

Universidade de São Paulo
Faculdade de Saúde Pública

**Uso da técnica de *linkage* nos sistemas de informação
em saúde: aplicação na base de dados do Registro de
Câncer de Base Populacional do
Município de São Paulo**

Stela Verzinhasse Peres

**Tese apresentada ao Programa de Pós-
Graduação em Saúde Pública para
obtenção do título de Doutor em Saúde
Pública.**

Área de Concentração: Epidemiologia
**Orientadora: Profa. Dra. Maria do
Rosário Dias de Oliveira Latorre.**
**Co-Orientadora: Profa. Dra. Márcia
Furquim de Almeida**

São Paulo
2011

Uso da técnica de *linkage* nos sistemas de informação em saúde: aplicação na base de dados do Registro de Câncer de Base Populacional do Município de São Paulo

Stela Verzinhasse Peres

Tese apresentada ao Programa de Pós-Graduação em Saúde Pública da Faculdade de Saúde Pública da Universidade de São Paulo para obtenção do título de Doutor em Saúde Pública.

**Área de Concentração: Epidemiologia
Orientadora: Profa. Dra. Maria do Rosário Dias de Oliveira Latorre.
Co-Orientadora: Profa. Dra. Márcia Furquim de Almeida**

São Paulo

2011

É expressamente proibida a comercialização deste documento, tanto na sua forma impressa como eletrônica. Sua reprodução total ou parcial é permitida **exclusivamente** para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da tese/dissertação.

Aos meus pais

Marlene Verzinhasse Peres e Celoni Peres

Agradecimentos

À Prof^a Dr^a Maria do Rosário Dias de Oliveira Latorre pelos grandes ensinamentos ao longo destes anos. Agradeço pelo carinho, atenção, confiança e por me impulsionar na carreira profissional.

À Prof^a Dr^a Márcia Furquim de Almeida pela preciosa co-orientação e pelas discussões enriquecedoras para a finalização desta tese.

Aos membros da banca pelas valiosas contribuições no desenvolvimento e conclusão desta pesquisa, Prof^a Dr^a Claudia Medina Coeli, Prof^a Dr^a Bernadette Cunha Waldvogel, Prof^a Dr^a Zilda Pereira da Silva e Prof^a Dr^a Karina de Cássia Braga Ribeiro.

Ao Dr. Mauro Tanigushi e à Dr^a Margarida Lira pela confiança e presteza na disponibilidade das bases de dados.

À Monica La Porte pela ajuda inestimável para a finalização deste trabalho.

A todos os Professores do Departamento de Epidemiologia da Faculdade de Saúde Pública, meus eternos mestres!

Às "meninas insuperáveis" da Secretaria de Pós-Graduação, Renilda Maria Figueiredo Shimono, Vânia Santos, Maria Aparecida Mendes, Alessandra B. Frutuoso e Ângela Maria P.S. de Andrade. Muito obrigada pelo carinho, compreensão, competência, respeito e dedicação... adoro vocês!

À Elizabete Margarida Nassar Ribeiro, por toda dedicação e competência.

Aos meninos da informática, Daniel, Rodrigo e Pascoal por resolverem todas as "panes" dos computadores e pela amizade.

À equipe do Registro de Câncer de Base Populacional de São Paulo, Aryane, Thamires, Iza, Lucinda, Rita e Nazaré, pela colaboração em todas as etapas desta tese, pela amizade e por muitos momentos de descontração.

A Carolina Terra Luizaga e Vinicius Leati, pela amizade e ajuda incansável na elaboração dos bancos de dados.

Aos amigos da Área Técnica e do Núcleo de Indicadores e Sistemas de Informação do Instituto Israelita de Responsabilidade Social do Hospital Albert Einstein. Especialmente à Alessandra Cristina F. Martins, por compreender meus dias de sono,

minhas ausências e, ainda assim, sempre me incentivar para a conclusão deste trabalho.

Aos grandes amigos que fiz na Faculdade de Saúde Pública, Neuber, Shamy, Sofia, Suzana, Carmen, Luciana, Fernanda Cruciani, Débora, Lizis, Thaís, Maria, Michel e Tati.

Às melhores amigas, Aline, Luana e Fernanda, por fazerem parte da minha trajetória, pois sem vocês seria impossível terminar esta tese. Muito obrigada pelas conversas sérias e educativas, muito obrigada pelas besteiras e risadas... enfim, por estarem aqui!

À Milene Mitsuyuki Foschini, minha grande amiga, sempre tirando as minhas dúvidas e me fazendo companhia nas horas mais difíceis. Ao seu marido Maurício Foschini, por me aturar.

À Dona Teresa e à Silvana, por me acolherem como filha em Barretos. Agradeço pelo carinho, amizade e almoços maravilhosos.

Aos meus queridos irmãos, Jane, Patrícia e Ulisses. Aos meus cunhados, Rui, Domingos e Cristine. Aos meus sobrinhos, Marcello, Isabella, Bárbara, Graziella, José, Guilherme e Renato. Aos meus cachorros Veron e Pequena. Obrigada pela força e por compreenderem minha ausência.

Aos meus pais, Marlene e Celoni, por serem os melhores pais do mundo!

"O que pode mudar seu pensamento, pode mudar seu destino"

Stephen Covey

Resumo

PERES,SV. Uso da técnica de *linkage* nos sistemas de informação em saúde: aplicação na base de dados do Registro de Câncer de Base Populacional do Município de São Paulo [tese de doutorado]. São Paulo: Faculdade de Saúde Pública da USP; 2011.

Introdução- A disponibilidade de grandes bases de dados informatizadas em saúde tornou a técnica de relacionamento de fontes de dados, também conhecida como *linkage*, uma alternativa para diferentes tipos de estudos. Esta técnica proporciona a geração de uma base de dados mais completa e de baixo custo operacional.

Objetivo- Investigar a possibilidade de completar/aperfeiçoar as informações da base de dados do RCBP-SP, no período de 1997 a 2005, utilizando o processo de *linkage* com três outras bases, a saber: Programa de Aprimoramento de Mortalidade (PRO-AIM), Autorização e Procedimentos de Alta Complexidade (APAC-SIA/SUS) e Fundação Sistema Estadual de Análise de Dados (FSEADE).

Métodos- Neste estudo foi utilizada a base de dados do RCBP-SP, composta por 343.306 com casos incidentes de câncer do Município de São Paulo, registrados no período de 1997 a 2005, com idades que variaram de menos de um a 106 anos, de ambos os sexos. Para a completude das informações do RCBP-SP foram utilizadas as bases de dados, a saber: PRO-AIM, APAC-SIA/SUS e FSeade. Foram utilizadas as técnicas de *linkage* probabilística e determinística. O *linkage* probabilístico foi realizado pelo programa ReLink III versão 3.1.6. Quanto ao *linkage* determinístico as rotinas foram realizadas em *Visual Basic*, com as bases hospedadas em *SQL Server*. Foram calculados os coeficientes brutos de incidência (CBI) e mortalidade (CBM) antes e após o *linkage*. A análise de sobrevida global foi realizada pela técnica de Kaplan-Meier e para na comparação entre as curvas, utilizou-se o teste de *log rank*. Foram calculados os valores da área sob a curva, sensibilidade e especificidade para determinar o ponto de corte do escore de maior precisão na identificação dos pares verdadeiros.

Resultados- Após o *linkage*, verificou-se um ganho de 101,5% para a variável endereço e 31,5% para a data do óbito e 80,0% para a data da última informação. Quanto à variável nome da mãe, na base de dados do RCBP-SP antes do *linkage* esta informação representava somente 0,5%, tendo sido complementada, no geral, em 76.332 registros. A análise de sobrevida global mostrou que antes do processo de *linkage* havia uma subestimação na probabilidade de estar vivo em todos os períodos analisados. No geral, para a análise de sobrevida truncada em sete anos, a probabilidade de estar vivo no primeiro ano de seguimento antes do *linkage* foi menor quando comparada a probabilidade de estar vivo ao primeiro ano de seguimento após o *linkage* (48,8% x 61,1%; $p < 0,001$).

Conclusão- A técnica de *linkage* tanto probabilística quanto determinística foi efetiva para completar/aperfeiçoar as informações da base de dados do RCBP-SP. Além do mais, o CBI apresentou um ganho de 4,3%. Quanto ao CBM houve um ganho de 25,8%. Após o uso da técnica de *linkage*, foi verificado que os valores para a sobrevida global estavam subestimados para ambos os sexos, faixas etárias e para todas as topografias de câncer.

Descritores: Sistemas de Informação em Saúde; Registro de Câncer de Base Populacional; Qualidade da Informação; *Linkage* Probabilístico; *Linkage* Determinístico.

Abstract

PERES, SV. The use of the linkage technique for health information systems: application in the database of the São Paulo Population-based Cancer Registry [doctoral thesis]. São Paulo: Faculdade de Saúde Pública da USP; 2011.

Introduction – The availability of large computerized databases on health has turned the record linkage technique, into an alternative for different study designs. This technique provides the generation of a more complete database, at low operational cost. **Objective** – to investigate the possibility of completing/improving information from the database of the RCBP-SP, in the period between 1997 and 2005, using the record linkage technique with other three databases, namely: Mortality Improvement Program (PRO-AIM), Authorization of Highly Complex Procedures (APAC-SIA/SUS) and State System of Data Analysis (FSEADE), comparing different strategies. **Methods** – In this study we used the database of the RCBP-SP composed of 343,306 incident cancer cases in the Municipality of São Paulo registered in the period between 1997 and 2005 with ages ranging from under one to 106 years, of both sexes. To complete the database of the RCBP-SP three databases were used, namely: PRO-AIM, APAC-SIA/SUS and FSEADE. Both probabilistic and deterministic record linkage were used. Probabilistic linkage was performed using the Reclink III software, version 3.1.6. As for the deterministic record linkage, the routines were run in the Visual Basic and databases were hosted on a SQL Server. Before and after record linkage, crude incidence (CIR) and mortality rates (CMR) were calculated. The overall survival analysis was performed using the Kaplan-Meier technique and for the comparison between curves, the log rank test was employed. In order to determine the most precise cut-off scores in identifying true matches, we calculated the area under the curve, as well as, sensitivity and specificity.

Results – After record linkage, it was verified a gain of 101.5% for the variable address, 31.5% for death date and 80,0% for date of latest information. As for the variable mother's name, in the database of the RCBP-SP before record linkage, this information represented only 0.5%, having been completed, in general, in 76,332 registries. The overall survival analysis showed that before the record linkage there was an underestimation of the probability of being alive for all periods assessed. In general, for the truncated survival at seven years, the probability of being alive at the first year of follow up before record linkage was lower when compared to the probability of being alive at the first year of follow up after record linkage (48.8% x 61.1%; $p < 0.001$). **Conclusion** – Both the probabilistic and deterministic record linkage were effective to complete/improve information from the database of the RCBP-SP. Moreover, the CIR had a gain of 4,3%. As for the CMR, there was a gain of 25.8%. After using the record linkage technique, it was verified that values for overall survival were underestimated for both sexes, all age groups, and cancer sites.

Descriptors: Health Information Systems; Population-based Cancer Registry; Information Quality; Probabilistic Record Linkage; Deterministic Record Linkage.

ÍNDICE

1. INTRODUÇÃO	1
1.1 Sistemas de Informação em Saúde	1
1.2 Sistemas de Informação em Saúde no Brasil	2
1.2.1 Sistemas de Informação em Saúde em São Paulo	7
1.3 Relacionamento de bases de dados em Saúde	9
1.4 Acurácia da técnica de <i>linkage</i>	16
1.5 Registro de Câncer de Base Populacional (RCBP)	21
1.5.1 O Registro de Câncer de Base Populacional em São Paulo (RCBP-SP)	22
2. OBJETIVOS	25
3. MATERIAL E MÉTODOS	26
3.1 Base de dados do Registro de Câncer de Base Populacional do Município de São Paulo (RCBP - SP)	26
3.2 Bases de dados utilizadas no processo de <i>linkage</i>	26
3.3 Preparação das bases de dados	27
3.4 Padronização	30
3.4.1 Variável: sexo	33
3.4.2 Variável: data de nascimento	34
3.4.3 Variável: ano de nascimento	34
3.4.4 Variável: data do óbito	35
3.4.5 Variável: Nome do paciente	35
3.4.6 Rotinas	37
3.5 A execução do <i>linkage</i> probabilístico	37
3.5.1 Blocagem	37
3.5.2 Pareamento	38
3.6 A execução do <i>linkage</i> determinístico	49
3.7 Análise Estatística	53
3.8 Considerações Éticas	54

4. RESULTADOS	55
4.1 Resultados dos relacionamentos	55
4.2 Completitude da informação	61
4.3 Coeficientes brutos de incidência (CBI) e de mortalidade (CBM)	64
4.4 Probabilidade de sobrevida global acumulada antes e após o <i>linkage</i>	68
4.5 Proposta de pontos de corte nos escores calculados nos processos de <i>linkage</i> probabilístico	84
5. DISCUSSÃO	88
6. CONCLUSÕES	109
7. REFERÊNCIAS	110
ANEXOS	123
Anexo 1 – Declaração de Óbito (DO)	124
Anexo 2 – Ficha de Solicitação da APAC-SIA/SUS	125
Anexo 3 – Ficha de Notificação (RCBP-SP)	126
Anexo 4 – Tabela das 24 sub-bases de dados	127
Anexo 5 – Termo de Cooperação Técnica entre a FSP/USP e a FSEADE	129
Anexo 6 – Rotinas	133
Anexo 7 – Quadros dos <i>linkages</i> entre o RCBP-SP <i>versus</i> PRO-AIM e Apac-SIA/SUS	135
Anexo 8 – Aprovação do Comitê de Ética em Pesquisa (Faculdade de Saúde Pública/USP)	159
Anexo 9 – Aprovação do Comitê de Ética em Pesquisa (Secretaria Municipal da Saúde de São Paulo)	160
CURRÍCULO LATTES	

Lista de Quadros

Quadro 1	Padronizações segundo bases de dados, RCBP-SP, PRO-AIM e APAC-SIA/SUS.	32
Quadro 2	Padronizações da base de dados do RCBP-SP para sua vinculação à base de dados de óbitos da FSEADE.	33
Quadro 3	Estratégias de Blocação.	38
Quadro 4	Parâmetros para a construção dos escores de pareamento.	40
Quadro 5	Fatores de ponderação de concordância e discordância.	41
Quadro 6	Descrição dos campos de blocação e dos campos de pareamento e os escores máximos e mínimos calculados.	42
Quadro 7	RCBP-SP <i>versus</i> PRO-AIM para a topografia C18.0 a C18.9 (colón). RCBP-SP, 1997 a 2005.	46
Quadro 8	Estratégias para o processo de <i>linkage</i> probabilístico, segundo desempenho no relacionamento entre RCBP-SP <i>versus</i> PRO-AIM.	48
Quadro 9	Estratégias para o processo de <i>linkage</i> probabilístico, segundo desempenho no relacionamento entre RCBP-SP <i>versus</i> APAC-SIA/SUS.	49
Quadro 10	Critérios elaborados para o <i>linkage</i> entre a base de dados do RCBP-SP e a FSEADE.	51

Lista de Tabelas

Tabela 1	Número de registros vinculados e de pares verdadeiros entre o <i>linkage</i> da base de dados do RCBP-SP e a base de dados da FSeade, segundo ano de ocorrência do óbito.	52
Tabela 2	População do Município de São Paulo, entre os anos de 1997 a 2005.	53
Tabela 3	Número e percentual de pacientes com informação não ignorada, segundo bases de dados relacionadas e para a base final. RCBP-SP, 1997 a 2005.	62
Tabela 4	Número e percentual de óbitos, segundo causa básica e base de dados de relacionamento. RCBP-SP, 1997 a 2005.	63
Tabela 5	Número e percentual pacientes segundo status, antes e após o linkage entre o RCBP-SP com as bases de dados do PRO-AIM, Apac-SIA/SUS e FSeade. RCBP-SP, 1997 a 2005.	64
Tabela 6	Número e percentual de pacientes segundo status, antes e após o linkage. RCBP-SP, 1997 a 2005.	66
Tabela 7	Coefficientes brutos de incidência e mortalidade por câncer (por 100.000 habitantes) para o período de 1997 a 2005, antes e após o linkage. RCBP-SP, 1997 a 2005.	67
Tabela 8	Probabilidade de sobrevida global acumulada (%) geral e segundo sexo, antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	69
Tabela 9	Probabilidade de sobrevida global acumulada (%) segundo faixa etária, antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	72

Tabela 10	Probabilidade de sobrevida global acumulada (%), segundo topografia, antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	75
Tabela 11	Probabilidade de sobrevida global acumulada (%) geral e segundo sexo, antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	77
Tabela 12	Probabilidade de sobrevida global acumulada (%) segundo faixa etária, antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	80
Tabela 13	Probabilidade de sobrevida global acumulada (%), segundo topografia, antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	82
Tabela 14	Descrição dos pontos de corte, valores de sensibilidade e especificidade para o linkage RCBP-SP versus PRO-AIM. RCBP-SP, 1997 a 2005.	86
Tabela 15	Descrição dos pontos de corte, valores de sensibilidade e especificidade para o linkage RCBP-SP versus APAC-SIA/SUS. RCBP-SP, 1997 a 2005.	87

Lista de Figuras e Gráficos

Figura 1	RCBP-SP <i>versus</i> PRO-AIM. RCBP-SP, 1997 a 2005.	56
Figura 2	RCBP-SP <i>versus</i> APAC-SIA/SUS. RCBP-SP, 1997 a 2005.	59
Figura 3	RCBP-SP <i>versus</i> FSEADE. RCBP-SP, 1997 a 2005.	60
Gráfico 1	Probabilidade de sobrevida global acumulada (%), antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	70
Gráfico 2	Probabilidade de sobrevida global acumulada (%) segundo sexo, antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	71
Gráfico 3	Probabilidade de sobrevida global acumulada (%), segundo faixa etária, antes e após o linkage, truncada em quatro anos. RCBP-SP, 1997 a 2005.	73
Gráfico 4	Probabilidade de sobrevida global acumulada (%), antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	78
Gráfico 5	Probabilidade de sobrevida global acumulada (%) segundo sexo, antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	79
Gráfico 6	Probabilidade de sobrevida global acumulada (%), segundo faixa etária, antes e após o linkage, truncada em sete anos. RCBP-SP, 1997 a 2005.	81

Lista de Siglas e Abreviaturas

Aids -	<i>Acquired Immunodeficiency Syndrome</i>
AIH -	Autorizações de Internação Hospitalar
ANS -	Agência Nacional de Saúde
APAC -	Autorização de Procedimentos Ambulatórios de Alta Complexidade/Custo
APAC-Mag -	APAC Magnético
BIP-Aids -	Base Integrada Paulista de Aids
CBI -	Coefficiente bruto de incidência
CBM -	Coefficiente bruto de mortalidade
CD4+ -	<i>Cluster of differentiation 4</i>
CD8+ -	<i>Cluster of differentiation 8</i>
CEInfo -	Coordenação de Epidemiologia e Informação
CEP -	Código de Endereçamento Postal
CHI -	<i>Community Health Index</i>
CID-10 -	10 ^a Revisão da Classificação Internacional de Doenças
CID-O -	3 ^a Classificação Internacional de Doenças para Oncologia
CIH -	Comunicação de Internação Hospitalar
CNES -	Cadastro Nacional de Estabelecimentos de Saúde
CPF -	Cadastro de Pessoas Físicas
DATASUS -	Departamento de Informática do SUS
DIR-	Direção Regional de Saúde

DNV -	Declaração de Nascido Vivo
DST	Doença Sexualmente Transmissível
DO -	Declaração de Óbito
ES -	Espírito Santo
FOSP -	Fundação Oncocentro de São Paulo
FSEADE -	Fundação Sistema Estadual de Análise de Dados
GIA -	Gerência de Informação Assistencial
GISA -	Geoprocessamento e Informação Socioambientais
HIV -	<i>Human Immunodeficiency Virus</i>
IARC -	<i>International Agency for Research on Cancer</i>
IBGE -	Instituto Brasileiro de Geografia e Estatística
IESS -	Instituto de Estudos de Saúde Suplementar
INCA-	Instituto Nacional de Câncer
INSS -	Instituto Nacional de Seguridade Social
Lacen-CE	Laboratório Central do Ceará
MS -	Ministério da Saúde
NAA -	Núcleo de Apoio Administrativo
NAT -	Núcleo de Assessoria Técnica
NDC -	Núcleo de Divulgação e Comunicação
OMS/ <i>WHO</i> -	Organização Mundial da Saúde/ <i>World Health Organization</i>
OPAS/PAHO -	Organização Pan-americana de Saúde/ <i>Pan American Health Organization</i>
ORLS -	<i>Oxford Record Linkage Study</i>

PAPPE -	Unidade de Planejamento, Avaliação, Pesquisas e Programas Especiais
PELL -	<i>Pregnancy to Early Life Longitudinal Data Project</i>
PRO-AIM -	Programa de Aprimoramento das Informações de Mortalidade
RCBH -	Registro de Câncer de Base Hospitalar
RCBP -	Registro de Câncer de Base Populacional
RCBP-SP -	Registro de Câncer de Base Populacional do Município de São Paulo
RG -	Registro Geral
ROC -	<i>Receiver Operating Characteristic</i>
SAS -	Secretaria de Atenção à Saúde
SES -	Secretaria Estadual de Saúde
SIA/SUS -	Sistema de Informações Ambulatoriais do Sistema Único de Saúde
SIAB -	Sistema de Informação da Atenção Básica
SICLOM -	Sistema de Controle e Logística de Medicamentos
SIGA -	Sistema Integrado de Gestão do Atendimento
SIH/SUS -	Sistema de Informações Hospitalares do Sistema Único de Saúde
SIM -	Sistema de Informação sobre Mortalidade
SINAN -	Sistema de Informação de Agravos de Notificação
SINASC -	Sistema de Informações sobre Nascidos Vivos
SIPACS -	Sistema de Informação do Programa de Agentes Comunitários de Saúde

SIS -	Sistemas de Informação em Saúde
SISCEL -	Sistema de Controle de Exames Laboratoriais da Rede Nacional de contagem de Linfócitos CD4+/CD8+ e Carga Viral
SISCOLO -	Sistema de Informação do câncer do colo do útero
SISPRENATAL -	Sistema de Acompanhamento do Programa de Humanização no Pré-Natal e Nascimento
SMS -	Secretaria Municipal de Saúde
SNT -	Sistema Nacional de Transplante
SVE -	Sistema de Vigilância Epidemiológica
SP -	São Paulo
SPSS -	<i>Statistical Package for the Social Sciences</i>
SQL -	<i>Structured Query Language</i> Linguagem de Consulta Estruturada
SUS -	Sistema Único de Saúde
SVS -	Secretaria de Vigilância em Saúde
TRE -	Tribunal Regional Eleitoral
TRS -	Terapia Renal Substitutiva
UF -	Unidade Federativa
WALDS -	<i>Western Australian Data Linkage System</i>

1. INTRODUÇÃO

1.1 Sistemas de Informação em Saúde

O sistema de informação pode ser compreendido como um conjunto inter-relacionado e organizado de pessoas e componentes informatizados que coletam e/ou recuperam, processam, armazenam e disseminam informações destinadas a auxiliar nas tomadas de decisões. Os sistemas de informação podem ser classificados como: 1) operacionais, que processam operações relacionadas a transações e procedimentos de sistemas de arrecadação, de saúde e de educação; 2) gerenciais, que processam grupos de dados, transformando-os em informações para gestão de serviços, por exemplo, número de serviços pendentes, número de obras realizadas, entre outros; e 3) estratégicos, que trabalham com informações em nível macro voltadas para funções organizacionais, isto é, analisam o impacto dos serviços prestados e relacionam o custo-benefício das ações (LAUDON e LAUDON, 2004 p.5; REZENDE, 2007).

