9.8. Then no true matches were found up to score 7.06, where 16 were found. The first six false matches appeared at score 6.02, and from there numbers increased very slowly. It was only at the very lowest scores that the majority of the false matches were found. Sensitivity at score 9.08 was 86.2% (95%CI: 80.5-90.7) and at score 7.06 was 94.4% (95%CI: 90.2-97.2), the maximum value obtained. Specificity was 100% (95%CI: 98.2-100.0) up to score 6.02. Such results would point to a natural choice of a score cut-off value able to discriminate pairs into true and false matches, as the last true matches were found at score 7.06 and the first false matches appeared only at score 6.02. Using a cut-off inside this range, the positive predictive value was 100% (95%CI: 98.2-100.0).

The manual review revealed why 16 true matches were found only at score 7.06. The problem was in the variable name. In one of the records of these pairs, one element of the name was missing, e.g. in one record one woman possibly had her maiden name (Maria Antonia Lima) and in the other her married name (Maria Antonia Lima Castro – this is a fictitious name).

Discussion

The probabilistic linkage process used in the present study aims to help track blood bank patients from the FPS database in the SIM. In order to assess the accuracy of this linkage process, we assembled a file in which we knew beforehand the death status of its recorded individuals. By doing so, we were able to know which of the obtained pairs were true or false matches. This strategy of using a "known vital status" in order to ascertain linkage accuracy has been used in a number of studies, including a few Brazilian ones^{2,3}.

The accuracy of our linkage process was very high, with a sensitivity of 94% and a specificity of 100% using a score in between 7.06 and

6.02 as the cut-off point. Our results are slightly better than those of a Brazilian study that used RECLINK-III and the same matching variables in order to link the AIDS surveillance database with SIM, where they found a sensitivity of 87.6% and a specificity of 99.6%.² However, apart from the differences in the size of our reference file and their surveillance database, we have to bear in mind that while our reference file did not have any missing or incomplete values in the matching variables, theirs was a much bigger database that was likely to have many records with these types of errors. The use of large databases obviously also makes the manual review process a real challenge. Our results are comparable to those of another Brazilian study that used an in-house deterministic linkage process in order to link data from two study cohorts of HIV-infected patients with SIM, where they found a sensitivity of 96.5% and a specificity of 99.6%.²

Although sensitivity was high, the linkage process was still unable to find 11 individuals known to be dead from the reference file in the SIM database. There are at least four possible explanations for this finding: (1) the death of these individuals had not been registered in the SIM database, (2) the death of these individuals had been registered in the SIM database, but their records were among those that were discarded before the linkage process either because they had missing data for the variable name, or had been wrongly recorded as belonging to an individual younger than 18 years of age or had missing data for death certificate IDs, (3) the death of these individuals had been registered in the SIM database, but information available at the three matching variables differed to such an extent between the reference and the comparison files that the pair obtained was classified with the last possible score, i.e. -10.52, in which case it was not submitted to manual review, and therefore was not found, and (4) pairs were not formed because records were blocked by the variable gender, present in all three of our blocking steps.

The former explanation suggests that there is under notification in SIM. SIM has undergone significant improvements in coverage and overall data quality over the last decade²¹, but mostly these improvements were needed for remote rural areas of the Northeastern and Northern regions of the country, and for individuals who had died outside hospitals and were at the extremes of age²². This was not the case for the 196 individuals from our reference file who were known to have died during their hospitalization at the biggest public university hospital in the city of São Paulo, where they had certainly been granted a death certificate. Another possibility is that

these deaths had not yet been registered at the SIM database due to notification delays. But we also think this is unlikely because they happened from 2002 to 2005, and the SIM database was extracted in April, 2007.

The second and third explanations refer to quality issues in the SIM database, which may have prevented us from reaching a higher sensitivity. The sensitivity of record linkage studies is very much dependent on the quality of its data sources. We believe it is likely that the 11 missing individuals were included in the 16.7% of SIM records that had been discarded from our linkage process due to missing data on the variables name, or that had been wrongly recorded as belonging to an individual younger than 18. On the other hand, considering the nature of the probabilistic linkage and the fact that the last score represents pairs that had total discordance of all three matching variables, it is most unlikely that any true matches for these 11 individuals would be found at the score -10.51, which was not manually reviewed.

Gender was chosen as a good blocking variable to be used in combination with others, as we believed not many records were likely to be wrongly classified or to have missing values in this field. However rare, such mistakes may have happened, even though it is unlikely that this possibility alone would explain all of the 11 missing individuals. In any case, in our future linkages, we will not have the same variable participating in all of our blocking steps.

With only 396 individuals coming from two data sources (FPS and HC-FMUSP databases), our reference file was relatively small and therefore not very heterogeneous. If we were to repeat our linkage process using the whole or a more representative sample of the FPS database, the bigger number of records would obviously increase the variability of the values presented in our matching variables, which would mostly increase the number of false matches, and possibly increase the scores of these false matches^{3,17}. As this could decrease the specificity observed for a given score value, perhaps a natural choice of a score cut-off point, as we had in the present study, would not be presented. The immediate consequence would be the need to check whether the same cut-off point used in the pres-

ent study could apply to the new exercise. This would be accomplished by manually reviewing at least some of the pairs in the so-called grey zone. Importantly, while linking records from primary or secondary data sources to a mortality database, we are dealing with a situation in which we obviously expect that only one true match will be found per individual, unless duplicates exist in one or both files. Today, Reclink removes from both databases the records that are part of a pair considered a true pair in a blocking step and on the next blocking step these records will not form pairs. As a suggestion to improve and accelerate the manual review in each blocking step, Reclink could automatically remove from the list of pairs that will be submitted to the manual review all pairs with lower scores that contain one or the other of two records of pairs that have reached a higher score. This strategy would significantly decrease the burden of the manual review.