Os Sistemas de Informação em Saúde (SIS) são instrumentos para aquisição, organização e análise de dados relacionados aos problemas de saúde de uma população, visando o planejamento e a operacionalização de ações para uma área geográfica definida (DESHAIES e SEIDMAN, 1971; WHITE, 1980).

Na segunda metade da década de 1990, a Organização Mundial da Saúde (WHO/PAHO 1997) definiu o SIS como aquele que tem o propósito de selecionar os dados pertinentes a serviços e transformá-los em informações necessárias para processos de decisões. Estas informações referem-se às organizações e aos

indivíduos e são úteis para o planejamento, administração e avaliação dos serviços de saúde. Considera-se, ainda, que a transformação de um dado bruto em informação exija, além da análise, a divulgação e as recomendações para ações efetivas.

Neste sentido, gerar a informação torna-se um processo muito mais complexo e dinâmico, na medida em que envolve componentes tecnológicos, econômicos, políticos, conceituais e ideológicos. Assim, os SIS's podem tornar-se instrumento a serviço das políticas públicas de modo a atender aos interesses específicos e práticas institucionais (MS, 2008).

1.2 Sistemas de Informação em Saúde no Brasil

A história dos SIS no Brasil é recente. Na década de 1970, o governo brasileiro, por meio do Ministério da Saúde (MS), interessou-se em criar um sistema de informação para auxiliar as decisões junto à área de saúde. No ano de 1973, com a Resolução 274 da Secretaria Geral do Ministério, foi delegada à Unidade de Planejamento, Avaliação, Pesquisas e Programas Especiais (PAPPE), com a colaboração da Organização Pan-americana de Saúde (OPAS), a organização dos SIS para o Brasil. Neste mesmo ano, foi realizada em Brasília, entre 27 de agosto e 6 de setembro, a Primeira Reunião de Chefes de Programas de Sistemas de Informação no Setor Saúde, com representantes dos Estados de Pernambuco, Pará, Guanabara, São Paulo, Rio Grande do Sul e Distrito Federal (SILVEIRA, 1974).

No entanto, a produção das informações se deu de maneira centralizada, principalmente, por falta de recursos humanos e tecnológicos no país como um todo (BALDIJÃO, 1992). Somente em 1988, com a implantação do Sistema Único de

Saúde (SUS), os SIS's passaram por um processo gradual de descentralização, voltado para a gestão e definição de prioridades em nível municipal (ALMEIDA, 1998).

Atualmente, o Ministério da Saúde dispõe de sete Secretarias e, dentre estas, duas são responsáveis pelos principais SIS. São elas: a Secretaria de Vigilância em Saúde (SVS) e a Secretaria de Atenção à Saúde (SAS). A SVS é responsável pelo gerenciamento dos SIS's para uso em epidemiologia, com a finalidade de resumir, organizar, analisar e divulgar informações epidemiológicas de interesse do país e da relação com países vizinhos. Além disso, está sob sua responsabilidade construir e revisar indicadores epidemiológicos para o monitoramento do quadro sanitário nacional, a fim de se desenvolver e difundir técnicas aplicáveis ao melhor funcionamento do Sistema Único de Saúde (SUS) (DATASUS, 2008a).

A SAS atua no gerenciamento de políticas públicas de atenção à saúde no âmbito do SUS. A esta secretaria compete monitorar, avaliar e divulgar informações sobre a atenção básica como, por exemplo, do Programa de Saúde da Família, atenção especializada como o Sistema Nacional de Transplante, Atenção Hospitalar e programas voltados a populações específicas como mulher, idoso, jovem e adolescente, em situação de vulnerabilidade e em situação de violência. Também sob sua gerencia estão a atualização do Cadastro Nacional de Estabelecimentos de Saúde (CNES) e o monitoramento de recursos humanos e financeiros (DATASUS, 2011a).

No Brasil, são seis os SIS's mais utilizados para vigilância e avaliação de serviços de abrangência nacional sob responsabilidade do MS. A seguir será feita breve descrição dos mesmos.

1) O Sistema de Informação sobre Mortalidade (SIM)

Criado pelo MS, em 1975/76, o SIM possui cobertura populacional e tem como objetivo registrar os dados sobre mortalidade em nível nacional, de maneira abrangente e confiável, propiciando a produção de estatísticas de mortalidade e a construção dos principais indicadores de saúde. O instrumento básico de coleta é a Declaração de Óbito (DO) (Anexo 1), emitido em três vias pela unidade notificadora, podendo estas serem hospitais e outros estabelecimentos de saúde e Institutos de Medicina Legal. A primeira é direcionada as Secretarias Estaduais ou Municipais, a segunda é encaminhada pela família aos Cartórios de Registro Civil e, a terceira, permanece na unidade notificadora (LAURENTI et al., 2005; DATASUS, 2008b).

2) O Sistema de Informações sobre Nascidos Vivos (SINASC)

Implantado pelo MS, em 1990, o SINASC possui cobertura populacional e tem por objetivo reunir informações sobre os nascimentos vivos ocorridos em todo território nacional, fornecendo dados referentes à saúde materno-infantil. O instrumento de coleta é a Declaração de Nascido Vivo (DNV) emitida em três vias, preenchida pelos hospitais e outros estabelecimentos de saúde. A primeira via deste documento é encaminhada aos órgãos responsáveis pelo processamento da informação, a segunda via é entregue à família para realizar o registro nos Cartórios de Registro Civil e, a terceira, permanece no estabelecimento de saúde (DATASUS, 2008c).

3) O Sistema de Informação de Agravos de Notificação (SINAN)

Este sistema, implantado em 1993, é de abrangência populacional e tem como objetivo coletar os dados gerados, rotineiramente, pelo Sistema de Vigilância Epidemiológica, nos níveis federal, estadual e municipal. As informações obtidas através das unidades básicas de saúde, hospitais ou outros estabelecimentos cadastrados, permitem conhecer a incidência e a prevalência dos agravos de notificação compulsória (MS, 2007).

A notificação de um caso suspeito e/ou confirmado é feita em uma ficha padronizada pela SVS/MS, em duas vias pré-numeradas, sendo a primeira digitada na própria unidade e enviada para as SMS, Secretaria Estadual de Saúde e MS. A segunda via permanece arquivada na própria unidade notificadora. Para cada agravo existe um modelo específico de ficha que se configura como um roteiro de investigação (MS, 2007).

Cabe ressaltar que, além dos agravos que constam na lista nacional de doenças de notificação compulsória, é facultado aos Estados e Municípios incluir outros problemas de saúde que são objeto de vigilância sistemática na região (MS, 2007).

4) O Sistema de Informação Hospitalares do Sistema Único de Saúde (SIH/SUS)

Implantado em 1983, este sistema tem como objetivo principal o pagamento das internações efetuadas pela rede hospitalar, mas, também, é utilizado na análise da

morbidade hospitalar. A coleta destas informações é feita por meio das Autorizações de Internação Hospitalar (AIH), que são emitidas pela rede hospitalar federal, estadual, municipal e privada (MS, 2006).

Cabe ressaltar que, as informações geradas por este sistema é limitada à população que faz uso do SUS, em todo território nacional, e de hospitais privados, porém vinculados ao SUS. (MS, 2006).

5) O Sistema de Informações Ambulatoriais do Sistema Único de Saúde (SIA/SUS)

O SIA/SUS foi implantado em todo país, pelo MS, em 1993, com o objetivo de registrar os atendimentos e procedimentos realizados em cada estabelecimento de saúde e processar a produção dos mesmos. Os dados gerados neste sistema são referentes à população atendida pelo SUS e permitem a análise de alguns indicadores de saúde, como o total de consultas médicas, número de exames solicitados, perfil dos clientes atendidos nos diferentes estabelecimentos de saúde, entre outros (MS, 2008).

A Autorização de Procedimentos Ambulatórios de Alta Complexidade/Custo (APAC) é um SIS integrado ao SIA/SUS. Foi criado em 1996 pelo MS, com o objetivo de registrar os atendimentos e procedimentos de alto custo e complexidade. A emissão de uma APAC (Anexo 2) é feita a partir de um laudo em duas vias, autorizada por profissionais de hospitais públicos e, em nível central, pelas Secretarias Estaduais e Municipais de Saúde e seus órgãos administrativos. Para o registro da informação é utilizado um aplicativo magnético (APAC-Mag) enviado à

SES ou SMS, dependendo da gestão do estabelecimento. Por meio da APAC são fornecidas as autorizações para diversos procedimentos como a quimioterapia e a radioterapia e fornecimento de medicamentos (MS, 2008).

6) O Sistema de Informação da Atenção Básica (SIAB)

O SIAB foi implantado em 1998, em substituição ao Sistema de Informação do Programa de Agentes Comunitários de Saúde (SIPACS). Dentre seus objetivos está agregar e processar as informações obtidas pelo cadastramento familiar, acompanhamento dos grupos de risco e registro de atividades e procedimentos. Por meio do SIAB obtêm-se dados sobre condições de moradia, saneamento, situação de saúde, produção e composição das equipes de saúde (DATASUS, 2011b).

Os instrumentos de coleta são conhecidos como Ficha A (cadastramento das famílias), Ficha B (acompanhamento domiciliar de pessoas com hipertensão, com diabetes, portadoras de tuberculose, hanseníase e mulheres em gestação), Ficha C (ficha de acompanhamento da criança) e Ficha D (ficha de registro de atividades, procedimentos e notificação). Os dados são digitados no aplicativo denominado Sistema Integrado de Gestão do Atendimento (SIGA) e são enviados às Secretarias Municipais e Estaduais e MS (DATASUS, 2011b).

1.2.1 Sistemas de Informação em Saúde em São Paulo

No Estado de São Paulo, é a Fundação Sistema Estadual de Análise de Dados (FSEADE) a instituição responsável por compilar, analisar e divulgar as informações

sobre condições de vida, economia, educação, processos eleitorais, de justiça e segurança, perfil populacional, estatísticas vitais, estatísticas de saúde e de trabalho dos Municípios paulistas. Esta fundação está vinculada à Secretaria Estadual de Planejamento e Desenvolvimento Regional do Estado de São Paulo (FSEADE, 2011).

A FSEADE disponibiliza à SES-SP, desde 1978, os bancos de dados anuais de nascimentos e óbitos, assim como os indicadores de mortalidade e natalidade dos Municípios paulistas. Em 2005, em parceria com a SES-SP, foi realizado um novo projeto entre estas instituições para a elaboração de uma base de dados unificada de nascimentos e óbitos, possibilitando uma base mais completa e precisa. Esta vinculação permite identificar os eventos não captados por um dos sistemas, unificar todas as variáveis demográficas e epidemiológicas e aprimorar a rotina de verificação de consistência (WALDVOGEL et al., 2008).

Quanto ao Município de São Paulo, o órgão responsável pela consolidação, análise e divulgação de diferentes tipos de informação é a Coordenação de Epidemiologia e Informação (CEINFO), que inclui as seguintes gerências: 1) Gerência de Informação Assistencial (GIA), que é responsável pela análise das bases de dados do sistema de informação cadastral e de produção de serviços do SUS, sendo os ambulatoriais (SIA, APAC, Sistema de Acompanhamento do Programa de Humanização no Pré-Natal e Nascimento - SISPRENATAL - SIAB e Sistema de Informação do câncer do colo do útero - SISCOLO) e os hospitalares (SIH e Comunicação de Internação Hospitalar - CIH) e cadastral (CNES); 2) Geoprocessamento e Informação Socioambientais (GISA), que é responsável pelo georeferenciamento das informações sociodemográficas, ambientais e de redes de

serviços; 3) Programa de Aprimoramento das Informações de Mortalidade (PRO-AIM), que é responsável pelo processamento, análise e divulgação das informações sobre mortalidade referente a todos os óbitos ocorridos no Município de São Paulo; e 4) Gerência SINASC, que é responsável pelo processamento, análise e divulgação das informações sobre nascidos vivos no Município de São Paulo (CEINFO, 2011).

Além destas gerências, a CEINFO possui mais três núcleos de suporte aos sistemas de informação, Núcleo de Apoio Administrativo (NAA), Núcleo de Assessoria Técnica (NAT) e Núcleo de Divulgação e Comunicação (NDC) (CEINFO, 2011).

1.3 Relacionamento de bases de dados em Saúde

A disponibilidade de grandes bases de dados, informatizadas, em saúde tornou a técnica de relacionamento de bases de dados, também conhecida como *linkage*, uma alternativa para diferentes tipos de estudos. Esta técnica proporciona a geração de uma base de dados mais completa e de baixo custo operacional (ALMEIDA e MELLO JORGE, 1996; HOWE, 1998; FSEADE, 2010).

No processo de *linkage* dois métodos são utilizados: o determinístico ou lógico e o probabilístico. O relacionamento determinístico identifica os indivíduos de diferentes bases de dados a partir de uma ou várias variáveis. Por exemplo, nome, data de nascimento e data do óbito, ou um identificador comum como número do cartão SUS ou RG ou CPF (MORAES et al., 2009). Por outro lado, o relacionamento probabilístico faz uso de um escore que mostra o quão verossímilhanete os registros pareados pertencem à mesma pessoa (STEVENS, 2007). Este método se baseia na

padronização, blocagem e pareamento de registros (FELLEGI e SUNTER, 1969; JARO, 1989).

A vantagem do *linkage* determinístico está na certeza em identificar um par verdadeiro quando este apresenta um identificador comum ou um conjunto de variáveis com boa qualidade de preenchimento. Outra vantagem está na sua operacionalização, pois não há a necessidade de cálculos complexos de probabilidade ou de programas específicos para a combinação entre as bases de dados. No entanto, a ausência de um identificador único ou um conjunto de variáveis que possam, de forma única e precisa, identificar um mesmo indivíduo nas bases de dados relacionadas podem dificultar o processo. Nesta situação, o *linkage* determinístico poderia apresentar dificuldades em localizar registros com forte probabilidade de serem os mesmos, mas que apresentam diferenças nos campos, em exemplo desta situação, é a ocorrência de nome incompleto, mas com endereço igual e data de nascimento com valores trocados. Isso ocorre, pois não existe ponderação para os tipos de variáveis utilizadas no relacionamento (MACHADO, 2004 e TROMP et al., 2011).

Quanto ao *linkage* probabilístico, a sua vantagem está na capacidade de identificar um indivíduo em diferentes bases de dados mesmo que estes não possuam identificadores unívocos comuns e apresentem problemas de inconsistência, erros e informações não declaradas. A desvantagem desta técnica é o poder de discriminação das variáveis utilizadas no processo de relacionamento, pois as probabilidades são influenciadas pelo número de valores válidos e sua uniformidade. Por exemplo, pares podem ser classificados como verdadeiros quando na realidade os registros pertencem a indivíduos diferentes. Este tipo de erro é conhecido como erro de

homônimos. Em relacionamentos de bases de dados que utilizam poucos campos e os mesmos apresentam baixo poder discriminatório, os erros de homônimos tendem a aumentar o número de falso-positivos (MCGLYNN et al., 1998; CAMARGO e COELI, 2000, MACHADO e HILL 2004; TROMP et al., 2011).

Porém, relacionar registros de diferentes fontes, independente da técnica de *linkage*, mostra-se uma tarefa pouco trivial, pois as bases de dados são geradas com finalidades distintas, não havendo padronização entre os campos das variáveis ou na digitação dos dados coletados. Além disso, são frequentes os erros de digitação e a presença de valores ignorados (CAMARGO e COELI, 2000).

Em um estudo de revisão de literatura, MACHADO (2004) descreve a aplicabilidade da técnica de *linkage*, tanto determinística quanto probabilística, utilizadas em pesquisas que focavam desfechos de saúde na infância, realizadas nos Estados Unidos, Canadá, Escócia, Suécia, Noruega, Dinamarca, Inglaterra, Japão e Brasil. Dos nove países citados, houve uma variação entre o uso da técnica de *linkage*. Contudo, pode-se verificar que em países como Suécia, Noruega, Dinamarca e Japão em que as pessoas possuem um número único de identificação ou os registros dos dados são de boa qualidade houve uma predileção pelo *linkage* determinístico.

O Canadá mostrou-se precursor no uso desta técnica com experiências bem sucedidas desde a década de 1970. O *National Cancer Institute of Canada*, em conjunto com o *Statistics Canada*, relacionaram os dados do registro nacional de casos incidentes de câncer, que possui informações disponíveis a partir do ano de 1969, com a base de dados sobre mortalidade, que possui informações disponíveis desde 1940. Neste trabalho foi utilizado o relacionamento probabilístico, assim

recomendado pelo Dr. Howard Newcombe, em 1959, introdutor do método. O objetivo foi mostrar a técnica de *linkage* probabilístico como uma alternativa para o relacionamento de bases de dados, a fim de se obterem informações sobre o seguimento de pacientes/pessoas em estudos de coorte (HOWE, 1998).

A partir desta experiência, HOWE (1998) cita dois estudos mostrando que o processo de *linkage* serviu, principalmente, para demonstrar a possibilidade de aquisição de dados consistentes e de qualidade para estudos de coorte. No *The Fluoroscopy Study*, que verificou a associação entre a exposição à radiação ionizante com a presença de câncer, o relacionamento possibilitou mostrar uma relação do tipo dose resposta entre a exposição e a presença de câncer de mama. Outro trabalho foi *The Eldorado Study*, que relacionou a exposição ao urânio dos trabalhadores da companhia “*Eldorado Resources Limited*” com a presença de câncer de pulmão.

Diversos outros autores utilizaram esta técnica em seus estudos. Na Escócia, um estudo relacionou a base de dados *Community Health Index* com os dados do censo escocês por meio da técnica de *linkage* probabilístico. O objetivo foi gerar uma base de dados que permitisse analisar a incidência e a mortalidade por infarto agudo do miocárdio nos sul-asiáticos em relação aos não sul-asiáticos ao longo do tempo. (FISCHBACKER et al., 2007).

Em outra pesquisa, BARFIELD et al., (2008) utilizaram o *linkage* entre a base de dados do *Pregnancy to Early Life Longitudinal Data Project* e a base de dados de informações sobre altas hospitalares e nascimentos do Estado de Massachusetts para avaliar os serviços de intervenção precoce. Nesta pesquisa os autores buscaram identificar as características maternas e de nascimentos dos recém

nascidos com muito peso baixo ao nascer (< 1.200 kg), que precisariam de intervenção precoce, a fim de melhorar a rapidez no acesso à estas crianças.

No estudo conduzido por WOTTON et al. (2011) o *linkage* probabilístico foi realizado para relacionar as bases de dados hospitalares e de mortalidade do *Oxford Record Linkage Study* entre 1963 e 1998, e entre 1999 e 2008, com o objetivo de analisar o risco de câncer em pessoas acima dos 30 anos de idade, admitidas em hospitais por diabetes mellitus e residentes na região Sul da Inglaterra, entre os anos de 1963 e 2008.

No Brasil, estudos também mostram a aplicabilidade da técnica de *linkage*. VIDAL et al. (2006), com o objetivo analisar a mortalidade e as causas básicas de óbito em uma coorte de pacientes idosos que foram submetidos a cirurgia de reparação de fratura de quadril, utilizaram a técnica de *linkage* probabilístico entre o SIH e o SIM. A vinculação destas bases de dados possibilitou identificar que a maioria das mortes ocorrem após a alta hospitalar, reforçando a importância do relacionamento destas bases de dados.