Another consequence of repeating our linkage process using the whole or a more representative sample of the FPS database, which would very likely have a different death ratio, would naturally be the change of the positive predictive value for the chosen cut-off point¹⁸. Most importantly, such accuracy changes would also impart on how the longitudinal studies that will depend on the linkage results would be interpreted. As mentioned in other similar studies³, it is best to use a more stringent cut-off point so as to maximize the specificity of the matching process, avoiding false matches even while accepting some level of failures to find true matches. The reason for this is that if we are to compare risk factors for death among blood donors or recipients, false matches obtained as a non-differential error (i.e. the same error rate across the risk factor levels) would bias both the risk difference and the risk ratio towards the null hypothesis, whereas failing to find some true matches would only bias the risk difference but not the risk ratio¹⁹.

In conclusion, we believe our record linkage process can be used in studies that aim to track blood bank patients from the FPS database and, if necessary, from other blood bank databases that collect similar data, in the SIM, if special care is taken in the selection of the cut-off point, so as to guarantee that a high specificity be maintained.

Resumen

La relación probabilística (RP) se basa en una puntuación que se calcula en función de la similitud entre variables de emparejamiento. Los resultados de los tests sobre diferentes enfermedades están a disposición de la población de donantes de sangre. En el presente artículo se describe la precisión de un proceso de RP para identificar a donantes de sangre de la Fundação Pró-Sangue (FPS) en el Sistema de Información de Mortalidad (SIM). Se llevó a cabo la RP del SIM y de un banco compuesto por individuos vivos (200 donantes de sangre en 2007) y muertos (196 pacientes del Hospital de Clínicas de São Paulo, que murieron entre 2001 y 2005). El método consistió en depurar los bancos de datos y RP en tres etapas de bloqueo, comparando las variables nombre, nombre de la madre y fecha de nacimiento para determinar un punto de corte. Para el punto de corte 7:06, la especificidad y sensibilidad del método fue de un 94,4% (IC95%: 90,0-97,0) y 100% (IC95%: 98,0-100,0), respectivamente. Este método puede ser utilizado en más estudios con el fin de identificar a los pacientes FPS en el SIM.

Registro Médico Coordinado; Donantes de Sangre; Tasa de Mortalidad

Contributors

L. Capuani contributed to the conception and design of the study; acquisition, analysis, and interpretation of data for the study; drafting the study, and revising it for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work. A. L. Bierrenbach contributed to the conception and design of the study; analysis and interpretation of data for the study; drafting the study and critically revising it for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the study. F. Abreu contributed to the drafting the study and critically revising it for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the study. P. L. Takecian contributed to the analysis and interpretation of data for the study; critically revising it for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the study. J. E. Ferreira and E. C. Sabino contributed to the conception and design of the study; analysis and interpretation of data for the study; critically revising it for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the study.

Acknowledgments

This work is supported and funded by the National Institute of Allergy and Infectious Disease – Grant Number 1P50AI098461-01 REVISED.

References

1. Bierrenbach AL, de Oliveira GP, Codenotti S, Gomes AB, Stevens AP. Duplicates and misclassification of tuberculosis notification records in Brazil, 2001-2007. *Int J Tuberc Lung Dis* 2010; 14:593-9.
2. Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC, et al. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol* 2008; 168:1326-32.
3. Fonseca MGP, Coeli CM, Lucena FFA, Veloso VG, Carvalho MS. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad Saúde Pública* 2010; 26:1431-8.
4. Migowski A, Chaves RB, Coeli CM, Ribeiro AL, Tura BR, Kuschner MC, et al. Accuracy of probabilistic record linkage in the assessment of high-complexity cardiology procedures. *Rev Saúde Pública* 2011; 45:269-75.
5. de Oliveira GP, Pinheiro RS, Coeli CM, Barreira D, Codenotti SB. Mortality information system for identifying underreported cases of tuberculosis in Brazil. *Rev Bras Epidemiol* 2012; 15:468-77.
6. Balabram D, Turra CM, Gobbi H. Survival of patients with operable breast cancer (Stages I-III) at a Brazilian public hospital: a closer look into cause-specific mortality. *BMC Cancer* 2013; 13:434.
7. Oliveira SB, Merchán-Hamann E, Amorim LDAE. HIV/AIDS coinfection with the hepatitis B and C viruses in Brazil. *Cad Saúde Pública* 2014; 30:433-8.

8. Lynch J, Stuckler D. In God we trust, all others (must) bring data. *Int J Epidemiol* 2012; 41:1503-6.
9. Gonzalez TI, Sabino EC, Capuani L, Liu J, Wright DJ, Walsh JH, et al. Blood transfusion utilization and recipient survival at Hospital das Clinicas in Sao Paulo, Brazil. *Transfusion* 2012; 52:729-38.
10. Sabino EC, Ribeiro AL, Salemi VM, Di Lorenzo Oliveira C, Antunes AP, Menezes MM, et al. Ten-year incidence of Chagas cardiomyopathy among asymptomatic *Trypanosoma cruzi*-seropositive former blood donors. *Circulation* 2013; 127:1105-15.
11. Busch M, Custer B. Health outcomes research using large donor-recipient databases: a new frontier for assessing transfusion safety and contributing to public health. *Vox Sang* 2006; 91:282-4.
12. van Hest NA, Smit F, Baars HW, De Vries G, De Haas PE, Westenend PJ, et al. Completeness of notification of tuberculosis in The Netherlands: how reliable is record-linkage and capture-recapture analysis? *Epidemiol Infect* 2007; 135:1021-9.
13. Fournel I, Schwarzwinger M, Binquet C, Benzenine E, Hill C, Quantin C. Contribution of record linkage to vital status determination in cancer patients. *Stud Health Technol Inform* 2009; 150:91-5.
14. Roberts SE, Williams JG, Meddings D, Goldacre MJ. Incidence and case fatality for acute pancreatitis in England: geographical variation, social deprivation, alcohol consumption and aetiology – a record linkage study. *Aliment Pharmacol Ther* 2008; 28:931-41.
15. Donati S, Senatore S, Ronconi A. Maternal mortality in Italy: a record-linkage study. *BIOG* 2011; 118:872-9.
16. World Health Organization. Assessing tuberculosis under-reporting through inventory studies. Geneva: World Health Organization; 2013.
17. Christen P, Goiser C. Quality and complexity measures for data linkage and deduplication. In: Guillet F, Hamilton HJ, editors. *Quality measures in data mining*. Berlin: Springer Berlin Heidelberg; 2007. p. 127-51.
18. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Delineando a pesquisa clínica: uma abordagem epidemiológica*. 3a Ed. Porto Alegre: Editora Artmed; 2008.
19. Camargo Jr. KR, Coeli CM. *Reclink*: aplicativo para o relacionamento de bases de dados, implementando o método *probabilistic record linkage*. *Cad Saúde Pública* 2000; 16:439-47.
20. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995; 14:491-8.
21. Jorge MH, Laurenti R, Di Nubila HB. Death and its epidemiological investigation: considerations about some relevant aspects. *Rev Bras Epidemiol* 2010; 13:561-76.
22. Franca E, de Abreu DX, Rao C, Lopez AD. Evaluation of cause-of-death statistics for Brazil, 2002-2004. *Int J Epidemiol* 2008; 37:891-901.

Submitted on 17/Feb/2014
 Final version resubmitted on 20/May/2014
 Approved on 04/Jun/2014

Apêndice 2: Tabela 9 Completa: Causas básicas de morte no SIM dos pares verdadeiros por CID 10

Grupo	Soro +	Soro -	CID10	Descrição Causa Básica	TOTAL
	32	3		Capítulo I - Algumas doenças infecciosas e parasitárias	35
4	2	1	A162	Tuberculose pulmonar, sem menção de confirmação bacteriológica ou histológica	3
4	1	2	B207	Doença pelo HIV resultando em infecções múltiplas	3
4	1	0	B208	Doença pelo HIV resultando em outras doenças infecciosas e parasitárias	1
4	1	0	B222	Doença pelo HIV resultando em síndrome de emaciação	1
4	1	0	B569	Tripanossomíase africana não especificada	1
1	22	0	B572	Doença de Chagas (crônica) com comprometimento cardíaco	22
1	3	0	B573	Doença de Chagas (crônica) com comprometimento do aparelho digestivo	3
1	1	0	B574	Doença de Chagas (crônica) com comprometimento do sistema nervoso	1
	22	17		Capítulo II - Neoplasias [tumores]	39
2	1	0	C159	Neoplasia maligna do esôfago, não especificado	1
2	4	2	C169	Neoplasia maligna do estômago, não especificado	6
2	0	1	C189	Neoplasia maligna do cólon, não especificado	1
2	0	1	C20	Neoplasia maligna do reto	1
2	1	0	C250	Neoplasia maligna da cabeça do pâncreas	1
2	4	2	C259	Neoplasia maligna do pâncreas, não especificado	6
2	0	1	C329	Neoplasia maligna da laringe, não especificada	1
2	2	2	C349	Neoplasia maligna dos brônquios ou pulmões, não especificado	4
2	0	2	C439	Melanoma maligno de pele, não especificado	2
2	1	0	C449	Neoplasia maligna da pele, não especificada	1
2	1	3	C509	Neoplasia maligna da mama, não especificada	4
2	2	0	C539	Neoplasia maligna do colo do útero, não especificado	2
2	1	0	C61	Neoplasia maligna da próstata	1
2	0	1	C760	Neoplasia maligna da cabeça, face e pescoço	1
2	1	0	C80	Neoplasia maligna, sem especificação de localização	1
2	1	1	C859	Linfoma não-Hodgkin de tipo não especificado	2
2	1	1	C900	Mieloma múltiplo	2
2	1	0	C97	Neoplasias malignas de localizações múltiplas independentes (primárias)	1
2	1	0	D371	Neoplasia de comportamento incerto ou desconhecido do estômago	1
Continua					