A pesquisa realizada por PEREIRA et al. (2007) teve como objetivo estimar as taxas de mortalidade neonatal e pós-neonatal na população do Município do Rio de Janeiro, no período de julho de 1999 a 2001, e compará-las com alguns indicadores de acesso e tipo de serviço e descrever a causa básica dos óbitos, segundo o perfil da maternidade. Os dados sobre a mortalidade foram obtidos pelo processo de *linkage* probabilístico entre o estudo citado e a base de dados do SIM. Com o *linkage*, foram identificados mais 20% de óbitos, com a taxa de mortalidade neonatal precoce atingindo 6,8 por mil nascidos vivos.

Em outro estudo, com o objetivo de criarem uma base de dados nacional em Terapia Renal Substitutiva (TRS), CHERCHIGLIA et al. (2007) utilizaram o processo de *linkage* probabilístico para relacionar as bases de dados da APAC-SIA/SUS, SIM e do SIH. Esta nova base permitiu obter informações completas sobre o paciente, como sexo, data de nascimento, modalidade de entrada e saída da TRS, diagnóstico na entrada, desfecho (óbito, tratamento, perda de seguimento) e gastos. A partir deste relacionamento foi possível reconstruir toda trajetória dos pacientes com TRS, viabilizando a construção de indicadores econômico-epidemiológicos que podem auxiliar na melhoria de políticas voltadas para esta população.

No estudo de SOUSA e PINHEIRO (2011) foi utilizada a técnica de *linkage* probabilístico, entre o SIM, SIH/SUS e SINAN, no período de 2002 a 2004, em residentes do Município do Rio de Janeiro, para identificar o sub-registro de mortes e de admissão hospitalar por tuberculose. Observou-se que, neste período, dos 542 óbitos por tuberculose, 43,2% não foram registrados no SINAN. Quanto às admissões hospitalares, das 1.079, 22,1% não estavam notificadas. Verificou-se, também, que a subnotificação variou conforme as regiões administrativas de saúde, mostrando que a subnotificação de óbito passou de 50% em algumas regiões e a subnotificação da admissão hospitalar ficou entre 12,7% e 37,8%.

ALMEIDA e MELLO JORGE (1996) realizaram estudo com o objetivo de testar a utilização do *linkage* determinístico entre as bases de dados do SIM e do SINASC, para analisar a mortalidade neonatal em Santo André, Município da Grande São Paulo. As autoras concluíram que o uso do *linkage* foi efetivo, pois permitiu a análise dos riscos relativos e das probabilidades de morte entre os nascidos vivos.

WALDVOGEL (2002) fez uso do *linkage* determinístico entre as bases de dados do Instituto Nacional de Seguridade Social (INSS) e do SIM, com o objetivo de propor uma metodologia para identificar e mensurar casos fatais de acidentes de trabalho. A vinculação destas bases de dados possibilitou identificar os casos fatais cobertos e os não-cobertos pelo INSS, bem como as diferentes exposições ao risco de acidentes fatais, de acordo com o perfil da população analisada. Segundo a autora, nesta pesquisa o relacionamento de bases de dados foi uma metodologia viável, de baixo custo monetário e de benefício social.

Outro estudo utilizou a técnica de *linkage* determinístico para comparar as informações das declarações de nascidos vivos (SINASC) com as obtidas na pesquisa “Mortalidade Perinatal na Região Sul do Município de São Paulo: Um Estudo Caso-controle de Base Populacional”, validada por meio de entrevistas com as mães e dados registrados nos prontuários hospitalares. Neste estudo, os autores relataram que a paridade materna e a presença de anomalias congênicas mostram elevada ausência de registro e que o SINASC superestima em cerca de duas vezes a presença de nascimentos de mães de baixa escolaridade (ALMEIDA et al., 2006).

SILVA et al. (2009) utilizaram a técnica do *linkage* determinístico com o objetivo de analisar a qualidade da informação fornecida pelo SINASC e pelo SIM. Foram investigadas 11.182 DNV entre os anos de 2000 e 2002 e 233 DO de menores de um ano entre os anos de 2000 e 2003, no Município de Maracanaú, Ceará. A partir deste estudo foi identificado um sub-registro de 92 nascimentos pelo SINASC e 69 casos de DNV duplicadas.

Na pesquisa realizada por SILVA et al. (2010) a técnica de *linkage* determinístico foi utilizada para gerar uma base de dados que possibilitou analisar o

perfil dos recém-nascidos, das mães e da mortalidade neonatal precoce, segundo complexidade do hospital e vínculo com o SUS, na região Metropolitana de São Paulo. Nesta pesquisa os autores utilizaram a base de dados unificada de eventos vitais gerada pela FSEADE por meio da vinculação determinística, a partir dos dados coletados nos cartórios de registro civil, no SIM e no SINASC. Para a identificação do estabelecimento de saúde, também foi realizado o *linkage* determinístico entre o registro de nascidos vivos e a base de dados do CNES.

No estudo TANCREDI et al. (2010) os autores trabalharam com a Base Integrada Paulista de Aids (BIP-Aids) para determinar a sobrevida de pessoas vivendo com HIV/Aids e avaliar os fatores prognósticos para estes pacientes, residentes no Município de São Paulo, entre os anos pares de 1990 a 2002. A BIP-Aids foi elaborada por meio do *linkage* determinístico entre os óbitos existentes nos sistemas de estatísticas vitais da FSEADE que apresentavam menção de Aids ou HIV entre as causas de morte, e dos casos notificados de Aids registrados pelo SINAN. Destaca-se que esta base de dados possibilitou não só analisar a sobrevida, mas os dados de mortalidade e morbidade de pacientes vivendo com HIV/Aids, produzindo informações epidemiológicas, subsídios para ações de prevenção e para o monitoramento da gestão de políticas públicas (WALDVOGEL et al., 2010).

1.4 Acurácia da técnica de *linkage*

Nos últimos anos, diferentes estudos foram desenvolvidos para avaliar a acurácia da técnica de *linkage* na obtenção de informações longitudinais. OBERAIGNER e STUHLINGER (2005) aplicaram o *linkage* probabilístico na base

de dados do Registro de Câncer de Tyrol, Áustria, para os anos de 1999 a 2003. Os pesquisadores objetivaram a padronização da base de dados e a verificação interna de duplicidades. Concluíram que este método de *linkage* possui precisão adequada quanto à identificação de pares, sendo que 89,0% dos casos foram identificados automaticamente.

Estudo realizado na Escócia verificou a validade do *linkage* probabilístico a partir do relacionamento de uma coorte de mulheres nascidas entre 1950 e 1956, em Aberdeen, com os dados de nascimentos do Sistema de Registro de Maternidade Escocês, no período de 1969 a 1999. Para validar as informações foram enviados questionários para 5.540 mulheres, com questões específicas sobre o número de partos, número de filhos e detalhes sobre os nascimentos, como data, sexo e local. Os pesquisadores observaram uma concordância de 89%, entre os dados obtidos pelo *linkage* quando comparados às respostas do questionário e sensibilidade de 92% (NITSCH et al., 2006).

Em Amsterdam, Holanda, foi realizada uma pesquisa para validar a técnica de *linkage* probabilístico. Os autores relacionaram três bases de dados pertencentes ao *Dutch Perinatal Registries*, compostas por informações de parteiras, obstetras e pediatras, relacionando dados sobre a mãe, a criança e o nascimento. A partir do banco de dados criado com o relacionamento das três bases, os pesquisadores fizeram entrevistas via telefone para confirmar os resultados. Concluíram que a técnica de *linkage* probabilístico foi capaz de gerar um banco de dados de alta qualidade (MÉRAY et al., 2007).

Na pesquisa de FOURNEL et al. (2009), que relacionou os dados de pacientes com câncer hospitalizados no *Institut Gustave Roussy*, entre 1998 e 2000,

com a base de dados de mortalidade da França, a técnica de *linkage* probabilístico mostrou excelente desempenho para 10.089 pacientes pareados, apresentando sensibilidade de 97,2% e especificidade de 99,4% na identificação de um par verdadeiro, a partir do relacionamento entre as variáveis nome, sobrenome, data de nascimento e código do local de nascimento.

Outro estudo que validou a técnica de *linkage* probabilístico foi realizado por HERK-SUKEL et al. (2010) na Holanda. Nesta pesquisa foram relacionadas as bases de dados do *Eindhoven Cancer Registry* com a base da *PHARMO Record Linkage System*, que compreende informações de várias bases de dados, tais como a utilização de drogas, dentro e fora de ambulatórios, dados de hospitalizações e de laboratórios. Para validação foi utilizada uma amostra de 11.989 pacientes a partir do relacionamento entre as bases de dados e, como padrão ouro, foram utilizados dados originais de pacientes, obtidos de outra fonte de informação. Os resultados apresentaram uma sensibilidade de 98,3% e especificidade de 99,5% do método na identificação de um par verdadeiro, a partir das variáveis nome completo e endereço.

Na pesquisa de COUTINHO e COELI (2006), os autores avaliaram a acurácia do método de relacionamento probabilístico na identificação de óbitos em uma coorte de idosos, por meio do cruzamento entre o banco de dados de declarações de óbitos do Estado do Rio de Janeiro com um banco de dados produzido especialmente para o estudo. Foi observado que o método atingiu uma sensibilidade de 85,5% e uma especificidade de 99,4%, sendo os valores preditivos positivo e negativo de 98,1% e 94,9%, respectivamente. Assim, concluiu-se que o relacionamento probabilístico de registros pode ser uma ferramenta útil para a identificação de óbitos.

O estudo de FONSECA et al. (2010) verificou a acurácia do *linkage* probabilístico para a identificação de óbitos entre os casos de Aids notificados no SINAN. O *linkage* foi realizado entre as bases de dados de pessoas vivendo com HIV/Aids, (formada pelos casos do SINAN), do sistema de Controle de Exames Laboratoriais da Rede Nacional de contagem de Linfócitos CD4+/CD8+ e Carga Viral (SISCEL) e do Sistema de Controle e Logística de Medicamentos (SICLOM) com a base de dados do SIM. Para o processo de validação foram selecionados todos os casos de Aids com datas de óbitos registradas no SINAN-AIDS entre 2002 e 2005, totalizando 19.750 e casos de pessoas vivendo com HIV/Aids em 2006, pelo SISCEL, totalizando 38.675. Foi observado que o método apresentou uma sensibilidade de 87,6%, especificidade de 99,6% e valor preditivo positivo de 99,2%.

MIGOWSKI et al. (2011) avaliaram a acurácia do *linkage* probabilístico para identificar mortes em pacientes submetidos a procedimentos cardíacos complexos. Este *linkage* foi realizado entre a base de dados de cirurgia cardíaca de um instituto de cardiologia no ano de 2005 e a base de dados de autorização de admissão hospitalar. Ao final, foram relacionados 225 pares verdadeiros e, somente, em 99 pacientes a informação pôde ser confirmada. A validação foi realizada por meio de entrevistas telefônicas para identificar o status vital destes pacientes, apresentando valores para sensibilidade de 90,6% e especificidades de 100%.

Em relação à validade da técnica de *linkage* determinístico, no estudo de LI et al. (2006), envolvendo a população de *Calgary Health Region*, Alberta, foram utilizadas três bases de dados do Canadá, a saber, registro populacional, registro de óbitos em hospitais e registro de Estatísticas Vitais. O *linkage* entre as bases de dados do registro populacional e do registro de Estatísticas Vitais apresentou uma precisão

de 96,9% ao identificar pares verdadeiros, em 2001 e 2002. Para o mesmo período, observa-se que houve uma precisão de 98,9% na identificação dos pares verdadeiros no *linkage* entre a base do registro de óbitos em hospitais com a base do registro de Estatísticas Vitais.

PACHECO et al. (2008) desenvolveram um algoritmo para o *linkage* determinístico num estudo com o objetivo recuperar informações sobre o status vital de pessoas vivendo com HIV. Os dados para a validação deste algoritmo foram obtidos de pacientes que vieram a óbito de dois estudos de coorte realizados no Rio de Janeiro e da base de dados de mortalidade do Estado do Rio de Janeiro, entre os anos de 1999 e 2006. Pela seleção automática esta técnica apresentou uma sensibilidade de 92% e uma especificidade de 100%. Os autores concluíram que o algoritmo aplicado na técnica de *linkage* determinístico foi acurado, capaz de resgatar dados de seguimento perdidos.

No estudo desenvolvido pela FSEADE em parceria com o Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto Jones dos Santos Neves e a SES do Espírito Santo, os pesquisadores realizaram a integração da base de dados do MS e IBGE para os eventos vitais, nascimento e óbito, ocorridos no Estado do Espírito Santo, com o objetivo de testar a metodologia de *linkage* determinístico, utilizando o número das DNV e DO. Foi verificado que o *linkage* determinístico indicou um alto grau de precisão para os casos vinculados, tendo 96,3% dos pares formados com os nascimentos, a partir do número da DNV. Quanto aos óbitos, foram confirmados como pares verdadeiros 97,2% dos registros vinculados com o número da DO (FSEADE, 2009).

1.5 Registro de Câncer de Base Populacional (RCBP)

Os RCBP's são instituições que coletam, classificam, armazenam, processam e analisam dados de casos novos de câncer, de maneira sistemática e contínua, oriundos de todas as fontes possíveis de notificação e referente a uma área geográfica definida. Produzem informações sobre a incidência, a mortalidade e sobrevida por câncer de uma população específica, em área geográfica delimitada. Os RCBP's têm como finalidade promover a vigilância epidemiológica do câncer e contribuir para o planejamento dos serviços de saúde (LATORRE, 2001; MIRRA, 2005; INCA, 2010).

O desenvolvimento dos RCBP's é antigo. O primeiro relato é datado do século XVIII, na cidade de Londres (Inglaterra), porém, não houve continuidade deste projeto. Somente, no ano de 1899, em Hamburgo (Alemanha), um inquérito populacional foi conduzido por Alexander Katz, seguido de um estudo com médicos de todo o país, a fim de se estabelecer a prevalência do câncer. Tal fato incentivou outros países europeus a realizarem estudos similares (WAGNER, 1991).

No entanto, o primeiro RCBP com coleta de dados contínua foi criado nos Estados Unidos, no Estado de Massachusetts, em 1927 e somente a partir da década de 1940 foram estabelecidos os padrões conhecidos atualmente (WAGNER, 1991).

No Brasil, os primeiros RCBP's foram implantados em Recife (1967) e em São Paulo (1969). Já no fim da década de 1980 havia seis RCBP's ativos (Recife, São Paulo, Fortaleza, Porto Alegre, Goiânia e Belém), cobrindo, aproximadamente, 11% da população do país. Segundo dados do INCA – Instituto Nacional de Câncer (2010), no momento, existem 17 RCBP's em atividade nas cidades de Aracajú,

Belém, Belo Horizonte, Campo Grande, Cuiabá, Curitiba, Fortaleza, Goiânia, Jaú, João Pessoa, Manaus, Natal, Palmas, Porto Alegre, Recife, Salvador e São Paulo (INCA, 2010).

1.5.1 O Registro de Câncer de Base Populacional em São Paulo (RCBP-SP)

A idéia do RCBP-SP surgiu em 1º de janeiro de 1963. Nos três primeiros anos de experiência, após um projeto piloto (1963-1965), o RCBP-SP mostrou-se viável desde que obtivesse um patrocínio oficial. Atualmente, o RCBP-SP mantém convênio com a Prefeitura do Município de São Paulo através da Secretaria Municipal de Saúde e é financiado, também, pelo MS. A Secretaria de Estado da Saúde fornece parte dos dados, através da Fundação Oncocentro de São Paulo (FOSP) e o Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo sedia o RCBP-SP (RCBP-SP, 2011).

O RCBP-SP é um dos registros mais importantes e antigos do país, localizado na cidade com maior população e de maior poder econômico, concentrando um grande número de casos incidentes de câncer. A cobertura deste registro refere-se ao Município de São Paulo que possui uma área geográfica de 1.522,986 km², essencialmente urbana (99%), correspondendo a 0,6% da área total do Estado de São Paulo e 0,02% do Brasil. De acordo com o último censo (2010) a população do Município é de 11.244.369 habitantes (IBGE, 2011), o que representa 27,3% da população do Estado e 5,9 % do país.

Atualmente, o RCBP-SP dispõe de 301 fontes de notificação, sendo 168 hospitais, 76 clínicas, 35 laboratórios de anatomia patológica, 2 serviços de verificação de óbitos e 20 Registros de Câncer de Base Hospitalar (RCBH). Os dados provenientes dos RCBH's são fornecidos pela Fundação Oncocentro de São Paulo (FOSP).

A coleta dos casos de câncer é realizada de duas formas: ativa e passiva. A coleta ativa compreende 245 fontes de notificação que são visitadas, periodicamente, por uma equipe de profissionais do RCBP-SP que registra as informações em ficha uma ficha padronizada (Anexo 3). As áreas de coleta de dados são divididas em cinco, a saber, Centro, Norte, Sul, Leste e Oeste (MIRRA et al., 2007).

A equipe deste registro é composta por uma coordenadora, uma vice-coordenadora, uma supervisora de base de dados, uma supervisora das fontes de notificação, cinco digitadores, sete registradores, um analista de sistema e uma comissão de assessoria científica (RCBP-SP, 2011).

Na base de dados do RCBP-SP são registrados os tumores malignos com comportamentos “*in situ*” e invasivo. Cabe ressaltar que, ao apresentar mais de uma neoplasia maligna primária os pacientes são registrados separadamente, porém com o mesmo número de registro. A classificação da topografia e da morfologia é feita conforme a 3^a Classificação Internacional de Doenças para Oncologia – CID-O. Para a confirmação do diagnóstico são utilizados exames clínico, de imagem, macroscópico, observação no ato cirúrgico, histológico e necrópsia (MIRRA et al., 2007; RCBP-SP, 2011).

De acordo com a última publicação do RCBP-SP, as incidências de câncer foram de 283,0 e 304,2 casos novos, respectivamente, por 100.000 homens por

100.000 mulheres. Os coeficientes de mortalidade foram de 116,6 e 99,7 óbitos, respectivamente, por 100.000 homens e por 100.000 mulheres (RCBP-SP, 2011).

Além da incidência e mortalidade, cabe aos RCBP's produzirem informações sobre a sobrevivência, e até o momento, esta análise não pode ser realizada neste registro, pois faltam informações sobre o seguimento destes pacientes. Para esta análise seria necessário conhecer o status vital (vivo ou morto) e a data da última informação e este foi o estímulo para desenvolver o presente trabalho.

Fazia-se necessário atualizar estas informações e, dado o volume de casos que deveriam ser constantemente levantados em prontuários, tal trabalho mostrava-se uma tarefa difícil e onerosa. Destaca-se que no período avaliado, de 1997 a 2005, o RCBP-SP apresentava 65% dos registros sem informação de seguimento, 76,4% de valores ignorados para o endereço e o preenchimento de menos de 1% do nome da mãe. Isso porque estas informações passaram a ser coletadas somente nos últimos cinco anos.

Desta maneira, sabendo-se da validade da técnica de *linkage* na melhora da qualidade da informação e de sua utilidade na completitude de bases de dados de estudos longitudinais com baixo custo operacional, decidiu-se fazer este estudo. Assumiu-se que, o processo de *linkage* possibilitará a estruturação de uma nova base de dados mais consistente e completa, que contribuirá para a organização de um sistema de informação epidemiológico de câncer no Município de São Paulo.

2. OBJETIVOS

2.1 Investigar a possibilidade de completar/aperfeiçoar as informações da base de dados do RCBP-SP, no período de 1997 a 2005, utilizando o processo de *linkage* com três outras bases, a saber: PRO-AIM, APAC-SIA/SUS e Fundação SEADE.

2.2 Calcular os coeficientes brutos de incidência e de mortalidade por câncer utilizando a base de dados do RCBP-SP, no período de 1997 a 2005, antes e após o processo de *linkage*.

2.3 Estimar a probabilidade de sobrevida global acumulada por câncer, segundo sexo, faixa etária e topografia utilizando a base de dados do RCBP-SP, no período de 1997 a 2005, antes e após o processo de *linkage*.

2.4 Propor pontos de corte nos escores calculados nos processos de *linkage* probabilístico da base de dados do RCBP-SP versus PRO-AIM e APAC-SIA/SUS, para as diversas topografias de câncer, calculando sua acurácia na detecção de pares verdadeiros.

3. MATERIAL E MÉTODOS

3.1. Base de dados do Registro de Câncer de Base Populacional do Município de São Paulo (RCBP - SP)

Neste estudo utilizou-se a base de dados do RCBP-SP com casos incidentes registrados no período de 1997 a 2005. Esta base de dados é composta por 343.306 casos incidentes de câncer (C00 a C80.9), com idades que variaram de menos de um a 106 anos, de ambos os sexos, todos residentes no Município de São Paulo.

3.2. Bases de dados utilizadas no processo de *linkage*

Para melhorar a completude das informações da base de dados do RCBP-SP foram utilizadas mais três bases de dados, a saber: 1) do Programa de Aprimoramento de Mortalidade (PRO-AIM), referente aos óbitos ocorridos no Município de São Paulo 2) da Autorização e Procedimentos de Alta Complexidade (APAC-SIA/SUS), referente a todos os procedimentos de alta complexidade e custo realizados no Município de São Paulo, ambas gerenciadas pelo Centro Epidemiológico de Informação do Município de São Paulo (CEINFO) e 3) a base de dados de óbito do Estado de São Paulo fornecida pela Fundação SEADE (FSEADE).

3.3. Preparação das bases de dados

Para o processo de *linkage* probabilístico e determinístico foi necessário realizar uma reestruturação das bases de dados do RCBP-SP, PRO-AIM, APAC-SIA/SUS e FSEADE para que estas apresentassem as mesmas características quanto ao tipo de caractere, formato e tamanho de campo.

A base de dados do RCBP-SP foi composta pelo número do registro, nome completo do paciente, nome da mãe, data de nascimento, idade, sexo, cor, Estado civil, nacionalidade, naturalidade, local de residência, profissão, fonte de notificação, topografia, morfologia, data do diagnóstico, data da última informação e data do óbito.

Para realizar o processamento do *linkage*, ela foi desmembrada em 24 sub-bases de acordo com a Classificação Internacional de Doenças Oncológicas (CID-O) 3ª edição. Esta divisão foi realizada para otimizar os relacionamentos. A tabela que descreve estas sub-bases encontra-se no Anexo 4.

O PRO-AIM forneceu onze bases de dados compreendendo os anos de 1997 a 2007. A escolha deste período foi para dar dois anos além do seguimento do RCBP-SP a fim de captar os possíveis casos de óbitos. Estas bases continham as variáveis: causa básica do óbito, data do óbito, naturalidade, nome do falecido, data de nascimento, sexo, raça/cor, Estado civil, escolaridade, endereço de residência (nome da rua, bairro/distrito, Município de residência, CEP e UF) e nome da mãe.

Estas bases foram unidas no programa *Statistical Package for the Social Sciences-SPSS* versão 17.0 para *Windows*, totalizando 799.052 óbitos. Destes, foram excluídos 24.758 óbitos fetais e 6.542 casos que estavam sem informação ou

apresentavam a palavra “indigente” no campo nome, totalizando 767.752 óbitos. Foram identificados 64 indivíduos com o campo “sexo” vazio. Nestes registros a variável “sexo” foi preenchida baseada no nome do indivíduo e somente 21 ficaram sem identificação do “sexo”, permanecendo o campo em branco, pois o nome poderia ser atribuído tanto para homens quanto para mulheres.

O CEINFO forneceu 54 bases de dados da APAC-SIA/SUS com todos os procedimentos de alta complexidade realizados de agosto de 2003 a dezembro de 2007, realizados no Município de São Paulo. A escolha deste período foi devido a disponibilidade dos dados e os dois anos além do seguimento do RCBP-SP foi para identificar os registros mais recentes destes pacientes. Estas bases continham as variáveis: nome do paciente, data de nascimento, sexo, CPF, nome da mãe, endereço, vila ou bairro, CEP, UF, número do Município, tipo de procedimento/tratamento realizado, data da última informação e motivo da finalização do tratamento (alta ou óbito). Estas bases também foram agregadas em uma única composta por 31.743.533 pacientes. A partir desta etapa, realizou-se o processo de identificação de pessoas que apareceram mais de uma vez na base da APAC-SIA/SUS, dado que a cada tratamento ou renovação de terapia existe um novo registro. Neste trabalho cada paciente teve um único registro com a data mais recente da última informação. Para esta identificação considerou-se a igualdade das variáveis nome do paciente, data de nascimento, CPF - para os que possuíam esta informação- e nome da mãe. Após este processo, a base final constou de 863.735 pacientes.

O *linkage* determinístico entre as bases de dados do RCBP-SP e a de óbito da FSEADE foi realizado nesta fundação. Para o envio da base de dados do RCBP-SP, o primeiro passo foi agrupar as 24 sub-bases. Após esta etapa, foram retirados 91.065

registros pareados no *linkage* entre as 24 sub-bases e a base de dados de óbitos do PRO-AIM. Além destes registros, foram retirados 15.334 registros que apresentavam as informações completas para nome do paciente, data de nascimento, endereço completo e data do óbito. Ao final, a base de dados encaminhada à FSEADE foi composta por 236.907 registros.

As variáveis que integraram a base de dados do RCBP-SP entregue à FSEADE foram: ano de registro do caso novo, número do registro de entrada na base de dados, nome do paciente, data de nascimento, idade, sexo, cor, Estado civil, naturalidade, nacionalidade, endereço de residência, número da residência, complemento da residência, CEP, código do endereço residencial, profissão, data do diagnóstico, topografia, CID-10, morfologia, data da última informação, data do óbito, tipo de óbito, registro da fonte de notificação, número da fonte de notificação, bairro, cidade e Estado da fonte de notificação, telefone e CNES da fonte notificadora.

O primeiro processo realizado pela FSEADE foi a verificação de possíveis duplicidades na base de dados do RCBP-SP. Foram identificados por esta fundação 9.539 registros de possíveis casos duplicados, a partir da combinação das variáveis nome do paciente e data de nascimento.

Todos estes casos foram enviados à pesquisadora que fez a revisão manual. Destes, foram encontrados apenas dezoito casos duplicados, pois continham a mesma informação para as variáveis nome do paciente, data de nascimento, data do diagnóstico, CID-10 e morfologia.

A base de dados da FSEADE relacionada com o RCBP-SP foi composta pelo anos de 2000 a 2009, totalizando 2.308.061 óbitos. A escolha deste período foi

devido a disponibilidade dos dados, dado que a informação sobre os óbitos ocorridos no Estado de São Paulo começou a ser informatizada no ano de 2000. Para o *linkage* entre as bases de dados destas instituições a base de dados da FSEADE apresentou as seguinte divisões por ano de registro:

- 2000 - 127.111 registros;
- 2001 - 173.195 registros;
- 2002 - 238.877 registros;
- 2003 a 2008 - 1.505.804 registros;
- 2009 - 263.074 registros.

Observa-se que, para os anos de 2000 e 2001, há um número inferior de registros, pois muitos cartórios fora do Município paulista não eram informatizados. Com isso, os óbitos registrados nestes anos, são, em sua maioria, pertencentes ao Município de São Paulo.

O relacionamento entre as bases de dados do RCBP-SP e a base de dados de óbito da FSEADE não será restrito a esta pesquisa, pois foi assinado um Termo de Cooperação Técnica (Anexo 5) em 2011 entre estas instituições para que, anualmente, ocorra este processo de *linkage*.

3.4 Padronização

Com as quatro bases de dados prontas (RCBP-SP, PRO-AIM, APAC-SIA/SUS e FSEADE), deu-se início à padronização das mesmas em relação às variáveis nome do paciente, data de nascimento, sexo, data do óbito e causa básica.

Esta etapa foi executada nos programas SPSS versão 17.0 para *Windows* e Reclink III versão 3.1.6, para as bases de dados do RCBP-SP, PRO-AIM e APAC-SIA/SUS e para a base de dados da FSEADE todo o processo foi desenvolvido em *Visual Basic*, que conecta as bases hospedadas em *SQL Server*.

O processo de padronização foi realizado para homogeneizar as variáveis das bases, visando minimizar erros no processo de pareamento. O detalhamento destes processos estão nos Quadros 1 e 2.

A estratégia consistiu em transformar todos os caracteres numéricos em alfanuméricos, alfanuméricos minúsculos em maiúsculos, suprimir letras duplicadas, retiradas de caracteres diferentes de A-Z, retirar qualificações presentes nos campos de nome e sobrenomes (por exemplo, "já falecido", "brasileiros", "indigente"), eliminar espaços em branco no início dos campos, retirar qualquer tipo de pontuação (% , ; , ! , ?) e preposição e a substituição de letras.

Somente a variável causa básica não foi padronizada no sistema da FSEADE, pois existe uma norma para sua utilização, segundo a CID.

Quadro 1 – Padronizações segundo bases de dados, RCPB-SP, PRO-AIM e APAC-SIA/SUS.

Campos	Objetivos	RCPB-SP características		PRO-AIM Características		APAC características		Programa
		Original	Final	Original	final	original	final	
Sexo	Transformar em caracteres alfanuméricos de mesmo formato	1	1	M	1	M	1	SPSS
		2	2	F	2	F	2	
Data de nascimento	Transformar o tipo de variável (data para alfanumérico)	dd/mm/aaaa	dd/mm/aaaa					SPSS
	Retirar as barras e alterar a posição dos campos	dd/mm/aaaa	aaaammdd	ddmmaaaa	aaaammdd	ddmmaaaa	aaaammdd	Reclink
Ano de nascimento	Extrair o ano da data de nascimento	dd/mm/aaaa	aaaa					SPSS
				aaaammdd	aaaa	aaaammdd	aaaa	Reclink
	Gerar o ano de nascimento a partir de duas variáveis	ano registro - idade	aaaa					SPSS
Nome do paciente	Transformar caracteres minúsculos em maiúsculos	Maria	MARIA	Idem	Idem	idem	Idem	Reclink
	Transformar primeiro nome em código fonético	MARIA	M001	Idem	Idem	idem	Idem	
	Exemplo de grafias diferentes para mesmo fonema	WALTER	V002	Idem	Idem	idem	Idem	
		VALTER	V002	Idem	Idem	idem	Idem	
	Transformar último nome em código fonético	SILVA	S0101	Idem	Idem	idem	Idem	
Eliminar preposições	MARIA DA SILVA	MARIA SILVA	Idem	Idem	idem	Idem		

Quadro 2 - Padronizações da base de dados do RCBP-SP para sua vinculação à base de dados de óbitos da FSEADE.

Campos	Objetivos	RCBP-SP e FSEADE características	
		Original	Final
Data de nascimento	A padronização desta variável consiste em converter este campo para texto em outro formato	dd/mm/aaaa	ddmmaaaa
Nome do paciente	Transformar caracteres minúsculos em maiúsculos	Maria	MARIA
	Suprimir letrar duplicadas	ESTELLA	ESTELA
	Retirar caracteres diferentes de A-Z	MARI%A	MARIA
	Substituição das letras Y e W	YARA	IARA
	Substituição de duas ou mais letras que produzem um único som	THAMIRES	TAMIRES
		OCTAVIO	OTAVIO
	Eliminar preposições	MARIA DA SILVA	MARIA SILVA
Criar variáveis derivadas	MARIA DA SILVA	MARIA	
Data de óbito	A padronização desta variável consiste em converter este campo para texto em outro formato	dd/mm/aaaa	ddmmaaaa

3.4.1 Variável: sexo

Na base de dados do RCBP-SP, a variável sexo estava em caractere numérico representada pelos algarismos arábicos 1 e 2, respectivamente, para os sexos masculino e feminino. Neste caso, foram mantidos os algarismos, porém transformados em caracteres alfanuméricos.

Nas bases do PRO-AIM e APAC-SIA/SUS o campo sexo estava com caracteres alfanuméricos representados pelas letras M e F. Manteve-se o tipo de caractere e estes foram transformados em 1 para masculino e 2 para feminino.

3.4.2 Variável: data de nascimento

Para as datas de nascimento registradas no RCBP-SP foram realizadas duas padronizações. Numa primeira etapa, executada pelo programa SPSS 17.0 para *Windows*, as datas estavam formatadas como *data*, dispostas em dia, mês e ano (dd/mm/aaaa) e foram transformadas em caracteres alfanuméricos com a mesma disposição. Após este processo, esta variável recebeu outra padronização pelo programa Reclink III versão 3.1.6, no qual foram extraídas as barras para divisão entre dia, mês e ano e foram invertidas as posições em ano, mês e dia (aaaammdd).

A variável data de nascimento registrada tanto na bases de dados do PRO-AIM quanto na base da APAC-SIA/SUS estava formatada como alfanumérica, sem as barras divisórias. Contudo, foram alteradas as posições originais dia, mês e ano (ddmmaaaa) para ano, mês e dia (aaaammdd).

Na padronização executada pela FSEADE, as datas de nascimento foram convertidas da formatação *data* (dd/mm/aaaa) para campo texto em ddmmaaaa (dia, mês e ano).

3.4.3 Variável: ano de nascimento

A variável ano de nascimento foi gerada de duas formas na base do RCBP-SP. Primeiramente, para os casos em que havia a data de nascimento foram extraídos os quatro dígitos correspondentes ao ano. Nos casos em que não havia esta informação, esta variável foi gerada a partir da subtração da idade do paciente do ano do registro da base, chegando-se ao valor aproximado do ano de nascimento.

Nas bases do PRO-AIM e APAC-SIA/SUS foi realizada a extração do ano de nascimento diretamente pelo programa Reclink III versão 3.1.6.

3.4.4 Variável: data do óbito

Na padronização executada pela FSEADE, as datas de óbito foram convertidas da formatação *data* (dd/mm/aaaa) para campo texto em ddmmaaaa (dia, mês e ano). Para os anos de 2003 a 2008 foi criada a variável derivada ano do óbito composta por quatro caracteres em formato texto (aaaa).

3.4.5 Variável: Nome do paciente

O processo de padronização do nome do paciente foi realizado pelo programa Reclink III versão 3.16 pela opção “Subdivide Nome”. Neste processo foram retirados sinais de pontuação, preposições e acentos. Todos os nomes foram transformados em caixa alta e, após este processo, foram criadas seis variáveis a partir do nome completo.

Das seis variáveis, duas foram formadas pelo primeiro e último nome (FNOMEP e FNOMEU, respectivamente), mantendo-se as grafias originais. Outras duas variáveis foram formadas pelas iniciais do nome do meio (FNOMEI) e pelos sobrenomes qualificadores de geração (FNOMEA), isto é, JÚNIOR, NETO, FILHO e SOBRINHO. Por exemplo, Maria do Rosário Dias de Oliveira Latorre, ficaria MARIA RDO LATORRE. Quando não há, no nome da pessoa, sobrenome do meio

ou qualificador de geração, na base de dados, o campo destas variáveis se apresenta com o código asterisco (*).

As duas últimas variáveis geradas são denominadas como PBLOCO e UBLOCO. Estas variáveis são transformadas em códigos fonéticos (*soundex*), constituídos por quatro dígitos, sendo o primeiro uma letra e, os demais, números. Por exemplo, MARIA se transforma em M600. O código *soundex* permite que grafias diferentes, mas com mesmo som, tenham o mesmo formato; assim MANOEL ou MANUEL tornam-se M540.

Na padronização executada pela FSEADE foram suprimidas todas as duplicações de letras, por exemplo BB para B. Também foram retiradas preposições, caracteres não pertencentes ao alfabeto, espaços em branco e as qualificações presentes nos campos, por exemplo “indigente”. O nome e o sobrenome dos pacientes foram transformados em caixa alta, as letras Y e W foram transformadas, respectivamente, em I e V, TH por T, OCT por OT, etc.

Neste processo, também foram criadas as variáveis derivadas "iniciais do nome do paciente", "primeiro nome do paciente" e "segundo nome do paciente". Estas variáveis foram construídas a partir da variável nome do paciente. Nessa rotina, o sistema recodifica a variável contendo as primeiras letras do nome e dos sobrenomes de cada indivíduo, no caso das iniciais, e duas outras variáveis contendo somente o primeiro e o segundo nome dos pacientes.

3.4.6 Rotinas

A descrição das rotinas utilizadas no processo de limpeza e padronização das variáveis executadas no programa SPSS versão 17.0 para *Windows* encontra-se no Anexo 6. Optou-se por colocá-la em anexo, pois é extensa.

As rotinas utilizadas para padronização das variáveis pelo programa Reclink III versão 3.1.6 estão disponíveis no tutorial Reclink III: relacionamento probabilístico de registros (CAMARGO JR E COELI, 2007).

3.5 A execução do *linkage* probabilístico

Para o *linkage* probabilístico foram realizados os processos de blocagem e pareamento, descritos a seguir.

3.5.1 Blocagem

O processo de blocagem consiste na criação de blocos lógicos, mutuamente exclusivos e podem ser utilizadas diversas estratégias. A estratégia de blocagem é a união de mais de um bloco, onde cada bloco é representado por uma chave. Estas estratégias são empíricas e da escolha do pesquisador. Tenta-se unir as variáveis com menor percentual de erro, como sexo, nome e ano de nascimento e esta recomendação consta no tutorial do programa Reclink III (CAMARGO JR. e COELI, 2007). Neste trabalho foram testadas dez estratégias de blocagem descritas no Quadro 3. Por exemplo, para Maria do Rosário Dias de Oliveira Latorre, nascida

em 31 de outubro de 1955, a estratégia um (1) seria PBLOCO = M600, UBLOCO = L2043 e SEXO = 2. Para a estratégia dez (10) a formação da estratégia de blocagem seria FPRIMEIRONOME = MARIA, FULTIMONOME = LATORRE, SEXO = 2 e ANONASCIMENTO = 1995.

Quadro 3 - Estratégias de Blocagem.

Estratégias	Chaves de Blocagem
1.	PBLOCO + UBLOCO + SEXO
2.	PBLOCO + SEXO
3.	UBLOCO + SEXO
4.	PBLOCO + UBLOCO + SEXO + ANO DE NASCIMENTO
5.	PBLOCO + SEXO + ANO DE NASCIMENTO
6.	UBLOCO + SEXO + ANO DE NASCIMENTO
7.	FPRIMEIRO NOME + FULTIMO NOME + SEXO
8.	FPRIMEIRO NOME + SEXO + ANO DE NASCIMENTO
9.	FULTIMO NOME + SEXO + ANO DE NASCIMENTO
10.	FPRIMEIRO NOME + FULTIMO NOME + SEXO + ANO DE NASCIMENTO

3.5.2 Pareamento

Para dar início a esta fase, foram definidos os campos de pareamento para a construção dos escores que verificam o quão verossimilhante os registros pareados pertencem à mesma pessoa. Neste estudo, os campos escolhidos foram nome do paciente, data de nascimento e ano de nascimento. Destaca-se que o nome utilizado está com a grafia da base de dados original.

A partir disto, calculou-se um escore, que é a soma ponderada dos campos empregados no processo de pareamento. Os pesos foram construídos a partir dos conceitos de sensibilidade e especificidade de testes diagnósticos. A sensibilidade representa a probabilidade de identificar um par verdadeiro quando este é realmente verdadeiro e a especificidade representa a probabilidade de identificar um não par quando este é realmente não par.

O programa Reclink III versão 3.1.6 possui valores iniciais destas probabilidades, estimados em outro estudo (COELI e CAMARGO JR, 2007), descritos no Quadro 4. Com estes valores iniciais (chamados de sementes) aplicados nas bases de dados desta pesquisa, foram definidos novos valores de sensibilidade e 1-especificidade para a geração dos escores entre as bases de dados RCBP-SP *versus* PRO-AIM e RCBP-SP *versus* APAC-SIA/SUS.

Para o cálculo destas probabilidades, as bases de dados do PRO-AIM e APAC-SIA/SUS foram rodadas integralmente e da base de dados RCBP-SP foi retirada uma amostra de 10% de casos, proporcional aos anos analisados e às topografias.

Quadro 4 – Parâmetros para a construção dos escores de pareamento.

Bases de dados	Variáveis	Algoritmo	Sensibilidade (m_i)	1-especificidade (u_i)	Proporção mínima de concordância
Valores sementes (Tutorial Reclink)	Nome	Aproximado	92%	1%	85%
	Data de Nascimento	Caractere	90%	5%	65%
	Ano de Nascimento	Diferença	92%	4%	± 2
RCBP x PRO-AIM	Nome	Aproximado	96,2%	0,001%	85%
	Data de Nascimento	Caractere	92,5%	1,1%	65%
	Ano de Nascimento	Diferença	95,0%	6,4%	± 2
RCBP x APAC- SIA/SUS	Nome	Aproximado	87,1%	0,001%	85%
	Data de Nascimento	Caractere	83,0%	1,0%	65%
	Ano de Nascimento	Diferença	89,5%	5,3%	± 2

Após definidas as probabilidades foram gerados dois fatores de ponderação - concordância e discordância - calculados como logaritmo de base dois da razão de verossimilhança entre as probabilidades:

1) concordância

$$m_i \text{ e } u_i \quad (wc_i = \log_2 \left(\frac{m_i}{u_i} \right))$$

2) discordância

$$1 - u_i \quad (wd_i = \log_2 \left(\frac{(1 - m_i)}{(1 - u_i)} \right))$$

onde m_i = sensibilidade e u_i = 1- especificidade.

Assim, ao se comparar o campo nome de uma base de dados com o mesmo campo, em outra, e estes concordarem, o fator de ponderação aplicado será o de concordância. Para a situação inversa, aplica-se o fator de discordância. Contudo, devido a erros nos campos de comparação, muitas vezes decorrentes do processo de coleta e digitação de dados, utiliza-se o parâmetro denominado limiar aproximado, isto é, uma concordância mínima, que permite atribuir valores menores do que o observado para o escore total. Os fatores de concordância e discordância obtidos nesta pesquisa estão descritos no Quadro 5.

Quadro 5 – Fatores de ponderação de concordância e discordância.

Base de dados	Variáveis	Pesos da concordância	Pesos da discordância
RCBP x PRO-AIM	Nome	16,8947	-4,70399
	Data de Nascimento	6,3953	-3,7279
	Ano de Nascimento	3,88547	-4,24028
RCBP x APAC-SIA/SUS	Nome	16,6247	-2,96648
	Data de Nascimento	6,42096	-2,54357
	Ano de Nascimento	4,08494	-3,16782

A partir das estratégias de blocagem e dos valores para os escores de pareamento, foram definidos catorze estratégias que estão no Quadro 6.

Quadro 6 – Descrição dos campos de blocagem e dos campos de pareamento e os escores máximos e mínimos calculados.