Grupo	Soro +	Soro -	ICD10	Descrição Causa Básica	TOTAL
	0	2		Capítulo IV - Doenças endócrinas, nutricionais e metabólicas	2
4	0	2	E149	Diabetes mellitus não especificado - sem complicações	2
	1	0		Capítulo V - Transtornos mentais e comportamentais	1
4	1	0	F102	Transtornos mentais e comport devidos ao uso de álcool - síndrome de dependência	1
	0	1		Capítulo VI - Doenças do sistema nervoso	1
4	0	1	G062	Abscesso extradural e subdural não especificados	1
	58	20		Capítulo IX - Doenças do aparelho circulatório	78
3	1	0	I10	Hipertensão essencial (primária)	1
3	1	1	I110	Doença cardíaca hipertensiva com insuficiência cardíaca (congestiva)	2
3	3	0	I119	Doença cardíaca hipertensiva sem insuficiência cardíaca (congestiva)	3
3	1	0	I210	Infarto agudo transmural da parede anterior do miocárdio	1
3	23	11	I219	Infarto agudo do miocárdio não especificado	34
3	1	0	I248	Outras formas de doença isquêmica aguda do coração	1
3	2	0	I251	Doença aterosclerótica do coração	2
3	3	0	I259	Doença isquêmica crônica do coração não especificada	3
3	0	1	I340	Insuficiência (da valva) mitral	1
3	5	1	I420	Cardiomiopatia dilatada	6
3	1	0	I429	Cardiomiopatia não especificada	1
3	1	0	I443	Outras formas de bloqueio atrioventricular e as não especificadas	1
3	1	0	I500	Insuficiência cardíaca congestiva	1
3	1	0	I514	Miocardite não especificada	1
3	3	0	I517	Cardiomegalia	3
4	0	1	I608	Outras hemorragias subaracnóides	1
4	2	2	I619	Hemorragia intracerebral não especificada	4
4	0	1	I639	Infarto cerebral não especificado	1
4	2	0	I64	Acidente vascular cerebral, não especificado como hemorrágico ou isquêmico	2
4	2	0	I678	Outras doenças cerebrovasculares especificadas	2
4	1	0	I693	Seqüelas de infarto cerebral	1
4	1	0	I698	Seqüelas de outras doenças cerebrovasculares e das não especificadas	1
4	0	1	I713	Aneurisma da aorta abdominal, roto	1
4	2	0	I718	Aneurisma da aorta de localização não especificada, roto	2
4	1	1	I802	Flebite e tromboflebite de outros vasos profundos dos membros inferiores	2
	9	2		Capítulo X - Doenças do aparelho respiratório	11
4	1	1	J180	Broncopneumonia não especificada	2
4	1	0	J189	Pneumonia não especificada	1
4	1	0	J440	Doença pulmonar obstrutiva crônica com infecção respiratória aguda do trato respiratório inferior	1
4	3	0	J690	Pneumonite devida a alimento ou vômito	3
4	2	0	J81	Edema pulmonar, não especificado de outra forma	2
4	1	1	J90	Derrame pleural não classificado em outra parte	2

Continua

Grupo	Soro +	Soro -	ICD10	Descrição Causa Básica	TOTAL
	8	8		Capítulo XI - Doenças do aparelho digestivo	16
4	0	1	K256	Úlcera gástrica - crônica ou não especificada com hemorragia e perfuração	1
4	0	1	K265	Úlcera duodenal - crônica ou não especificada com perfuração	1
4	0	1	K550	Transtornos vasculares agudos do intestino	1
4	0	1	K559	Transtorno vascular do intestino, sem outra especificação	1
4	1	0	K562	Volvo	1
4	0	1	K650	Peritonite aguda	1
4	0	1	K659	Peritonite, sem outras especificações	1
4	1	0	K703	Cirrose hepática alcoólica	1
4	4	2	K746	Outras formas de cirrose hepática e as não especificadas	6
4	1	0	K810	Colecistite aguda	1
4	1	0	K85	Pancreatite aguda idiopática	1
	2	0		Capítulo XIV - Doenças do aparelho geniturinário	2
4	1	0	N12	Nefrite túbulo-intersticial não especificada se aguda ou crônica	1
4	1	0	N19	Insuficiência renal não especificada	1
	0	1		Capítulo XVII - Malformações congênitas, deformidades e anomalias cromossômicas	1
4	0	1	Q274	Ectasia venosa (flebectasia) congênita	1
Continua					

Grupo	Soro +	Soro -	ICD10	Descrição Causa Básica	TOTAL
	9	8		Capítulo XVIII - Sint., sinais e ach anormais de ex clínicos e de lab, não classif em outra parte	17
6	1	0	R571	Choque hipovolêmico	1
6	0	1	R688	Outros sintomas e sinais gerais especificados	1
6	3	1	R98	Morte sem assistência	4
6	5	6	R99	Outras causas mal definidas e as não especificadas de mortalidade	11
	18	41		Capítulo XX - Causas externas de morbidade e de mortalidade	59
5	0	1	V050	Pedestre traumatizado em colisão com trem [comboio] ou um veículo ferroviário - acidente não-de-tr	1
5	0	3	V093	Pedestre traumatizado em um acidente de trânsito não especificado	3
5	1	2	V274	Motociclista traumatizado colisão com um objeto fixo ou parado - condutor traumat - acidente de trânsito	3
5	0	1	V284	Motociclista traumatizado - acidente de transporte sem colisão - condutor traumat - acidente de trânsito	1
5	1	1	V299	Motociclista [qualquer] traumatizado em um acidente de trânsito não especificado	2
5	2	2	V892	Pessoa traumatizada em um acidente de trânsito com um veículo a motor não especificado	4
5	0	1	V899	Pessoa traumatizada em um acidente com um veículo não especificado	1
5	0	1	W119	Queda em ou de escadas de mão - local não especificado	1
5	1	1	W199	Queda sem especificação - local não especificado	2
5	1	0	W749	Afogamento e submersão não especificados - local não especificado	1
5	0	1	X094	Exposição a tipo não especificado de fumaça, fogo ou chamas - rua e estrada	1
5	1	0	X680	Auto-intoxicação por e exposição, intencional, a pesticidas - residência	1
5	1	1	X700	Lesão autoprovocada intencionalmente por enforcamento, estrangulamento e sufocação - residência	2
5	1	0	X709	Lesão autoprovocada intenci por enforc, estrang e sufocação - local não especificado	1
5	0	1	X934	Agressão por meio de disparo de arma de fogo de mão - rua e estrada	1
5	0	1	X950	Agressão por meio de disparo de outra arma de fogo ou de arma não especificada - residência	1
5	2	6	X954	Agressão por meio de disparo de outra arma de fogo ou de arma não especific - rua e estrada	8
5	1	5	X959	Agressão por meio de disparo de outra arma de fogo ou de arma não especific - local não especific	6
5	0	1	X990	Agressão por meio de objeto cortante ou penetrante - residência	1
5	1	0	X994	Agressão por meio de objeto cortante ou penetrante - rua e estrada	1
5	1	1	Y004	Agressão por meio de um objeto contundente - rua e estrada	2
5	0	1	Y008	Agressão por meio de um objeto contundente - outros locais especificados	1
5	0	1	Y049	Agressão por meio de força corporal - local não especificado	1
5	0	1	Y079	Outras síndromes de maus tratos por pessoa não especificada	1
5	0	1	Y090	Agressão por meios não especificados - residência	1
5	1	2	Y094	Agressão por meios não especificados - rua e estrada	3
5	1	1	Y099	Agressão por meios não especificados - local não especificado	2
5	0	1	Y199	Envenenamento por e expos a prod quím e subs nocivas e aos não especific, int não det - local não e	1
5	0	1	Y299	Contato com objeto contundente, intenção não determinada - local não especificado	1
6	0	1	Y340	Fatos ou eventos não especificados e intenção não determinada - residência	1
6	0	1	Y344	Fatos ou eventos não especificados e intenção não determinada - rua e estrada	1
6	2	0	Y349	Fatos ou eventos não especificados e intenção não determinada - local não especific	2
TOTAL	159	103			262

Apêndice 3: Autorização do uso dos bancos de dados utilizados neste estudo

PROJETO DE PESQUISA

Título: Mortalidade entre doadores de sangue soropositivos para Doença de Chagas (1996-2000) em São Paulo: um estudo de relacionamento probabilístico de dados.

Área Temática:

Pesquisador: Ester cerdeira Sabino

Versão: 1

Instituição: Fundação Faculdade de Medicina - ICESP

CAAE: 03882512.4.0000.0065

PARECER CONSUBSTANCIADO DO CEP

Número do Parecer: 70440

Data da Relatoria: 08/08/2012

Apresentação do Projeto:

Estudo coorte, retrospectivo, para obtenção do título de doutorado, para avaliar a morbi-mortalidade entre doadores de sangue soropositivos para Doença de Chagas (1996-2000) em São Paulo: um estudo de relacionamento probabilístico de dados. A introdução é bem detalhada, apresenta justificativa e embasamento à proposta do estudo. A metodologia apresenta clareza. A bibliografia é atualizada, completa e pertinente ao estudo. O TCLE não se aplica, pois trata-se de uma análise de banco de dados.

Objetivo da Pesquisa:

A descrição dos objetivos é suficiente e está adequada ao título.

Avaliação dos Riscos e Benefícios:

Riscos: O risco potencial para os sujeitos é mínimo, uma vez que as informações serão provenientes de base de dados já existentes. O único risco seria a perda acidental de confidencialidade referente ao status de óbito e causa de morte.

Benefícios: Conhecer a taxa de mortalidade e causas de morte entre os doadores de sangue soropositivos para a Doença de Chagas permitirá o aprimoramento de políticas de saúde pública e o melhor aconselhamento dos doadores.

Comentários e Considerações sobre a Pesquisa:

A produção científica do pesquisador capacita-o para orientar e desenvolver o estudo. O projeto foi bem escrito e é exequível. Descreve o plano de trabalho e o cronograma da análise estatística a ser utilizada.

Considerações sobre os Termos de apresentação obrigatória:

Nada consta.

Recomendações:

Nada consta.

Conclusões ou Pendências e Lista de Inadequações:

Aprovado.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Aprovado

SAO PAULO, 09 de Agosto de 2012

Assinado por:
Roger Chammas

São Paulo, 02 de julho de 2012.

Parecer apresentado à reunião da Comissão de Ética e Pesquisa do Depto de Moléstias Infecciosas e Parasitárias em 02/07/2012.

I - IDENTIFICAÇÃO

- **Título do Projeto:** “Mortalidade entre doadores de sangue soropositivos para Doença de Chagas (1996-2000) em São Paulo: um estudo de relacionamento probabilístico de dados.”

- **Pesquisador responsável:** Profa Dra Ester Cerdeira Sabino.

- **Pesquisador executante:** Dra Maria Ligia Damato Capuani.

- **Instituição onde será realizada a pesquisa:** Departamento de Moléstias Infecciosas e Parasitárias da Faculdade de Medicina da Universidade de São Paulo e Fundação Pró-Sangue HC-FMUSP.

- **Data de apresentação à Comissão de Ética e Pesquisa:**

06/06/2012

II – OBJETIVOS

Definir a taxa de mortalidade para doadores positivos para doença de Chagas em um período de 10 a 15 anos, através de consulta aos prontuários dos doadores e do registro de óbitos (SIM) do MS.

III - SUMÁRIO DO PROJETO:

III. 1 Descrição e caracterização da amostra:

Estudo de coorte retrospectivo de doadores soropositivos para a doença de Chagas atendidos na Fundação Pró-Sangue entre 1996 e 2000, e doadores soronegativos no mesmo período.

Critérios de inclusão:

Doadores soropositivos para a doença de Chagas atendidos na Fundação Pró-Sangue entre 1996 e 2000 e doadores soronegativos na proporção de 1:2.