	Chaves de Blocagem	Campo de Pareamento	Escores			
			RCBP-SP vs PRO-AIM		RCBP-SP vs APAC-SIA/SUS	
			Máximos	Mínimos	Máximos	Mínimos
1.	PBLOCO + UBLOCO + SEXO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
2.	PBLOCO + SEXO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
3.	UBLOCO + SEXO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
4.	FPRIMEIRO NOME + FULTIMO NOME + SEXO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
5.	PBLOCO + UBLOCO + SEXO +ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
6.	PBLOCO + SEXO +ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
7.	UBLOCO + SEXO +ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
8.	FPRIMEIRO NOME + FULTIMO NOME + SEXO + ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
9.	FPRIMEIRO NOME + SEXO + ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
10.	FULTIMO NOME + SEXO + ANO	NOME + DATA DE NASCIMENTO	23,28	-8,43	23,04	-5,51
11.	PBLOCO + UBLOCO + SEXO	NOME + ANO DE NASCIMENTO	20,78	-8,94	20,71	-6,13
12.	PBLOCO + SEXO	NOME + ANO DE NASCIMENTO	20,78	-8,94	20,71	-6,13
13.	UBLOCO + SEXO	NOME + ANO DE NASCIMENTO	20,78	-8,94	20,71	-6,13
14.	FPRIMEIRO NOME + FULTIMO NOME + SEXO	NOME + ANO DE NASCIMENTO	20,78	-8,94	20,71	-6,13

Definidas as estratégias de blocagem e os parâmetros de relacionamento, iniciou-se o processo de *linkage* entre dois arquivos. Nesta pesquisa, o primeiro relacionamento foi feito entre a base de dados do RCBP-SP e a base de dados do PRO-AIM e o segundo relacionamento entre a base de dados do RCBP-SP, sem os óbitos identificados no processamento anterior, e a base de dados da APAC-SIA/SUS.

Neste processo é criado um arquivo de relacionamento com as variáveis “posição do registro identificado na base de dados do RCBP-SP” e “posição do registro identificado na base de dados relacionadas” (PRO-AIM ou APAC-SIA/SUS). A partir deste arquivo, é gerado um arquivo denominado como “combinado”. Neste arquivo, as variáveis originais e padronizadas das bases de dados relacionadas são colocadas lado a lado. Também estão presentes, as variáveis que indicam a posição dos registros nas bases de dados padronizadas e o valor do score calculado.

As estratégias de blocagem e pareamento foram aplicadas nos relacionamentos do RCBP-SP (24 sub-bases) *versus* PRO-AIM e RCBP-SP (24 sub-bases) *versus* APAC-SIA/SUS. Para a geração do arquivo “combinado” do *linkage* entre as bases de dados desta pesquisa, foram eliminados do arquivo de “relacionamento” os escores de discordância. Ao final, foram criados arquivos com identificação de pares verdadeiros ou falsos. Para a classificação dos pares verdadeiros e falsos, foram conferidas, manualmente, as variáveis: nome do paciente, data de nascimento, endereço, causa básica de óbito e tratamento/procedimento realizado.

Foram considerados verdadeiros os pares que apresentassem, ao menos, uma das seguintes condições:

- 1) Nome, sobrenome e data de nascimento iguais, em nomes pouco comuns (isto é, João, José, Maria, Aparecida, Santos, Souza, Silva, etc). Os nomes e sobrenomes poderiam apresentar grafias diferentes.
- 2) Nome, sobrenome, dia e mês da data de nascimento iguais e o último dígito do ano de nascimento com variação de dois anos para mais ou para menos, em nomes pouco comuns. Os nomes e sobrenomes poderiam apresentar grafias diferentes.
- 3) Nome, sobrenome e ano de nascimento iguais, com dia e mês invertidos, com endereço compatível, em nomes pouco comuns. Os nomes e sobrenomes poderiam apresentar grafias diferentes.
- 4) Nome e sobrenome iguais, sem data de nascimento, com endereço compatível e causa básica de óbito ou tratamento por câncer, em nomes pouco comuns. Os nomes e sobrenomes poderiam apresentar grafias diferentes.
- 5) Nome e sobrenome iguais, sem data de nascimento, com idade compatível e endereço compatível e data do óbito igual, em nomes pouco comuns. Os nomes e sobrenomes poderiam apresentar grafias diferentes.
- 6) Nome e três ou mais sobrenome iguais, sem data de nascimento, com idade compatível, endereço compatível, data do óbito igual ou causa básica do óbito ou tratamento por câncer. Os nomes e sobrenomes poderiam apresentar grafias diferentes.

Por fim, foram identificadas as melhores estratégias, considerando o maior número de pares verdadeiros pareados, o menor tempo de execução para o relacionamento, menor número de duplicidades e a precisão, dado pelo escore máximo alcançado. Destaca-se que a cada novo processo de *linkage* as bases de dados eram relacionadas integralmente. Os 48 quadros gerados nesta etapa estão no Anexo 7. Para exemplificar, no Quadro 7, está o resultado do relacionamento RCBP-SP (cólón) *versus* PRO-AIM.

Considerando o número de pares verdadeiros, a estratégia 1 identificou 6.132 casos, com 36,46% de sucesso na leitura manual, levando 40:56 minutos neste processo, sendo esta estratégia adequada para esta localização de câncer. As estratégias 12 e 13, que têm maior percentual de sucesso, respectivamente, 47,6% e 42,3%, no entanto, só localizam, respectivamente, 1.090 e 1.749 pares verdadeiros. Destaca-se que o número de duplicidades gerada pela estratégia 1 não foi o maior (Quadro 7).

Quadro 7 - RCBP-SP *versus* PRO-AIM para a topografia C18.0 a C18.9 (colón). RCBP-SP, 1997 a 2005.

Colón (n=16.890)											
	Estratégias	tempo (hh:mm:ss)		Total	n	n de pares	Par	% de pares	%	duplicidades	escore máximo
		Relaciona	Combina	Minutos	linkado	no escore máximo	Verdade	no escore máximo	sucesso		
1.	P+U+S/N+DT	0:19:58	0:20:58	0:40:56	16819	3501	6132	20.82	36.46	85	23,28
2.	P+S/N+DT	0:27:01	0:21:06	0:48:07	15250	689	1131	4.52	7.42	5	23,28
3.	U+S/N+DT	0:27:04	0:20:06	0:47:10	14156	957	1826	6.76	12.90	16	23,28
4.	Fp+Fu+S/N+DT	0:18:10	0:20:43	0:38:53	13643	3651	4547	26.76	33.33	9	23,28
5.	P+U+S+A/N+DT	0:17:47	0:21:12	0:38:59	4192	2587	355	61.71	8.47	1	16,69
6.	P+S+A/N+DT	0:26:06	0:19:36	0:45:42	493	244	158	49.49	32.05	5	16,69
7.	U+S+A/N+DT	0:26:49	0:19:52	0:46:41	681	362	263	53.16	38.62	8	16,69
8.	Fp+Fu+S+A/N+DT	0:17:23	0:20:47	0:38:10	3765	2629	325	69.83	8.63	1	16,69
9.	Fp+S+A/N+DT	0:25:14	0:20:15	0:45:29	593	322	195	54.30	32.88	1	16,69
10.	Fu+S+A/N+DT	0:23:39	0:20:26	0:44:05	789	470	339	59.57	42.97	6	16,69
11.	P+U+S/N+A	0:19:40	0:21:26	0:41:06	12663	3656	5903	28.87	46.62	127	20,78
12.	P+S/N+A	0:27:01	0:20:12	0:47:13	1492	710	1090	47.59	73.06	12	20,78
13.	U+S/N+A	0:27:01	0:20:45	0:47:46	2307	976	1749	42.31	75.81	17	20,78
14.	Fp+Fu+S/N+A	0:18:25	0:21:22	0:39:47	10758	3814	4570	35.45	42.48	84	20,78

Legenda: P = PBLOCO; U = UBLOCO; S = SEXO; N = NOME; DT = DATA DE NASCIMENTO; Fp = FPRIMEIRO NOME; Fu = FULTIMO NOME; A = ANO DE NASCIMENTO.

Para a aplicação das catorze estratégias entre as bases de dados do RCBP-SP e APAC-SIA/SUS foram retirados todos os pares verdadeiros obtidos na primeira estratégia que consiste na combinação da chave de blocagem PBLOCO (primeiro nome em código fonético) + UBLOCO (último nome em código fonético) + S (sexo) com os campos de comparação N (nome do paciente) + DT (data de nascimento) das 24 sub-bases relacionadas com o PRO-AIM. Com as 24 sub-bases reduzidas, foi iniciado o processo para o *linkage* com a base de dados da APAC-SIA/SUS.

Os Quadros 8 e 9 apresentam, respectivamente, as estratégias escolhidas em cada *linkage* entre o RCBP-SP versus PRO-AIM e RCBP-SP versus APAC-SIA/SUS.

Quadro 8 – Estratégias para o processo de *linkage* probabilístico, segundo desempenho no relacionamento entre RCBP-SP *versus* PRO-AIM.

Topografia	Estratégia	Tempo total (mm:ss)	Par verdadeiro	% sucesso	Número duplicidades
C00.0 a C09.9	P+U+S/N+DT	14:40	2447	21,81	3
C10.0 a 11.9	P+U+S/N+DT	12:29	813	16,08	16
C12.0 a 14.8	P+U+S/N+DT	37:16	644	13,13	9
C15.0 a C15.9	P+U+S/N+DT	13:28	2687	23,23	10
C16.0 a C17.9	P+U+S/N+DT	14:26	7370	37,17	20
C18.0 a C18.9	P+U+S/N+DT	40:56	6132	26,46	85
C19.0 a C21.8	P+U+S/N+DT	13:43	3312	27,97	5
C22.0 a C24.9	P+U+S/N+DT	15:20	2307	26,99	4
C25.0 a C26.9	P+U+S/N+DT	15:41	3831	31,73	8
C30.0 a C33.9	P+U+S/N+DT	15:44	2060	21,51	7
C34.0 a C34.9	P+U+S/N+DT	15:15	9204	42,09	18
C37.0 a C39.9	P+U+S/N+DT	14:20	415	13,10	0
C40.0 a C42.4	P+U+S/N+DT	15:29	4874	35:89	8
C44.0 a C44.9 (8720 a 8780)	P+U+S/N+DT	15:30	803	13,98	3
C44.0 a C44.9	P+U+S/N+DT	17:43	4113	19,44	4
C47.0 a C49.9	P+U+S/N+DT	13:47	841	16,11	3
C50.0 a C50.9	P+U+S/N+DT	15:10	6989	52,49	16
C51.0 a C58.9	P+U/N+DT	38:25	7728	2,86	13
C60.0 a C60.9 e C62.0 a C63.9	P+U/N+DT	15:22	387	2,92	0
C61.9	P+U/N+DT	27:57	7731	3,49	10
C64.0 a C68.9	P+U+S/N+DT	16:29	3417	27,04	4
C69.0 a C72.9	P+U+S/N+DT	15:58	2980	27,16	11
C73.0 a C75.9	P+U+S/N+DT	16:02	741	14,60	3
C76.0 a C80.9	P+U+S/N+DT	15:50	7902	44,19	7

Quadro 9 – Estratégias para o processo de *linkage* probabilístico, segundo desempenho no relacionamento entre RCBP-SP *versus* APAC-SIA/SUS.

Topografia	Estratégia	Tempo total (hh:mm:ss)	Par verdadeiro	% sucesso	Número duplicidades
C00.0 a C09.9	P+U+S/N+A	15:50	580	2,47	83
C10.0 a 11.9	P+U+S/N+A	13:14	87	0,77	0
C12.0 a 14.8	P+U+S/N+DT	15:13	57	1,82	0
C15.0 a C15.9	P+U+S/N+DT	15:58	145	2,07	2
C16.0 a C17.9	P+U+S/N+A	17:31	393	1,50	8
C18.0 a C18.9	P+U+S/N+A	15:33	457	2,69	28
C19.0 a C21.8	P+U+S/N+DT	14:22	357	5,05	1
C22.0 a C24.9	P+U+S/N+A	13:24	61	0,48	0
C25.0 a C26.9	P+U+S/N+A	13:26	51	0,33	0
C30.0 a C33.9	P+U+S/N+A	16:54	360	1,91	12
C34.0 a C34.9	P+U+S/N+DT	15:41	556	6,74,	107
C37.0 a C39.9	P+U+S/N+DT	12:44	21	1,06	0
C40.0 a C42.4	P+U+S/N+DT	16:16	574	6,69	81
C44.0 a C44.9 (8720 a 8780)	P+U+S/N+DT	18:06	150	3,01	1
C44.0 a C44.9	P+U+S/N+DT	16:32	1032	5,85	18
C47.0 a C49.9	P+U+S/N+A	14:40	231	1,57	32
C50.0 a C50.9	P+U+S/N+DT	17:22	3419	21,28	138
C51.0 a C58.9	P+U/N+DT	06:22:35	1843	0,62	104
C60.0 a C60.9 e C62.0 a C63.9	P+U/N+A	13:51	167	0,52	1
C61.9	P+U/N+DT	55:28	1569	0,12	15
C64.0 a C68.9	P+U+S/N+A	06:49	308	1,29	0
C69.0 a C72.9	P+U+S/N+DT	16:39	276	4,02	8
C73.0 a C75.9	P+U+S/N+A	14:48	273	1,27	2
C76.0 a C80.9	P+U+S/N+DT	15:18	613	5,92	4

3.6 A execução do *linkage* determinístico

O *linkage* determinístico foi utilizado no relacionamento da base de dados do RCBP-SP com a base de óbitos do Estado de São Paulo, realizado pela FSEADE. O

linkage com esta base de dados teve como finalidade encontrar possíveis óbitos fora do Município de São Paulo.

Neste processo de *linkage*, foram elaborados dois critérios para a seleção de pares verdadeiros, utilizando as variáveis da etapa de padronização (nome do paciente, iniciais do nome do paciente, primeiro nome, segundo nome, data de nascimento, data do óbito e causa básica do óbito). Estes critérios possuem dois tipos de comparação: por igualdade e por semelhança. A comparação por igualdade seleciona as variáveis com preenchimento idêntico nas bases relacionadas e, na comparação por semelhança, o preenchimento deve apresentar um mínimo de concordância de 80% entre as variáveis relacionadas nas duas bases (MORAIS et al., 2010).

A seleção dos casos é realizada após a execução dos critérios que sugerem os possíveis pares a serem vinculados. Os critérios elaborados para esta pesquisa estão descritos no Quadro 10.

Após o *linkage*, de acordo com os 25 critérios estabelecidos, os pares pré-selecionados foram transferidos para uma planilha em Excel, na qual continha os registros vinculados para a seleção dos pares verdadeiros. Esta seleção foi feita de duas maneiras, automática e visual.

A seleção automática ocorreu quando todas as variáveis definidas para o processo de *linkage* coincidiram totalmente. Nesta pesquisa, esta seleção foi utilizada para o Critério 1 (Quadro 10). Por outro lado, quando um par selecionado apresentou uma concordância mínima de 80% entre as variáveis selecionadas para o processo de *linkage*, foi utilizada a análise visual, realizada pela pesquisadora.

Quadro 10 - Critérios elaborados para o *linkage* entre a base de dados do RCBP-SP e a FSEADE.

Critério	Definição
1	Nome do paciente, data de nascimento, data de óbito, causa básica
2	Iniciais do nome do paciente, 1º Nome do paciente, data de nascimento, data de óbito, causa básica
3	Iniciais do nome do paciente, 2º Nome do paciente, data de nascimento, data de óbito, causa básica
4	Nome do paciente, data de nascimento, data de óbito
5	Iniciais do nome do paciente, 1º Nome do paciente, data de nascimento, data de óbito
6	Iniciais do nome do paciente, 2º Nome do paciente, data de nascimento, data de óbito
7	Nome do paciente, data de nascimento, causa básica
8	Iniciais do nome do paciente, 1º Nome do paciente, data de nascimento, causa básica
9	Iniciais do nome do paciente, 2º Nome do paciente, data de nascimento, causa básica
10	Nome do paciente, data de óbito, causa básica
11	Iniciais do nome do paciente, 1º Nome do paciente, data de óbito, causa básica
12	Iniciais do nome do paciente, 2º Nome do paciente, data de óbito, causa básica
13	Nome do paciente, data de óbito
14	Iniciais do nome do paciente, 1º Nome do paciente, data de óbito
15	Iniciais do nome do paciente, 2º Nome do paciente, data de óbito
16	Data de nascimento, data de óbito, nome do paciente com 80% de grau de semelhança
17	Data de nascimento, causa básica, nome do paciente com 80% de grau de semelhança
18	Data de óbito, causa básica, nome do paciente com 80% de grau de semelhança
19*	Nome do paciente, data de nascimento, ano do óbito
20	Nome do paciente, data de nascimento
21	Iniciais do nome do paciente, 1º Nome do paciente, data de nascimento
22	Iniciais do nome do paciente, 2º Nome do paciente, data de nascimento
23	Data de nascimento, data de óbito, causa básica
24	Data de nascimento, data de óbito
25*	Iniciais do nome do paciente, 1º Nome do paciente, 2º Nome do paciente, causa básica, ano do óbito

*Esses critérios foram executados para a base de óbitos referentes aos anos de 2003 e 2008

Nesta análise visual, além da leitura das variáveis utilizadas no processo de *linkage*, a variável endereço de residência foi utilizada como critério para a confirmação de pares verdadeiros, além da variáveis causa básica de óbito, data e ano do óbito.

O tempo total para o processamento do *linkage* determinístico foi de 40 horas, sendo oito horas de trabalho em cinco dias úteis. Para a verificação visual, a pesquisadora levou, aproximadamente, 50 horas de trabalho.

Na Tabela 1 estão apresentados o registros vinculados e os pares verdadeiros encontrados nesta etapa, segundo ano da base de dados de óbitos do Estado de São Paulo.

Tabela 1 - Número de registros vinculados e de pares verdadeiros entre o *linkage* da base de dados do RCBP-SP e a base de dados da FSEADE, segundo ano de ocorrência do óbito.

Ano de vinculação	Registros vinculados para	Registros de pares
	a seleção de pares	verdadeiros
	n	n
2000	3.420	2.540
2001	1.882	1.297
2002	3.568	2.615
2003 a 2008	18.571	11.936
2009	3.039	2.075

3.7 Análise Estatística

Foram calculados os coeficientes brutos de incidência e mortalidade por câncer antes e após o *linkage*. No denominador, considerou-se a população total do Município de São Paulo de cada ano, extraída da página *on-line* da FSEADE (Tabela 2) e, no numerador, o total de casos do mesmo ano.

Tabela 2 – População do Município de São Paulo, entre os anos de 1997 a 2005.

Ano	População
1997	10.134.220
1998	10.233.627
1999	10.333.057
2000	10.426.384
2001	10.525.367
2002	10.613.691
2003	10.698.381
2004	10.782.296
2005	10.865.573

Fonte: FSEADE, 2011

A análise de sobrevida global foi realizada pelo estimador do produto limite de *Kaplan-Meier* e na comparação entre as curvas utilizou-se o teste de *log rank*. O tempo de sobrevida foi calculado como a diferença entre a data da última informação ou data do óbito e a data de diagnóstico e foi assumido um nível descritivo de 5%.

Para o total da base de dados foi estimada a probabilidade de sobrevida acumulada até quatro anos. Para aqueles pacientes com tempo de seguimento maior que quatro anos o tempo foi truncado e, todos foram considerados como censura.

Também foi estimada a probabilidade de sobrevida acumulada após sete anos. Nesta análise, o tempo foi truncado em sete anos, sendo os pacientes que vieram a óbito após este período considerados como censura. Nesta última análise, foram considerados apenas os pacientes com diagnóstico de câncer entre 1997 a 2002.

Foram propostos pontos de corte nos escores calculados nos processos de *linkage* probabilístico das bases de dados do RCBP-SP versus PRO-AIM e APAC-SIA/SUS, para as diversas topografias de câncer, calculando sua acurácia na detecção de pares verdadeiros. Foram calculados os valores da área sob a curva - *Receiver Operating Characteristic* (ROC) - e a sensibilidade e especificidade para determinar o ponto de corte do escore de maior precisão na identificação dos pares verdadeiros.

Estas análises foram realizadas no programa SPSS versão 17.0 para *Windows*.

3.8 Considerações Éticas

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da Universidade de São Paulo (Anexo 8) - protocolo 0086.0.207.000-08 e pela Secretaria Municipal da Saúde (SMS) – protocolo 0064.0.162.000-09 (Anexo 9).