Critérios de exclusão:

Não se aplica.

• **Adequação da metodologia**

A metodologia do estudo é adequada.

• **Adequação das condições**

A equipe proponente conta com infra-estrutura capacitada para realizar o estudo.

IV - OBSERVÂNCIA DA RESOLUÇÃO 196/96

• **Estrutura do protocolo**

O protocolo está bem estruturado.

• **Justificativa de uso de placebo:**

Não se aplica.

• **Justificativa de suspensão terapêutica (Wash-out):**

Não se aplica.

• **Análise de riscos e benefícios:**

Riscos: Não há riscos envolvidos neste estudo.

• **Retorno de benefícios para o sujeito e/ou para a comunidade:**

Não se aplica.

• **Adequação do Termo de Consentimento Livre e Esclarecido e forma de obtê-lo:**

Não há necessidade de TCLE.

FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO
DEPARTAMENTO DE MOLÉSTIAS INFECCIOSAS E PARASITÁRIAS

• **Informação adequada quanto ao financiamento:**

Não há necessidade de financiamento.

• **Outros centros, no caso de estudos multicêntricos**

V – PARECER FINAL:

Aprovado.

A Comissão aprovou o presente projeto com base no parecer do relator.

Profª Maria Aparecida Shikanai Yasuda
Coordenadora da Comissão Ética e Pesquisa
Depto de Moléstias Infecciosas e Parasitárias
da FMUSP

Memo.CAPPesq – 164/09

São Paulo, 12 de janeiro de 2009.

Ilma.Sra.
Dra. Ester Cerdeira Sabino
DD Pesquisador
Departamento Clínico Médica

Por determinação do Senhor Presidente, Prof. Dr. Eduardo Massad, encaminhamos para conhecimento, cópia do parecer nº **1006/2008**, da Comissão Nacional de Ética em Pesquisa – CONEP/MS referente à **APROVAÇÃO** do Protocolo de Pesquisa nº **0600/08** – intitulado **“Taansfusao de Sangue na cidade de Sao PAulo: uma análise da utilização do sangue em pacientes sobreviventes no Hospital da Clínicas”**.

Atenciosamente,

Julio Cesar
Secretária
Comissão de Ética para Análise de Projetos
de Pesquisa - CAPPesq

MINISTÉRIO DA SAÚDE
Conselho Nacional de Saúde
Comissão Nacional de Ética em Pesquisa - CONEP

PARECER Nº 1006/2008

Registro CONEP: 15099 (Este nº deve ser citado nas correspondências referentes a este projeto)

CAAE – 0543.0.015.000-08

Processo nº 25000.180664/2008-30

Projeto de Pesquisa: *"Transfusão de sangue na cidade de São Paulo, Brasil: uma análise da utilização do sangue e pacientes sobreviventes no Hospital das Clínicas"*

Pesquisador Responsável: Ester Cerdeira Sabino

Instituição: Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

CEP de origem: Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

Área Temática Especial: Cooperação estrangeira

Patrocinador: Fundação Faculdade de Medicina

Sumário geral do protocolo

No Brasil se conhece pouco sobre a utilização do sangue e as características dos receptores incluindo as indicações e a intensidade das transfusões realizadas.

Os objetivos do estudo são: 1) descrever os padrões de utilização de sangue no Hospital das Clínicas por meio de características demográficas dos receptores e respectivas indicações clínicas, baseadas no sistema informatizado do banco de sangue e nos registros do hospital; 2) determinar a taxa de sobrevida e alta hospitalar dos receptores e realizar o seguimento destes receptores no período de um ano, utilizando o sistema informatizado de mortalidade brasileira.

O estudo é retrospectivo transversal baseado em dados e registros informatizados relativos a todas as transfusões realizadas no complexo hospitalar do hospital das clínicas, durante o ano de 2004.

Os critérios de inclusão e exclusão estão descritos à página 27 do protocolo (numeração do CEP).

Os dados serão procedentes de dois bancos de dados já existentes. Os dados serão obtidos de registros informatizados que fazem parte da base de dados do sistema de informática do HC e da FPS/HSP. Haverá a compilação das informações relativas aos dados do perfil demográfico e transfusional de cada um dos receptores durante no ano de 2004 no HC.

O primeiro banco de dados será procedente da FPS/HSP irá incluir os seguintes elementos: número do registro hospitalar HC; número do registro do paciente na FPS/HSP; sexo; data de nascimento; data da transfusão; número do sangue e ou hemocomponente(s) transfundido(s); e tipo de componente (concentrado de hemácias, plaquetas) incluindo tipagem sanguínea e Rh. Utilizando o número do registro hospitalar do paciente do HC poderá ser associado as informações contidas nos registros médicos, que forneceram os seguintes elementos adicionais: código de diagnóstico; clínica ou divisão; resultado da alta ou óbito do paciente; data do óbito; e data da alta hospitalar. Para os pacientes que tiveram alta hospitalar, os pesquisadores irão cruzar os dados destes pacientes com os dados do banco de mortalidade do Ministério da Saúde.

Os bancos de dados serão compilados entre si por especialistas em informática que irão criar um novo banco de dados no qual o nome do paciente será retirado antes que seja enviado para a análise dos pesquisadores. Esse novo arquivo manterá o código numérico e os especialistas em informática poderão atualizar as informações, contudo, "os pesquisadores nunca terão acesso a esse código e ao nome do paciente" (página 28, numeração do CEP).