4. RESULTADOS

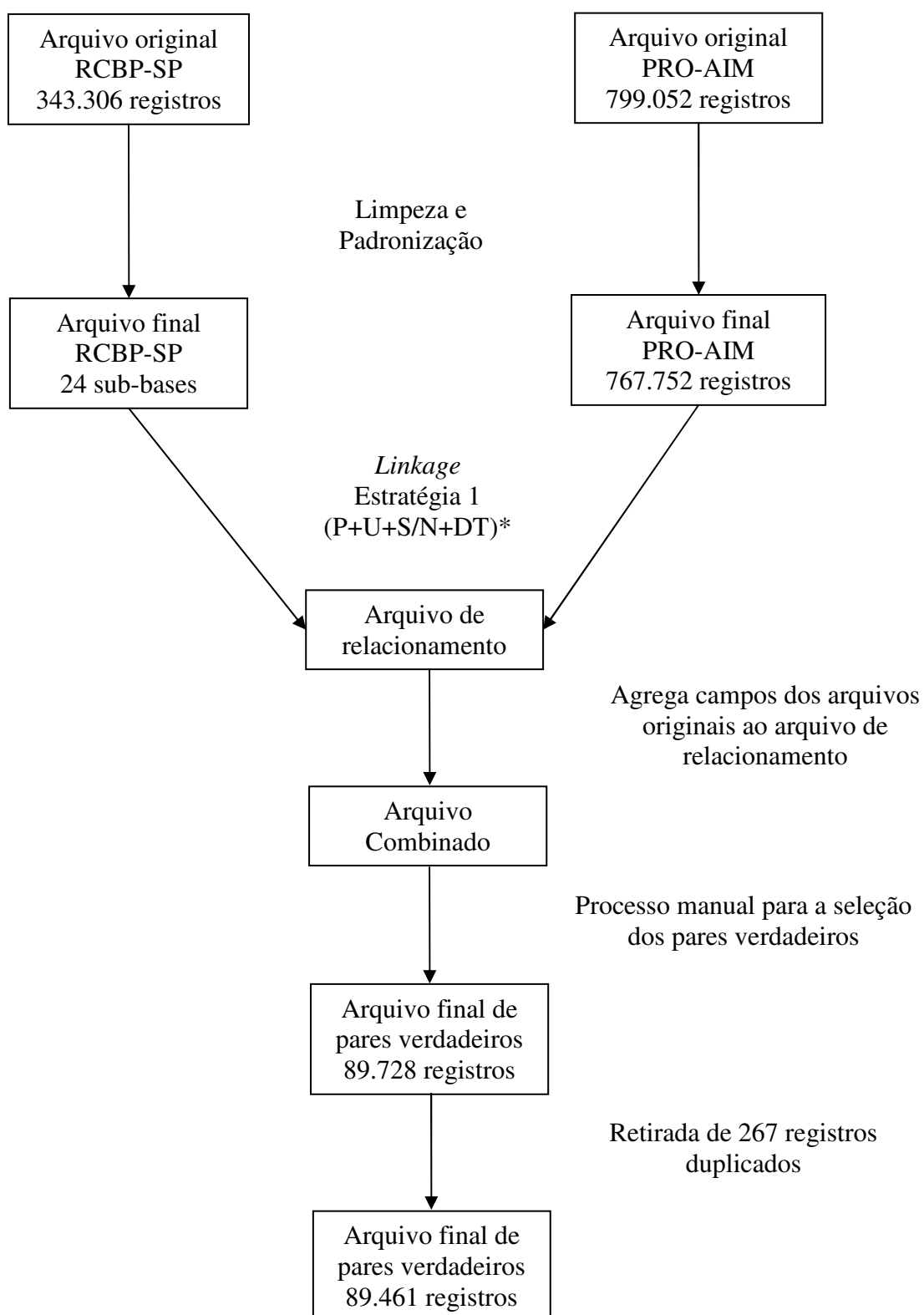
4.1 Resultados dos relacionamentos

Os resultados dos relacionamentos com a base de dados do RCPB-SP estão demonstrados nas Figuras 1 a 3. O total de pares verdadeiros encontrados nesta pesquisa foi resultado de duas estratégias de *linkage*: probabilístico e determinístico.

Após o processo de *linkage* probabilístico, entre as 24 sub-bases de dados do RCBP-SP *versus* PRO-AIM, foi criado um arquivo único de pares verdadeiros. Foram identificados 89.728 pares, sendo 267 casos duplicados. Ao final, o arquivo totalizou 89.461 pares verdadeiros (Figura 1). É importante reportar que foram encontrados 9.557 óbitos cuja causa básica foi câncer e não constavam da base de dados do RCBP-SP.

O total de pares verdadeiros para o *linkage* probabilístico foi resultado da estratégia de relacionamento composta pela chave de blocagem PBLOCO, UBLOCO e SEXO e pelas variáveis de pareamento nome e data de nascimento do paciente. No caso de topografias restritas a um único sexo, o campo de blocagem SEXO foi retirado. Destaca-se que esta estratégia de *linkage* foi a mais eficiente para todos os tipos de topografias de cânceres, assegurando o maior número de pares verdadeiros em um menor tempo de execução.

Figura 1 – RCBP-SP versus PRO-AIM. RCBP-SP, 1997 a 2005.



* P=PBLOCO; U=UBLOCO; S=SEXO; N=NOME DO PACIENTE; DT=DATA NASCIMENTO DO PACIENTE

Como um único paciente pode apresentar mais de um tumor - o que corresponde ao mesmo número de registro - as informações encontradas na base de dados do PRO-AIM foram completadas em 91.065 registros.

Das informações completadas, 16.985 (19,0%) casos não constavam como óbitos na base de dados do RCBP-SP. Dos 103.910 registros com data de óbito presentes nesta base de dados, 12.845 registros não foram relacionados, correspondendo 12,4%.

A partir do *linkage* entre as 24 sub-bases de dados do RCBP-SP e a base da APAC-SIA/SUS, também foi gerado um arquivo único com 13.571 de pares verdadeiros. Destes, 645 eram duplicidades, totalizando 12.926 casos (Figura 2). Nesta base de dados, cada registro relacionado foi identificado como único.

Para realizar o *linkage* entre estas bases, foi utilizada uma única estratégia de relacionamento, que foram distintas segundo a topografia. Em 12 das 24 sub-bases, a estratégia composta pela chave de blocagem PBLOCO + UBLOCO + SEXO com os campos de pareamento nome e data de nascimento foi a mais adequada e, em 10 sub-bases, a estratégia com a chave de blocagem PBLOCO + UBLOCO + SEXO com os campos de pareamento nome e ano de nascimento foi a mais adequada. Para a topografia C51.0 a C58.9 a melhor estratégia de blocagem foi composta por PBLOCO + UBLOCO + ANO DE NASCIMENTO com os campos de pareamento nome e data de nascimento. Para o caso das topografias restritas a um único sexo o campo de blocagem SEXO foi retirado. Desta forma, foi assegurado que o maior número de pares verdadeiros fosse agregado.

Dos 12.926 registros completados pelo *linkage* com a base de dados da APAC-SIA/SUS, 839 (6,5%) registros tinham a informação da data do óbito e

apenas 1.145 (8,9%) já constavam na base de dados do RCBP-SP com a data da última informação, tendo esta variável um ganho de 84,7%.

Destes, 56 (0,4%) foram considerados óbitos por câncer ou óbitos por não câncer (respectivamente, códigos 91 e 92) e, o restante, considerados vivos. Vale lembrar que na base de dados do RCBP-SP não há diferenciação entre óbito por câncer ou não. Destaca-se que, neste relacionamento, foram encontrados 2.094 registros de pacientes com a informação de residência no Município de São Paulo e que realizaram algum procedimento oncológico e que não estavam na base de dados do RCBP-SP.

Verificou-se que, tanto para o *linkage* probabilístico entre as 24 sub-bases de dados do RCBP-SP *versus* PRO-AIM, quanto para o *linkage* RCBP-SP *versus* APAC-SIA/SUS, as estratégias que apresentavam mais de três variáveis nas chaves de bloqueio mostraram-se com baixa precisão, alcançando um escore de 16,69, inferior aos escores máximos de 23,28 e 23,04, respectivamente, dos relacionamentos com o PRO-AIM e APAC-SIA/SUS.

A outra estratégia de *linkage* utilizada, nesta pesquisa, foi a técnica determinística para o relacionamento entre as bases de dados do RCBP-SP e FSEADE. Neste *linkage* foram vinculados 20.463 pares verdadeiros. Destes, 19.752 casos eram de pacientes com um único registro e foram identificados 711 (3,5%) pacientes com mais de uma topografia de câncer (Figura 3).

Dos registros relacionados, apenas 4.480 constavam com data de óbito na base de dados do RCBP-SP, completando esta informação em 80,9% (15.983) dos registros. Verificou-se que do total de registros completados, 15.631 (79,1%)

apresentaram São Paulo como Município de residência. Salienta-se que, destes registros, 4.447 (28,4%) foram identificados entre os anos de 2008 e 2009.

Figura 2 – RCBP-SP *versus* APAC-SIA/SUS. RCBP-SP, 1997 a 2005.

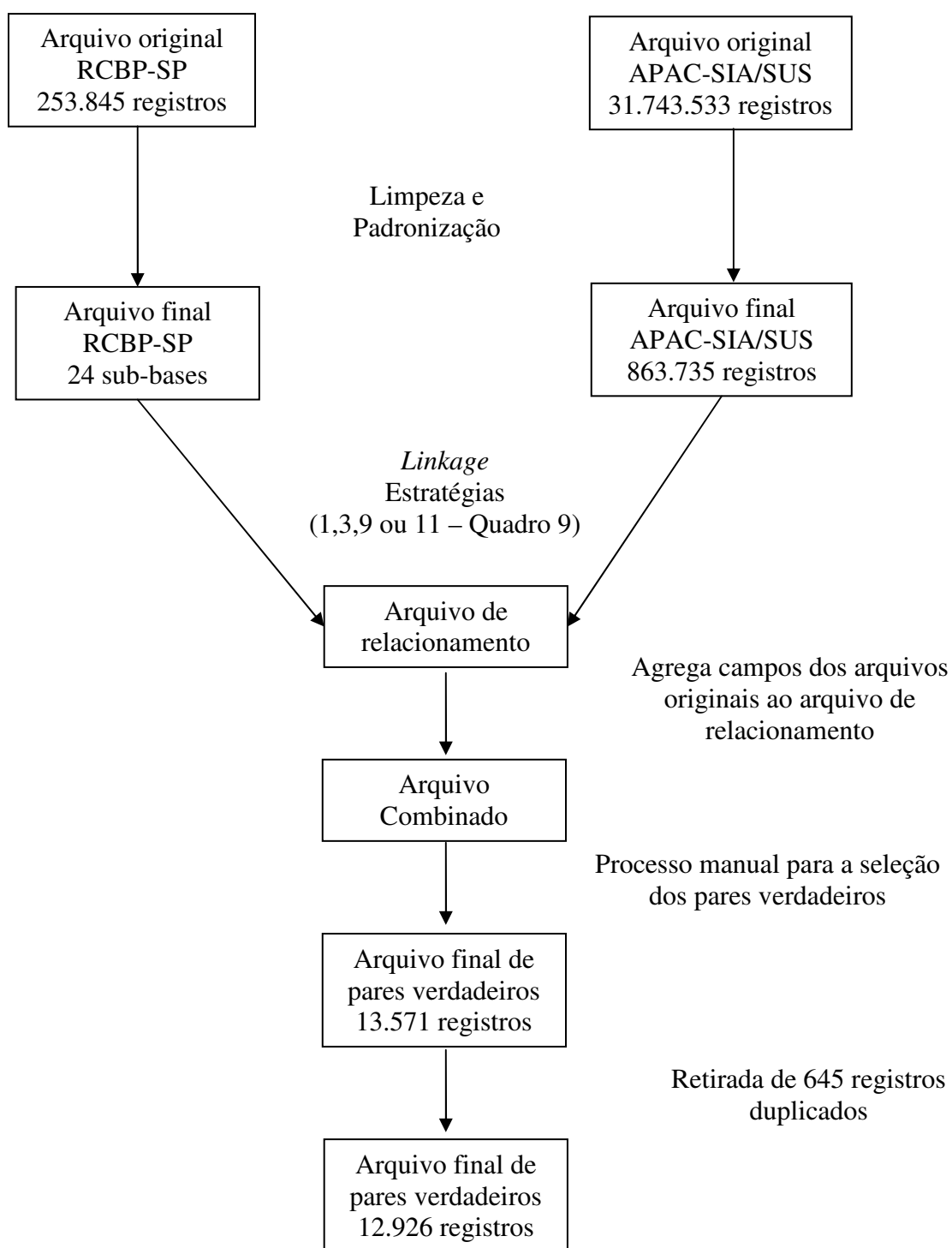
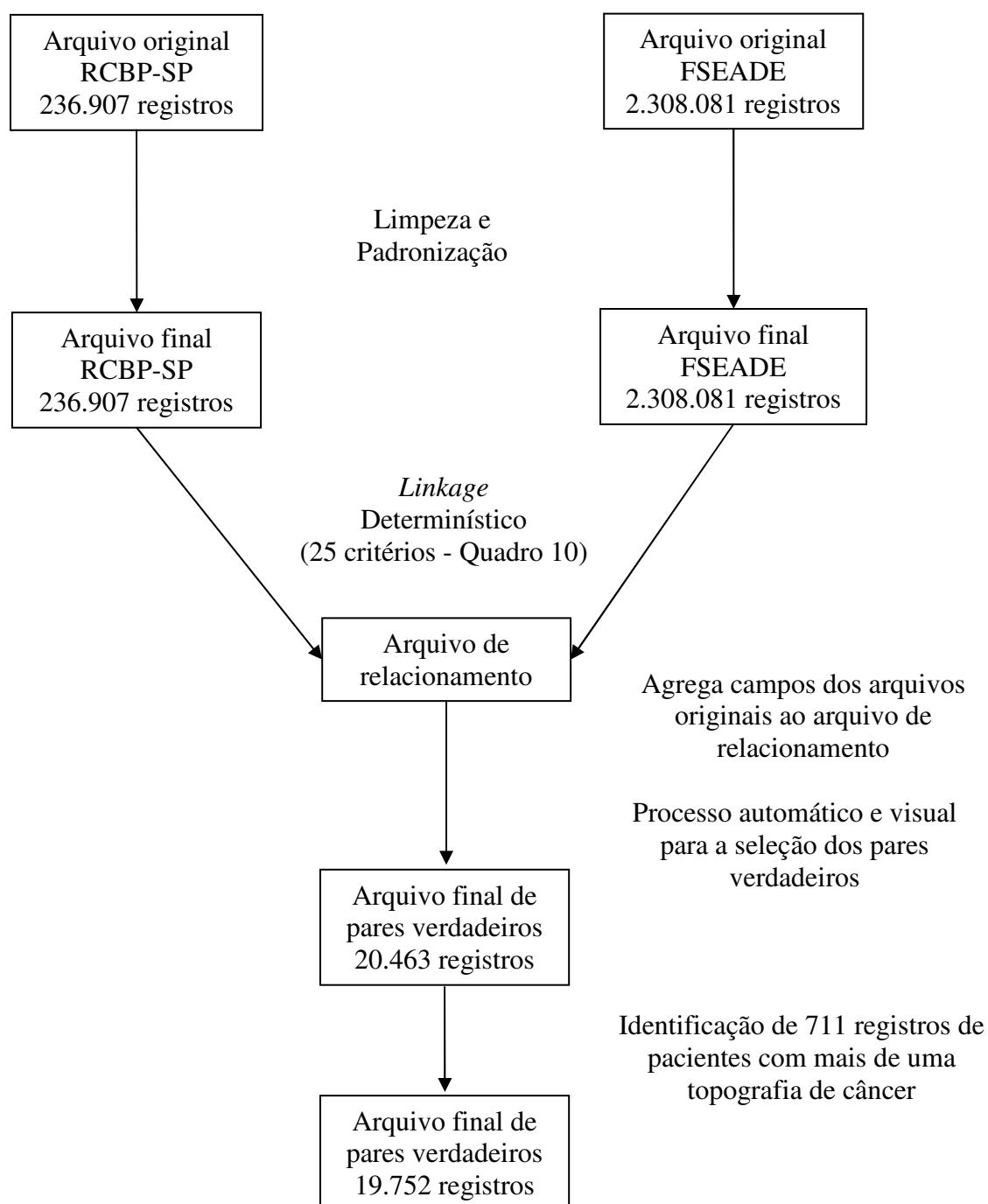


Figura 3 – RCBP-SP versus FSEADE. RCBP-SP, 1997 a 2005.



4.2 Completitude da informação

Após a identificação dos pares verdadeiros, foi realizado o procedimento de completitude das informações faltantes.

A primeira etapa da completitude foi checar se um mesmo paciente vinculado poderia ter entrado na base de dados do RCBP-SP pelos três *linkages* realizados. Nesta verificação identificou-se que os mesmos 1.720 pacientes foram encontrados pela APAC-SIA/SUS e pela FSEADE. Nestes casos, a informação que prevaleceu foi a obtida via FSEADE, dado que a data do óbito fecha o seguimento deste pacientes. Também foram encontrados 77 registros na APAC-SIA/SUS que foram relacionados com a base de dados de óbito do PRO-AIM. Da mesma forma, a informação considerada foi a obtida via PRO-AIM.

A Tabela 3 mostra a completitude da base de dados do RCBP-SP antes e após o *linkage*. Na coluna "Antes" estão apresentados os valores iniciais do RCBP-SP e na coluna denominada "Após" é apresentada a somatória entre quantidade de informação inicial com as informações completadas.

Ao se compararem os momentos "Antes" e "Após" da base de dados do RCBP-SP, no *linkage* do RCBP-SP com a base de dados do PRO-AIM, foi verificado um ganho de 69,2% para a variável endereço e 16,3% para a data do óbito. Quanto ao *linkage* com a base de dados da FSEADE, verificou-se ganhos de 24,9% e 15,4%, respectivamente, para as variáveis endereço e data de óbito (Tabela 3).

Quanto as variável nome da mãe e endereço, evidencia-se que antes do *linkage* estas informações apresentavam preenchimento, respectivamente, 23,6% e

0,5%. Destaca-se que, o nome da mãe não era uma informação de registro obrigatório (Tabela 3).

No geral, verifica-se que todos os relacionamentos foram importantes na estratégia para a completitude das informações da base de dados do RCBP-SP. Quanto ao seguimento, a data do óbito apresentou um ganho de 31,5%. Em relação ao *linkage* com a base de dados da APAC-SIA/SUS, a variável data da última informação teve um ganho na completitude do dado de 80,0%.

Tabela 3 – Número e percentual de pacientes com informação não ignorada, segundo bases de dados relacionadas e para a base final. RCBP-SP, 1997 a 2005.

Relacionamentos	Variáveis	Antes n	Após n	≠ percentual
PRO-AIM	Data de nascimento	220.176	224.719	2,1
	Endereço	81.057	137.118	69,2
	Nome da mãe	1.801	44.990	2.398,1
	Data do óbito	103.910	120.895	16,3
APAC-SIA/SUS	Data de nascimento	220.176	221.968	0,8
	Endereço	81.057	88.988	9,8
	Nome da mãe	1.801	14.477	703,8
	Data da última informação	11.462	20.628	80,0
FSEADE	Data de nascimento	220.176	223.288	1,4
	Endereço	81.057	101.209	24,9
	Nome da mãe	1.801	22.264	1.136,2
	Data do óbito	103.910	119.893	15,4
GERAL	Data de nascimento	220.176	229.470	4,2
	Endereço	81.057	163.310	101,5
	Nome da mãe	1.801	76.332	4.138,3
	Data do óbito	103.910	136.626	31,5
	Data da última informação	11.462	20.628	80,0

A partir do *linkage* entre a base de dados do RCBP-SP *versus* PRO-AIM e *versus* FSEADE, a variável causa do óbito foi agregada. Na Tabela 4 está apresentada a causa básica de óbito por base de dados. Observa-se que, dos casos relacionados, a causa básica de óbito, em sua maioria, foi decorrente ao câncer. É importante destacar que o *linkage* com a base de dados APAC-SIA/SUS não objetivou encontrar a causa básica do óbito. Mesmo assim, 56 casos foram identificados pelo motivo de saída do seguimento desta base.

Tabela 4 – Número e percentual de óbitos, segundo causa básica e base de dados de relacionamento. RCBP-SP, 1997 a 2005.

Relacionamentos	Causa básica do óbito				Total	
	Câncer		Outras causas		n	(%)
	n	(%)	n	(%)		
PRO-AIM (1997 a 2007)*	78.326	(87,6)	11.135	(12,4)	89.461	(100)
APAC-SIA/SUS (08/2003 a 2007)*	42	(75,0)	14	(25,0)	56	(100)
FSEADE (2000 a 2009)*	16.038	(81,2)	3.714	(18,8)	19.752	(100)
TOTAL	94.406	(86,4)	14.863	(13,6)	109.269	(100)

* período de abrangência

4.3 Coeficientes brutos de incidência (CBI) e de mortalidade (CBM)

Na Tabela 5 está apresentado o status final dos pacientes registrados na base de dados do RCBP-SP. Observa-se que houve um ganho de 3,4% de casos novos, passando 343.306 para 354.957 casos incidentes.

Destaca-se, também, que houve uma diminuição na perda de seguimento, após o *linkage*, tanto para o status vivo (antes = 3,3% e após = 5,8%) quanto para o status óbito (antes = 30,3% e após = 41,2%). Contudo, ainda há uma perda de seguimento de 53% de pacientes (Tabela 5).

Tabela 5 – Número e percentual de pacientes segundo status, antes e após o *linkage* entre o RCBP-SP com as bases de dados do PRO-AIM, APAC-SIA/SUS e FSEADE. RCBP-SP, 1997 a 2005.

Status final	antes		após	
	n	(%)	n	(%)
Vivo	11.462	(3,3)	20.628	(5,8)
Óbito	103.910	(30,3)	146.183	(41,2)
Ignorado	227.934	(66,4)	188.146	(53,0)
Total	343.306	(100)	354.957	(100)

A Tabela 6 mostra a diferença entre o total de casos, segundo status do paciente. Observa-se que um dos maiores ganhos na informação foi para o grupo de topografias C60.0 a C60.9/ C62.0 a C63.9, que corresponde ao câncer do trato

genito-urinário masculino. Os status vivo e óbito passaram, respectivamente, de 54 casos para 205 e de 358 para 1.285 casos.

Tabela 6 – Número e percentual de pacientes segundo status, antes e após o *linkage*. RCBP-SP, 1997 a 2005.

Topografia	Antes								Depois							
	vivo		Óbito		Ignorado		total		vivo		óbito		ignorado		total	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
C00.0 a C09.9	184	(2,0)	3.202	(34,4)	5.935	(63,7)	9.321	(100)	511	(5,4)	4.296	(45,0)	4.736	(49,6)	9.543	(100)
C10.0 a 11.9	34	(2,1)	892	(55,3)	687	(42,6)	1.613	(100)	101	(5,9)	1.156	(67,1)	466	(27,0)	1.723	(100)
C12.0 a 14.8	33	(2,4)	732	(53,0)	616	(44,6)	1.381	(100)	61	(4,2)	943	(65,4)	437	(30,3)	1.441	(100)
C15.0 a C15.9	58	(1,0)	3.497	(62,7)	2.024	(36,3)	5.579	(100)	103	(1,8)	4.319	(73,7)	1.438	(24,5)	5.860	(100)
C16.0 a C17.9	253	(1,4)	9.763	(52,7)	8.495	(45,9)	18.511	(100)	437	(2,3)	12.386	(64,5)	6.380	(33,2)	19.203	(100)
C18.0 a C18.9	273	(1,7)	6.548	(41,7)	8.863	(56,5)	15.684	(100)	524	(3,2)	8.965	(54,8)	6.860	(42,0)	16.349	(100)
C19.0 a C21.8	306	(2,9)	3.914	(36,9)	6.395	(60,2)	10.615	(100)	526	(4,8)	5.406	(49,3)	5.030	(45,9)	10.962	(100)
C22.0 a C24.9	27	(0,7)	2.624	(65,3)	1.370	(34,1)	4.021	(100)	72	(1,6)	3.593	(77,9)	947	(20,5)	4.612	(100)
C25.0 a C26.9	42	(0,7)	4.631	(78,5)	1.223	(20,7)	5.896	(100)	60	(0,9)	5.469	(85,5)	869	(13,6)	6.398	(100)
C30.0 a C33.9	128	(2,0)	2.606	(41,6)	3.526	(56,3)	6.260	(100)	372	(5,7)	3.432	(52,9)	2.680	(41,3)	6.484	(100)
C34.0 a C34.9	168	(1,0)	12.102	(69,3)	5.185	(29,7)	17.455	(100)	285	(1,5)	14.692	(79,8)	3.432	(18,6)	18.409	(100)
C37.0 a C39.9	6	(0,7)	434	(50,8)	414	(48,5)	854	(100)	27	(2,7)	646	(65,3)	317	(32,0)	990	(100)
C40.0 a C42.4	291	(2,6)	5.913	(52,1)	5.154	(45,4)	11.358	(100)	522	(4,6)	7.080	(62,2)	3.778	(33,2)	11.380	(100)
C44.0 a C44.9 (8720 a 8780)	104	(2,1)	684	(13,6)	4.246	(84,3)	5.034	(100)	239	(4,3)	1.644	(29,3)	3.720	(66,4)	5.603	(100)
C44.0 a C44.9	3.050	(4,3)	3.223	(4,6)	64.389	(91,1)	70.662	(100)	3.578	(5,1)	8.937	(12,6)	58.162	(82,3)	70.677	(100)
C47.0 a C49.9	55	(1,9)	915	(32,3)	1.860	(65,7)	2.830	(100)	201	(6,7)	1.278	(42,8)	1.508	(50,5)	2.987	(100)
C50.0 a C50.9	3.004	(7,3)	9.552	(23,1)	28.751	(69,6)	41.307	(100)	5.643	(13,1)	14.822	(34,3)	22.688	(52,6)	43.153	(100)
C51.0 a C58.9	1.129	(4,0)	7.500	(26,4)	19.730	(69,6)	28.359	(100)	2.632	(8,9)	10.638	(36,0)	16.285	(55,1)	29.555	(100)
C60.0 a C60.9/ C62.0 a C63.9	54	(2,5)	358	(16,6)	1.741	(80,9)	2.153	(100)	205	(7,0)	1.285	(43,9)	1.440	(49,1)	2.930	(100)
C61.9	885	(3,0)	6.428	(21,8)	22.142	(75,2)	29.455	(100)	2.051	(6,8)	10.632	(35,2)	17.553	(58,1)	30.236	(100)
C64.0 a C68.9	450	(3,6)	3.943	(31,9)	7.950	(64,4)	12.343	(100)	607	(4,9)	5.233	(42,3)	6.527	(52,8)	12.367	(100)
C69.0 a C72.9	97	(1,3)	3.484	(48,4)	3.618	(50,3)	7.199	(100)	314	(4,3)	4.161	(57,4)	2.776	(38,3)	7.251	(100)
C73.0 a C75.9	370	(3,2)	798	(6,8)	10.528	(90,0)	11.696	(100)	581	(4,9)	1.314	(11,2)	9.876	(83,9)	11.771	(100)
C76.0 a C80.9	461	(1,9)	10.167	(42,9)	13.092	(55,2)	23.720	(100)	976	(3,9)	13.856	(55,3)	10.241	(40,8)	25.073	(100)
Total sem pele não melanoma	8.412	(3,1)	100.687	(36,9)	163.545	(60,0)	272.644	(100)	17.050	(6,0)	137.246	(48,3)	129.984	(45,7)	284.280	(100)
Total	11.462	(3,3)	103.910	(30,3)	227.934	(66,4)	343.306	(100)	20.628	(5,8)	146.183	(41,2)	188.146	(53,0)	354.957	(100)

A seguir são apresentados os CBI e CBM, antes e após o *linkage*. Para o denominador foi considerada a população do Município de São Paulo entre os anos de 1997 a 2005. Para o cálculo destes coeficientes foram excluídos os casos de pele não melanoma (C44.0 a C44.9). Nesta análise, os numeradores são o total de casos novos (n antes = 272.644 e n após = 284.280) e o total de óbitos (n antes = 95.899 e n após = 120.652), ocorridos entre os anos de 1997 a 2005. Ressalta-se que no cálculo do CBM são considerados um único registro de cada paciente (Tabela 7).

O CBI, que era de 288,2 casos novos por 100.000 habitantes passou para 300,5 casos novos por 100.000 após o *linkage* com as bases de dados, havendo um ganho de +4,3% no coeficiente. Analisando os CBM, verifica-se que, no geral, há uma diferença percentual de 25,8 % (Tabela 7).

Tabela 7 – Coeficientes brutos de incidência e mortalidade por câncer (por 100.000 habitantes) para o período de 1997 a 2005, antes e após o *linkage*. RCBP-SP, 1997 a 2005.

Coeficientes	Antes		Após		≠ percentual
	n	coeficiente	n	coeficiente	
CBI	272.644	288,2	284.280	300,5	4,3
CBM*	95.899	101,4	120.652	127,5	25,8

*óbitos encontrados nas bases de dados do PRO-AIM, FSEADE e mais os 56 casos da APAC-SIA/SUS

4.4 Probabilidade de sobrevida global acumulada antes e após o *linkage*

Nesta análise de sobrevida são considerados apenas os registros com status conhecidos, a saber antes do *linkage* ($n_{\text{vivo}} = 11.462$ e $n_{\text{óbito}} = 103.9100$) e após o *linkage* ($n_{\text{vivo}} = 20.628$ e $n_{\text{óbito}} = 146.183$).

Também foram excluídos da análise, casos com datas erradas (negativas) e com tempo de seguimento zero antes e após o *linkage*, respectivamente, 43.261 e 54.735. Por fim, para a análise de sobrevida foram considerados 115.372 registros antes e 166.811 após o *linkage*.

Na Tabela 8 estão apresentadas as probabilidades de sobrevida global acumulada geral e segundo sexo, antes e após o *linkage*, truncada em quatro anos. Observa-se que antes do *linkage* há uma subestimação estatisticamente significativa da probabilidade de sobrevida. A análise com o total dos casos mostra que a probabilidade de estar vivo no primeiro ano de seguimento antes do *linkage* foi menor quando comparada a probabilidade de estar vivo ao primeiro ano de seguimento após o *linkage* (46,9% x 59,8%; $p < 0,001$), ressaltando-se que a subestimação da probabilidade de sobrevida acumulada aumenta conforme o tempo de seguimento. O Gráfico 1 mostra as curvas de sobrevida global acumulada geral, antes e após o *linkage*.

Na estratificação por sexo, também se observa a mesma tendência. Para o sexo feminino, no quarto ano de seguimento antes do *linkage* a probabilidade de sobrevida acumulada foi de 18,5%, enquanto que após o *linkage*, para o mesmo período, esta probabilidade foi de 29,7% ($p < 0,001$). Da mesma forma, para o sexo

masculino ao quarto ano de seguimento as probabilidades foram, respectivamente, de 10,9% antes e de 22,1% ($p < 0,001$). O Gráfico 2 apresenta estes resultados.

Tabela 8 – Probabilidade de sobrevida global acumulada (%) geral e segundo sexo, antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.

Variável	Linkage	n° de casos	N° de óbitos	Probabilidade de sobrevida global acumulada (% em anos)				p (K-M)
				1°	2°	3°	4°	
GERAL	Antes	72.111	57.027	46,9	29,5	20,4	14,6	<0,001
	Após	112.076	76.167	59,8	43,8	33,6	25,9	
SEXO								
Feminino	Antes	35.512	26.450	52,6	34,8	24,8	18,5	<0,001
	Após	56.516	35.610	64,6	48,7	37,9	29,7	
Masculino	Antes	36.599	30.577	41,6	24,5	16,2	10,9	<0,001
	Após	55.560	40.557	55,1	39,0	29,4	22,1	

Gráfico 1 – Probabilidade de sobrevida global acumulada (%), antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.

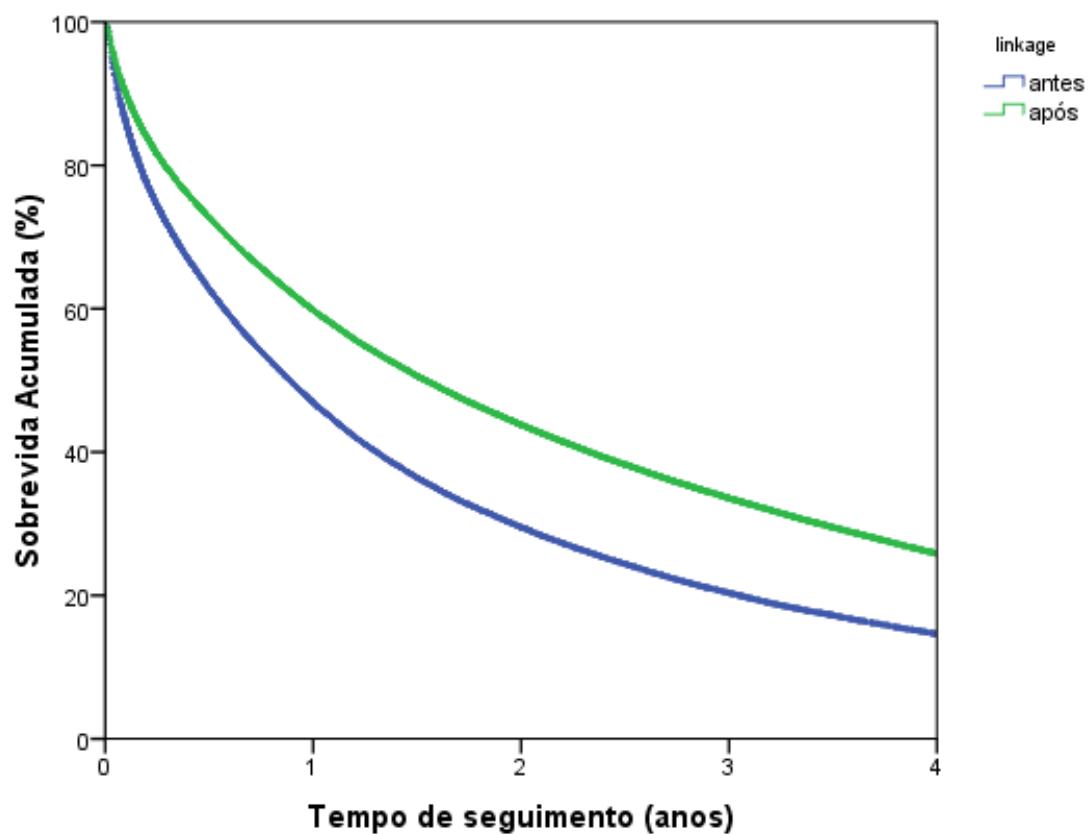
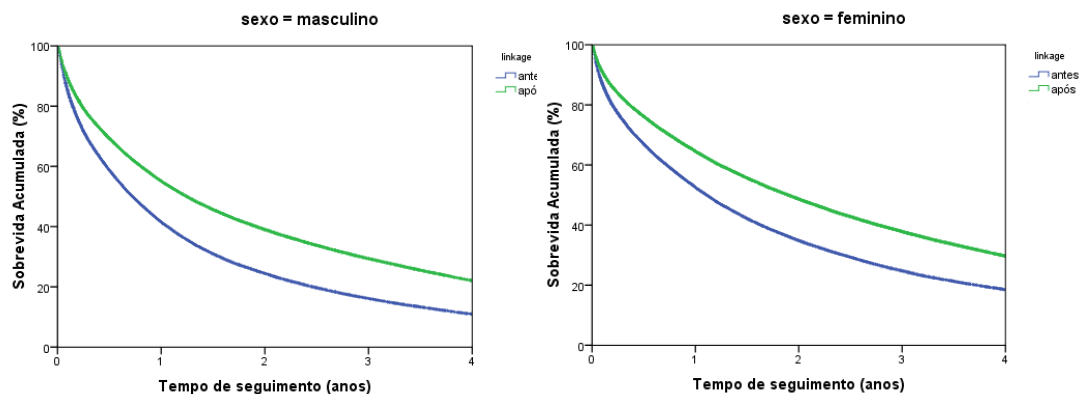


Gráfico 2 - Probabilidades de sobrevida global acumulada (%) segundo sexo, antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.

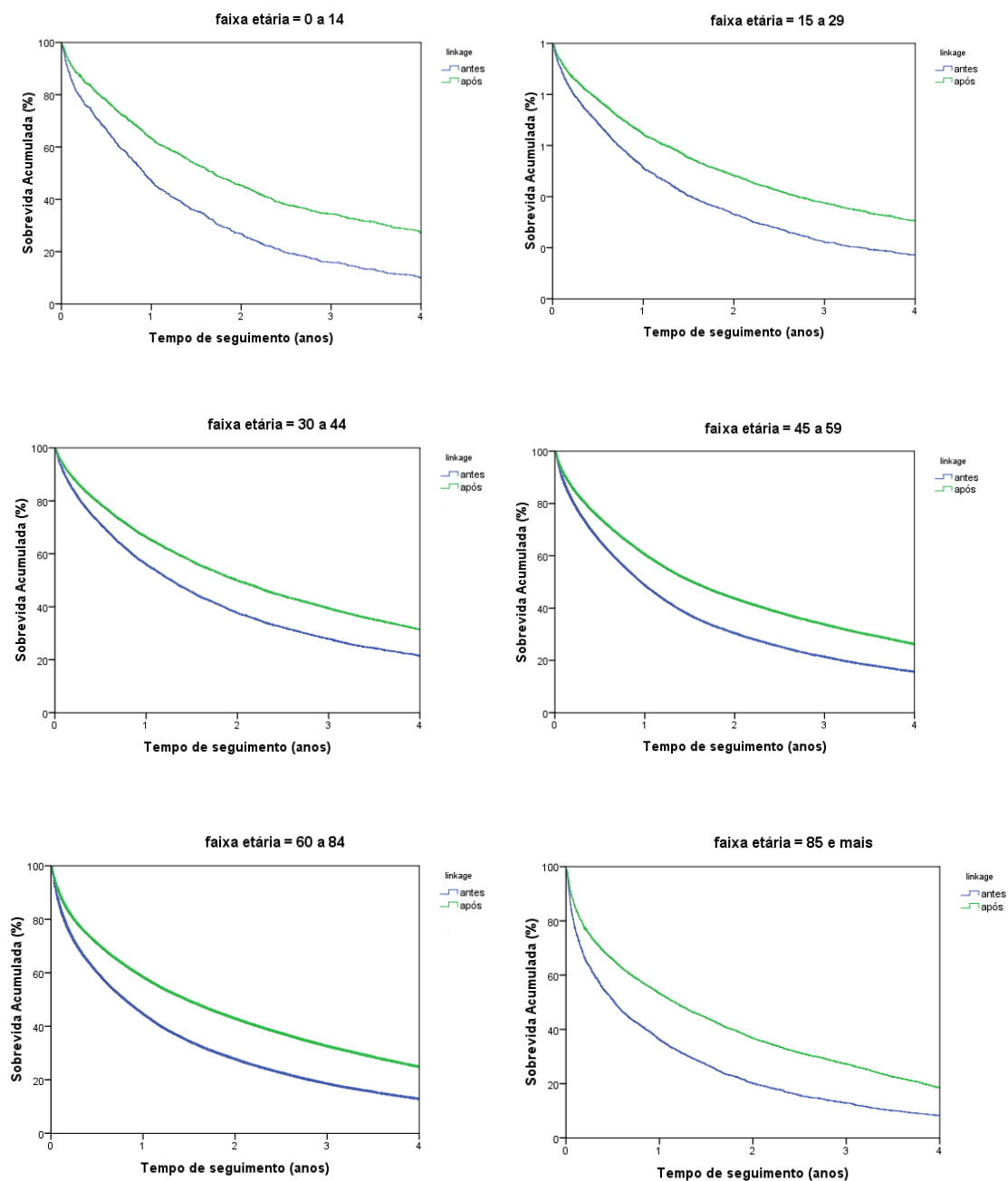


A Tabela 9 mostra as probabilidades acumuladas de sobrevida global após quatro anos estratificadas por faixa etária. Verifica-se que, para todas as faixas etárias, houve um ganho significativo nestas probabilidades. O Gráfico 3 mostra a sobrevida global segundo faixa etária, antes e após o *linkage*.

Tabela 9 – Probabilidade de sobrevida global acumulada (%) segundo faixa etária, antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.

Faixa Etária (anos)	Linkage	n° de Casos	n° de Óbitos	Probabilidade de sobrevida global acumulada (% em anos)				P (K-M)
				1°	2°	3°	4°	
0 – 14	Antes	938	766	47,2	26,8	16,0	10,2	<0,001
	Após	1.620	969	63,6	45,3	34,5	27,6	
15 – 29	Antes	1.777	1.310	51,2	33,0	22,3	17,2	<0,001
	Após	2.855	1.692	64,5	48,2	37,6	30,6	
30 – 44	Antes	6.915	4.902	56,0	37,6	27,9	21,5	<0,001
	Após	10.875	6.540	66,3	49,9	39,5	31,5	
45 – 59	Antes	18.667	14.547	48,7	30,4	21,4	15,7	<0,001
	Após	28.089	18.749	60,6	47,3	33,8	26,2	
60 – 84	Antes	39.861	32.506	44,7	27,7	18,5	12,8	<0,001
	Após	62.244	43.596	58,4	42,9	32,5	24,9	
85 e +	Antes	3.352	2.931	36,2	20,1	12,9	8,2	<0,001
	Após	5.835	4.580	53,2	36,8	27,3	18,6	

Gráfico 3- Probabilidade de sobrevida global acumulada (%), segundo faixa etária, antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.



A Tabela 10 mostra os resultados da análise de sobrevida segundo topografia. Observa-se que houve diferença estatisticamente significativa nas probabilidades de sobrevida acumulada global, antes e após o *linkage*, para todas as topografias de câncer. Para todas as localizações, antes do *linkage* as probabilidades de sobrevida global acumulada estavam subestimadas.

Tabela 10 - Probabilidades de sobrevida global acumulada (%), segundo topografia, antes e após o *linkage*, truncada em quatro anos. RCBP-SP, 1997 a 2005.