Será realizada pesquisa para identificar os óbitos que ocorreram nos 3 anos após a transfusão (até dezembro de 2006). Um arquivo indicando a data e a causa do óbito será combinado com o número do registro hospitalar do paciente. Esse arquivo será encaminhado

Cont. Parecer CONEP 1006/2008.

para os pesquisadores americanos para compilação, atualização e posterior análise do banco de dados. Com esta informação os pesquisadores americanos poderão determinar o status de 1 ano de sobrevivência após transfusão.

A pesquisadora informa que "em nenhum momento da pesquisa será liberado o nome do receptor ou do doador relacionado com a bolsa transfundida. O único momento em que haverá correlação entre a identificação dos receptores será quando iremos combinar os dados de mortalidade do Ministério da Saúde com os dados dos óbitos ocorridos no Hospital das Clínicas nos pacientes transfundidos. No entanto, ressaltamos que todos os dados serão mantidos em sigilo, em equipamentos de informática com acesso restrito".

A pesquisadora descreve que o benefício será a identificação de maneiras de aperfeiçoar as indicações de transfusão de sangue.

Local de realização

A folha de rosto informa que o número de sujeitos é de 30.000, que são todos os pacientes que receberam transfusão de sangue durante o período da pesquisa são elegíveis para participar do estudo.

O cronograma de execução do estudo informa que o início da coleta dos dados será em março de 2009.

Apresentação do protocolo

Os gastos previstos são da ordem de US 7200 (R\$ 13.000,00). Este recurso será repassado pelo Blood Systems Research Institute à Fundação Faculdade de Medicina. O orçamento financeiro está descrito à página 36 (numeração do CEP).

Consta no protocolo, à página 31, numeração do CEP, que "em se tratando de análise de banco de dados, não haverá necessidade de consentimento informado dos receptores". Há, ainda, um documento assinado pela pesquisadora informando que "por se tratar de uma revisão de prontuário de um grande número de casos, não poderemos solicitar um consentimento informado aos receptores de sangue".

Os currículos dos pesquisadores foram apresentados.

O protocolo de estudo foi classificado como cooperação estrangeira porque o grupo de pesquisadores norte-americanos será responsável pela análise do banco de dados, suporte logístico e estatístico, além do aporte financeiro. Consta um documento do Blood Systems Research Institute informando a referida cooperação estrangeira informando que o Instituto tem uma longa parceria com Ester C. Sabino e a Fundação Pró-Sangue na condução de estudos científicos de pesquisa sobre a segurança do sangue e medicina transfusional a qual inclui um histórico forte em publicações científicas".

Diante do exposto, a Comissão Nacional de Ética em Pesquisa - CONEP, de acordo com as atribuições definidas na Res. CNS 196/96, manifesta-se pela aprovação do projeto de pesquisa proposto.

Situação: Protocolo aprovado.

Brasília, 18 de dezembro de 2008.

Sybil Saddi Tannous
Coordenadora da CONEP/CNS/MS

SOLICITAÇÃO DE PESQUISA

DIAGNOSTICOS (somente pacientes internados):

histograma dos pacientes que foram a óbito na
clínica de hematologia
hematoma do Nome completo e data de
nascimento
se possível NOME COMPLETO DA MÃE

PERÍODO: 01 / 01 / 2003 a 31 / 12 / 2005
(dd) (mm) (aa) (dd) (mm) (aa)

CLÍNICAS Hematologia

VARIÁVEIS DE INTERESSE: ☒ IDADE ☒ SEXO ☐ COR
☐ PROFISSÃO ☐ DATA DE INTERNAÇÃO ☐ DATA DE SAÍDA
☐ CONDIÇÕES DE SAÍDA

PACIENTES ATENDIDOS NO AMBULATÓRIO

CLÍNICA: Unidades:

CLÍNICA: Unidades:

PERÍODO: / / a / /
(dd) (mm) (aa) (dd) (mm) (aa)

VARIÁVEIS DE INTERESSE: ☐ IDADE ☐ SEXO ☐ COR
☐ PROFISSÃO ☐ APENAS NÚMERO DE PACIENTES

OBSERVAÇÕES:

NOME: ESTER CERDEIRA SABINO Data: 16 / 08 / 2007
Assinatura: ESTER C. SABINO de Biologia Molecular FPG-USP Telefone: 30615544 2299
Professor Responsável: Dalton Chamone

MINISTÉRIO DA SAÚDE
SECRETARIA DE VIGILÂNCIA EM SAÚDE

**TERMO DE RESPONSABILIDADE DIANTE DA CESSÃO DAS BASES DE DADOS
NOMINAIS DE SISTEMAS DE INFORMAÇÃO GERENCIADOS PELA SECRETARIA DE
VIGILÂNCIA EM SAÚDE**

Pelo presente instrumento, na qualidade de responsáveis pela guarda e uso da(s) base(s) de dados do SIM, assumimos as seguintes responsabilidades;

- a) Utilizar esta(s) bases de dados única e exclusivamente para as finalidades descritas ao final deste documento;
- b) Guardar sigilo e zelar pela privacidade dos indivíduos relacionados/listados nesta base de dados;
- c) Não disponibilizar, emprestar ou permitir a pessoas ou instituições não autorizadas pela Secretaria de Vigilância em Saúde do Ministério da Saúde o acesso à esta(s) base(s) de dados;
- d) Não divulgar, por qualquer meio de divulgação, dados ou informações contendo o nome dos indivíduos ou outras variáveis que permitam a identificação do indivíduo e que afetem assim a confidencialidade dos dados contidos nesta(s) base(s) de dados;
- e) Não praticar ou permitir qualquer ação que comprometa a integridade desta(s) base(s) de dados;
- f) Não utilizar isoladamente as informações contidas nesta base de dados para tomar decisões sobre a identidade de pessoas falecidas/nascidas, para fins de suspensão de benefícios ou outros tipos de atos punitivos, sem a devida certificação desta identidade em outras fontes.