Topografia	Linkage	n° de	n° de	Probabilidade de sobrevida global acumulada (% em anos)				p (K-M)
		Casos	Óbitos	1°	2°	3°	4°	
C00.0 a	Antes	2.609	2.342	40,4	17,3	9,8	5,5	<0,001
C09.9	Após	3.769	2.855	53,8	33,1	24,5	17,7	
C10.0 a	Antes	629	578	35,1	15,0	8,0	4,7	<0,001
C11.9	Após	877	700	45,1	26,7	18,3	13,9	
C12.0 a	Antes	527	476	36,3	16,4	9,1	4,8	< 0,001
C14.8	Após	708	581	46,7	28,9	19,0	13,0	
C15.0 a	Antes	2.289	2.207	19,9	6,2	2,9	1,6	<0,001
C15.9	Após	2.875	2.625	28,9	14,2	9,2	6,4	
C16.0 a	Antes	6.400	6.044	25,2	10,6	5,3	2,7	<0,001
C17.9	Após	8.441	7.372	36,4	21,4	14,7	10,3	
C18.0 a	Antes	4.100	3.705	38,7	20,1	9,6	4,9	<0,001
C18.9	Após	6.070	4.959	51,3	33,3	21,1	13,8	
C19.0 a	Antes	3.198	2.678	52,1	30,6	16,9	10,0	<0,001
C21.8	Após	4.574	3.414	61,5	42,0	28,3	19,4	
C22.0 a	Antes	1.306	1.261	21,0	8,1	3,5	1,7	<0,001
C24.9	Após	1.735	1.554	33,0	18,9	12,1	8,6	
C25.0 a	Antes	1.690	1.623	16,9	6,0	3,1	2,0	<0,001
C26.9	Após	2.129	1.956	25,5	13,9	9,4	6,6	
C30.0 a	Antes	1.790	1.583	43,5	21,5	12,1	7,0	<0,001
C33.9	Após	2.617	1.945	55,9	36,5	26,0	19,1	
C34.0 a	Antes	7.283	6.925	27,5	11,8	5,9	3,3	<0,001
C34.9	Após	9.077	8.171	35,6	19,2	12,1	8,3	
C37.0 a	Antes	217	200	31,1	16,8	9,0	6,5	0,003
C39.9	Após	314	260	40,1	23,9	18,0	12,9	

(continua)

Topografia	Linkage	n° de Casos	n° de Óbitos	Probabilidade de sobrevida global acumulada (% em anos)				P (K-M)
				1°	2°	3°	4°	
C40.0 a	Antes	3.356	2.902	41,2	22,9	13,4	8,2	<0,001
C42.4	Após	4.731	3.699	52,3	35,0	25,0	17,5	
C44.0 aC44.9 (8720/8780)	Antes	603	450	61,9	41,1	26,6	16,7	<0,001
	Após	1.075	705	67,8	49,3	36,1	25,1	
C44.0 a	Antes	5.875	2.365	81,4	69,0	59,1	50,1	0,036
C44.9	Após	11.640	5.203	85,2	72,6	61,1	49,4	
C47.0 a	Antes	679	592	47,5	24,4	12,0	7,0	<0,001
C49.9	Após	1.048	724	61,6	43,1	31,2	24,0	
C50.0 a	Antes	9.140	5.311	73,6	56,3	43,3	33,9	<0,001
C50.9	Após	15.740	7.249	82,0	68,0	55,8	45,6	
C51.0 a	Antes	5.706	4.359	54,2	32,2	22,0	15,6	<0,001
C58.9	Após	9.372	5.745	66,5	47,3	35,5	27,2	
C60/C60.9	Antes	278	216	47,3	20,1	13,3	8,5	<0,001
C62./C63.9	Após	577	312	67,7	50,1	42,2	36,8	
C61.9	Antes	4.470	3.142	62,0	43,6	30,7	20,6	<0,001
	Após	8.985	4.893	76,4	61,4	48,9	38,4	
C64.0 a	Antes	2.885	2.205	52,9	34,7	24,5	18,3	<0,001
C68.9	Após	4.261	2.915	62,5	46,2	35,5	27,9	
C69.0 a	Antes	1.758	1.582	36,7	18,7	11,5	6,9	<0,001
C72.9	Após	2.637	2.050	49,9	32,3	24,3	17,8	
C73.0 a	Antes	835	417	69,9	54,6	45,5	37,1	<0,001
C75.9	Após	1.432	633	78,2	66,0	56,5	46,9	
C76.0 a	Antes	4.488	3.864	34,2	19,7	12,7	8,4	<0,001
C80.9	Após	7.392	5.647	47,6	32,9	24,2	18,1	

(continuação)

* a Tabela com a descrição das topografias encontra-se no Anexo 4.

Para a análise de sobrevida truncada em sete anos, também verificou-se que havia uma subestimação nas probabilidades de sobrevida global acumulada. Destaca-se que, no geral, antes do *linkage* a probabilidade de estar vivo em sete anos era de 7,8%, passando para 13,0% após o *linkage* (Tabela 11). Da mesma forma, a subestimação foi observada para ambos os sexos. Verifica-se que, entre os homens, após dois anos de seguimento, houve uma diferença percentual de 59,3%. A probabilidade de estar vivo antes do *linkage* era de 26,3% passando para 41,9% após o relacionamento ($p < 0,001$).

Tabela 11 – Probabilidade de sobrevida global acumulada (%) geral e segundo sexo, antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.

Variável	<i>Linkage</i>	n° de casos	n° de óbitos	Probabilidade de sobrevida global acumulada (% em anos)					p (K-M)
				1°	2°	3°	5°	7°	
GERAL	Antes	56.237	48.613	48,8	31,9	22,7	12,3	7,8	<0,001
	Após	78.831	64.120	61,1	46,6	37,0	23,2	13,0	
SEXO									
Feminino	Antes	27.865	22.900	54,7	37,6	27,6	16,4	11,8	<0,001
	Após	39.415	30.379	65,7	51,2	41,2	27,0	16,4	
Masculino	Antes	28.372	25.713	43,0	26,3	18,0	8,3	3,5	<0,001
	Após	39.416	33.741	56,5	41,9	32,8	19,5	9,7	

Gráfico 4 – Probabilidade de sobrevida global acumulada (%), antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.

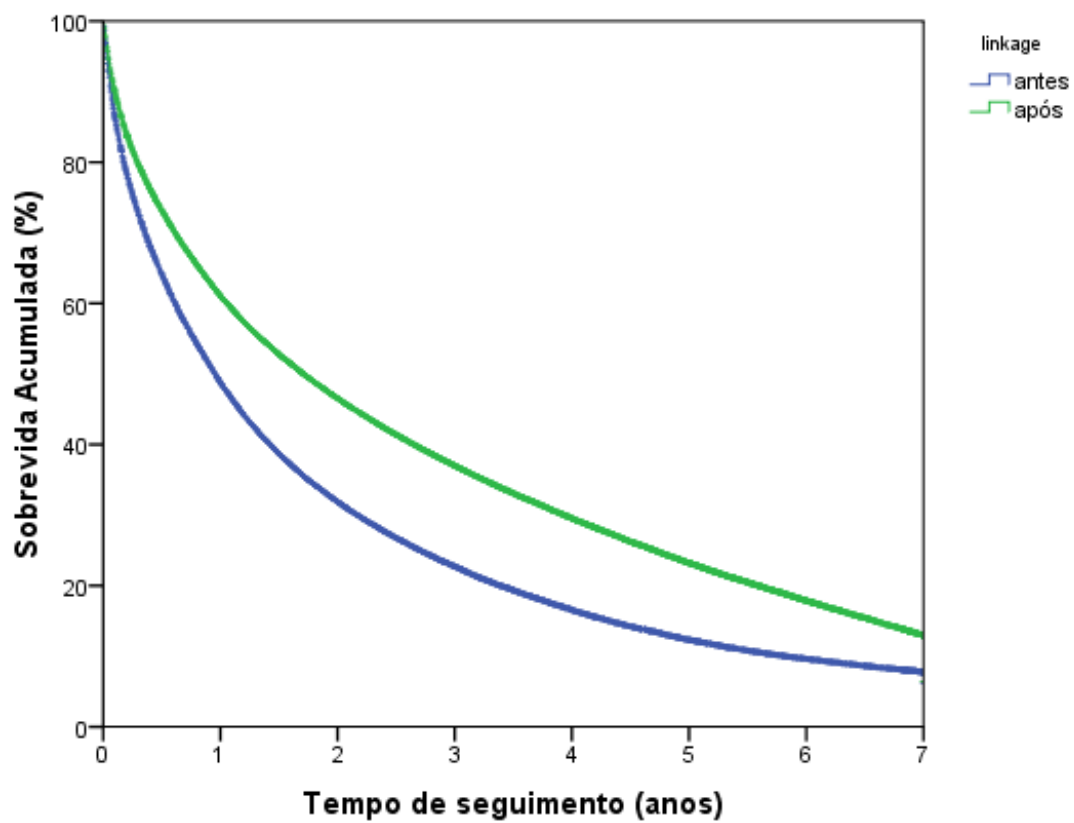
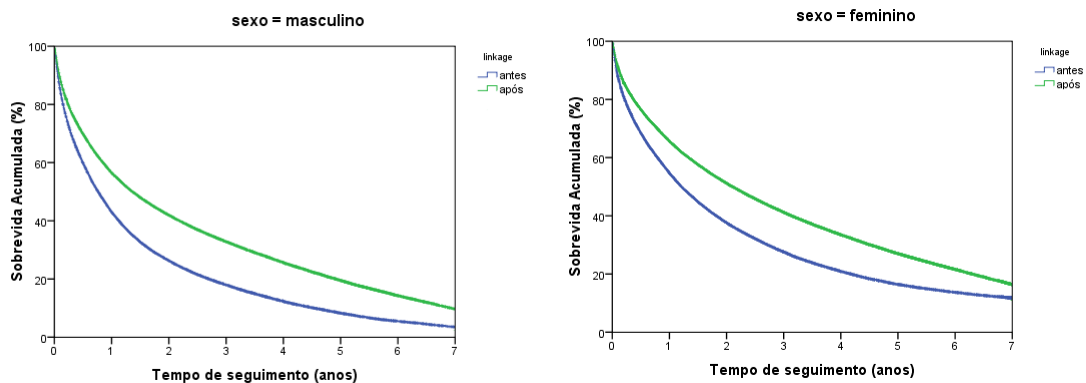


Gráfico 5 - Probabilidade de sobrevida global acumulada (%) segundo sexo, antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.



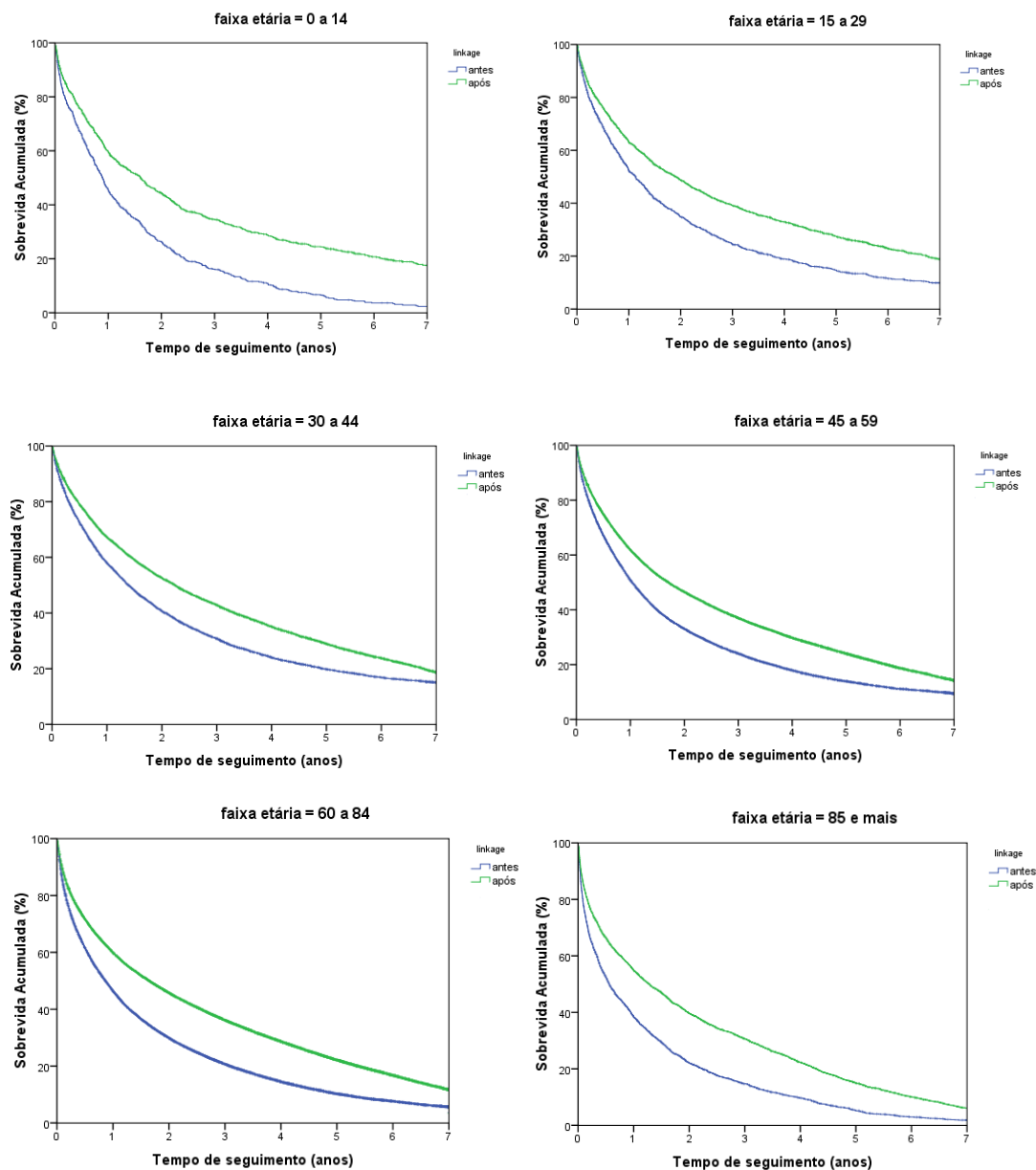
Na Tabela 12, observou-se que houve uma diferença estatisticamente significativa para todas as faixas etárias ($p < 0,001$) antes e após o *linkage*. Evidencia-se que a probabilidade de estar vivo ao sétimo ano passou de 2,3%, antes do *linkage* na faixa etária de 0 a 14 anos, para 17,6% após o *linkage*.

Para a faixa etária de 60 a 84 anos, verificou-se que ao segundo ano as probabilidades de sobrevida antes do *linkage* era de 29,9%, enquanto que após o relacionamento esta probabilidade ficou em 45,8%. O Gráfico 6 mostra estes resultados.

Tabela 12 – Probabilidade de sobrevida global acumulada (%) segundo faixa etária, antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.

Faixa Etária (anos)	Linkage	n° de Casos	n° de óbitos	Probabilidade de sobrevida global acumulada (% em anos)					p (K-M)
				1°	2°	3°	5°	7°	
0 – 14	Antes	711	648	45,6	26,3	16,2	6,5	2,3	<0,001
	Após	1.035	756	59,7	44,3	34,6	24,5	17,6	
15 – 29	Antes	1.408	1.147	52,3	34,8	24,6	14,7	9,9	<0,001
	Após	1.940	1.403	63,1	48,7	39,2	27,5	18,8	
30 – 44	Antes	5.532	4.300	58,0	40,6	30,8	19,8	15,0	<0,001
	Após	7.675	5.637	67,2	52,6	42,9	28,9	18,7	
45 – 59	Antes	14.503	12.374	50,9	33,1	24,1	13,9	9,5	<0,001
	Após	19.548	15.591	61,9	46,5	37,0	24,0	14,2	
60 – 84	Antes	31.157	27.716	46,4	29,9	20,7	10,3	5,7	<0,001
	Após	44.239	37.007	59,9	45,8	36,2	22,2	11,7	
85 e +	Antes	2.526	2.386	38,5	22,1	14,7	5,2	1,9	<0,001
	Após	4.021	3.699	54,9	39,7	30,7	15,0	6,1	

Gráfico 6 - Probabilidade de sobrevida global acumulada (%), segundo faixa etária, antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.



Em relação a análise sobrevida global acumulada estratificada por topografias de câncer (Tabela 13), verificou-se uma subestimação estatisticamente significativa nas probabilidades de sobrevida para todos os períodos analisados ($p < 0,001$).

Tabela 13 - Probabilidade de sobrevida global acumulada (%), segundo topografia, antes e após o *linkage*, truncada em sete anos. RCBP-SP, 1997 a 2005.

Topografia *	Linkage	n° de Casos	n° de Óbitos	Probabilidade de sobrevida global acumulada (% em anos)					p (K-M)
				1°	2°	3°	5°	7°	
C00.0 a	Antes	1.994	1.904	41,7	18,6	11,0	3,3	1,3	<0,001
C09.9	Após	2.635	2.328	54,5	34,7	26,4	14,7	7,5	
C10.0 a	Antes	456	440	36,2	17,3	10,1	4,5	1,2	<0,001
C11.9	Após	573	524	46,7	29,1	21,6	12,7	4,9	
C12.0 a	Antes	391	381	37,3	18,3	9,7	2,5	0,5	<0,001
C14.8	Após	500	468	48,9	33,6	23,3	10,9	4,1	
C15.0 a	Antes	1.706	1.686	20,1	6,8	3,6	1,1	0,3	<0,001
C15.9	Após	1.979	1.927	30,0	16,6	11,4	5,5	1,7	
C16.0 a	Antes	4.990	4.877	25,7	11,2	6,0	1,9	0,5	<0,001
C17.9	Após	6.029	5.693	36,8	23,0	16,6	9,0	3,9	
C18.0 a	Antes	3.116	2.994	40,4	21,9	11,1	2,8	0,6	<0,001
C18.9	Após	4.079	3.747	51,7	35,0	23,5	11,5	5,5	
C19.0 a	Antes	2.527	2.379	54,9	33,2	18,9	5,6	1,2	<0,001
C21.8	Após	3.239	2.883	63,0	44,7	30,9	15,6	7,1	
C22.0 a	Antes	988	976	21,5	9,0	4,1	0,5	0,2	<0,001
C24.9	Após	1.213	1.161	33,8	21,2	14,2	7,9	2,7	
C25.0 a	Antes	1.304	1.291	18,3	6,7	3,7	1,3	0,4	<0,001
C26.9	Após	1.524	1.485	26,7	15,4	10,8	5,1	1,7	
C30.0 a	Antes	1.381	1.326	44,0	23,2	13,6	4,7	1,0	<0,001
C33.9	Após	1.823	1.625	56,3	38,3	28,3	15,0	7,2	
C34.0 a	Antes	5.622	5.554	29,0	13,1	7,0	2,3	0,5	<0,001
C34.9	Após	6.543	6.317	37,2	21,6	14,3	7,0	2,6	
C37.0 a	Antes	154	150	32,5	20,8	12,2	6,0	1,5	<0,001
C39.9	Após	196	182	40,6	29,2	21,9	13,3	5,4	

(continua)

Topografia *	Linkage	n° de Casos	n° de Óbitos	Probabilidade de sobrevida global acumulada (% em anos)					p (K-M)
				1°	2°	3°	5°	7°	
				C40.0 a	Antes	2.604	2.448	42,7	
C42.4	Após	3.374	2.988	53,1	36,6	27,5	15,1	7,6	
C44.0 a	Antes	404	339	65,8	46,6	33,6	13,9	6,0	<0,001
C44.9 (8720/8780)	Após	662	535	69,0	53,1	41,4	21,5	11,5	
C44.0 a	Antes	4.642	2.300	82,9	71,7	62,4	44,7	31,6	<0,001
C44.9	Após	8.780	5.396	87,1	76,5	66,4	45,2	25,4	
C47.0 a	Antes	537	507	47,2	25,5	13,4	4,5	0,6	<0,001
C49.9	Após	746	606	61,7	44,9	33,6	20,7	11,7	
C50.0 a	Antes	7.414	5.188	74,8	59,0	46,5	30,4	23,5	<0,001
C50.9	Após	10.963	7.106	82,0	69,6	58,6	41,2	27,8	
C51.0 a	Antes	4.477	3.758	55,8	34,4	24,3	14,5	12,1	<0,001
C58.9	Após	6.242	4.858	66,7	48,6	37,8	24,8	16,3	
C60/C60.9	Antes	193	169	44,6	19,9	13,0	7,6	2,6	<0,001
C62./C63.9	Após	366	250	67,8	51,9	44,5	29,1	21,4	
C61.9	Antes	3.485	3.005	63,1	45,1	32,8	14,7	5,2	<0,001
	Após	6.326	4.850	77,5	64,1	52,7	32,8	17,3	
C64.0 a	Antes	2.384	2.042	55,7	37,5	26,7	13,3	4,9	<0,001
C68.9	Após	3.242	2.671	64,9	49,7	39,2	23,9	11,2	
C69.0 a	Antes	1.376	1.322	39,3	21,0	13,3	4,6	1,5	<0,001
C72.9	Após	1.874	1.663	52,1	36,2	28,0	16,2	7,6	
C73.0 a	Antes	633	393	70,3	55,4	46,8	32,8	22,9	<0,001
C75.9	Após	1.028	606	79,7	68,7	60,3	44,2	29,2	
C76.0 a	Antes	3.459	3.184	35,3	21,1	14,0	6,4	3,0	<0,001
C80.9	Após	4.895	4.251	48,6	35,5	27,5	16,7	8,0	

(continuação)

* a Tabela com a descrição das topografias encontra-se no Anexo 4.

4.5 Proposta de pontos de corte nos escores calculados nos processos de *linkage* probabilístico

Pelo exposto, verificou-se a relevância de se fazer o processo de *linkage* para a melhoria da qualidade da base de dados do RCBP-SP. No entanto, para se definirem os pares verdadeiros foi necessário analisar, manualmente, todos os registros relacionados a partir do menor escore de concordância.

Isso foi realizado em 648 bases de dados, sendo 24 sub-bases para o *linkage* probabilístico entre RCBP-SP *versus* PRO-AIM e as 24 sub-bases para o *linkage* entre o RCBP-SP *versus* APAC-SIA/SUS, onde cada base apresentou 14 estratégias de blocagem, exceto para as topografias exclusivas de um único sexo, tendo, estas, 10 estratégias de blocagem. Os tamanhos das bases de dados resultantes do processo de *linkage* probabilístico variaram de 187 a 1.309.283 registros, relacionados de acordo com o tamanho inicial de cada sub-base de dados.

Este trabalho levou dezoito meses para ser realizado por três pesquisadores. Por isso, decidiu-se verificar a possibilidade de se definirem pontos de corte nas probabilidades, com o objetivo de reduzir este trabalho. Nas Tabelas 14 e 15 estão os pontos de corte sugeridos.

Observa-se que o ponto de corte no escore 18 apresenta boa acurácia, com valores de sensibilidade que variaram de 73,7% a 96,7% e de especificidade de 98,5% a 99,4%. Destaca-se que abaixo deste ponto de corte, poucos valores verdadeiros foram encontrados. Por exemplo, para o pior valor de sensibilidade observado (73,7%) (sub-base C44.0 a C44.9 morfologia 8720 a 8780, pele melanoma), dos 803 pares verdadeiros, 592 estavam acima do escore 18. Outro fator

importante é que acima deste ponto de corte, a revisão se restringe a 624 pares, enquanto que a base de dados desta localização apresenta 5.745 pares.

Na sub-base de dados da topografia C51.0 a C58.9 (vulva, vagina, colo do útero, corpo do útero, útero