Desta forma, o(a) **Fundação Pró-Sangue – Hemocentro de São Paulo** assume total responsabilidade pelas consequências legais pela utilização indevida desta(s) bases de dados, por parte de servidores desta instituição ou por terceiros.

Base	Anos e abrangência (UF)
(X) SIM	1996 a 2009
() Sinasc	
() SINAN	
() _____	

Declaramos que esta(s) base(s) de dados será(ão) usada(s) única e exclusivamente para as seguintes finalidades:

1) Definir a sobrevida de doadores de sangue da Fundação Pro-Sangue, Hemominas e HEMOPE positivos para a doença de Chagas.

2) Avaliar a História Natural da doença de Chagas. Os bancos de sangue das instituições acima citados possuem cadastro completo dos seus doadores desde 1994 e estão trabalhando juntos em um projeto Internacional de Segurança transfusional. Este projeto prevê o desenvolvimento de um único data warehouse com dados dos três hemocentro que serão armazenados junto ao Laboratório de Banco de Dados do Departamento de Computação do Instituto de Matemática e estatística da USP. Para este estudo é importante sabermos quais doadores já morreram no período estudado.

3) Definir a sobrevida de receptores de sangue do Hospital das Clínicas de São Paulo.

São Paulo, 27 de agosto de 2010.

Técnico(s) Responsável(is) pelo uso e guarda da(s) base(s) de dados solicitada(s):

Nome: Ester Cerdeira Sabino

RG : __5223221-9 SSP-SP__ CPF: _075228968-35

Assinatura: _____

Dr. ESTER C. SABINO - CRM 51785

Instituição: Fundação Pró-Sangue – Hemocentro de São Paulo.

Responsável legal (nome): Alfredo Mendrone Jr.

RG: __9195478-2__ CPF: _084723508-43

Assinatura: _____

Dr. Alfredo Mendrone Junior
CRM 50264
Diretor Técnico Científico
Fundação Pró-Sangue / HSP

MINISTÉRIO DA SAÚDE
SECRETARIA DE VIGILÂNCIA EM SAÚDE

**TERMO DE RESPONSABILIDADE DIANTE DA CESSÃO DAS BASES DE DADOS
NOMINAIS DE SISTEMAS DE INFORMAÇÃO GERENCIADOS PELA SECRETARIA DE
VIGILÂNCIA EM SAÚDE**

Pelo presente instrumento, na qualidade de responsáveis pela guarda e uso da(s) base(s) de dados do SIM, assumimos as seguintes responsabilidades;

- a) Utilizar esta(s) bases de dados única e exclusivamente para as finalidades descritas ao final deste documento;
- b) Guardar sigilo e zelar pela privacidade dos indivíduos relacionados/listados nesta base de dados;
- c) Não disponibilizar, emprestar ou permitir a pessoas ou instituições não autorizadas pela Secretaria de Vigilância em Saúde do Ministério da Saúde o acesso à esta(s) base(s) de dados;
- d) Não divulgar, por qualquer meio de divulgação, dados ou informações contendo o nome dos indivíduos ou outras variáveis que permitam a identificação do indivíduo e que afetem assim a confidencialidade dos dados contidos nesta(s) base(s) de dados;
- e) Não praticar ou permitir qualquer ação que comprometa a integridade desta(s) base(s) de dados;
- f) Não utilizar isoladamente as informações contidas nesta base de dados para tomar decisões sobre a identidade de pessoas falecidas/nascidas, para fins de suspensão de benefícios ou outros tipos de atos punitivos, sem a devida certificação desta identidade em outras fontes.

Desta forma, o(a) **Departamento de Computação do Instituto de Matemática e Estatística da Universidade de São Paulo** assume total responsabilidade pelas consequências legais pela utilização indevida desta(s) bases de dados, por parte de servidores desta instituição ou por terceiros.

Base	Anos e abrangência (UF)
(X) SIM	1996 a 2006
() Sinasc	
() SINAN	

Declaramos que esta(s) base(s) de dados será(ão) usada(s) única e exclusivamente para as seguintes finalidades:

1) Definir a sobrevida de doadores de sangue da Fundação Pro-Sangue, Hemominas e HEMOPE positivos para a doença de Chagas.

2) Avaliar a História Natural da doença de Chagas. Os bancos de sangue das instituições acima citados possuem cadastro completo dos seus doadores desde 1994 e estão trabalhando juntos em um projeto Internacional de Segurança transfusional. Este projeto prevê o desenvolvimento de um único data warehouse com dados dos três hemocentro que serão armazenados junto ao Laboratório de Banco de Dados do Departamento de Computação do Instituto de Matemática e estatística da USP. Para este estudo é importante sabermos quais doadores já morreram no período estudado.

São Paulo, 13 de fevereiro de 2007.

Técnico(s) Responsável(is) pelo uso e guarda da(s) base(s) de dados solicitada(s):

Nome: Ester Cerdeira Sabino

RG : 5.223.221-9 SSP-SP CPF: 075.228.968-35

Assinatura:

Nome: João Eduardo Ferreira

RG : 13.814.187 CPF 041.577.558-20

Assinatura:

Instituição:

Responsável legal (nome): Routo Terada

RG: 3.726.749 CPF: 387.789.128-49

Assinatura: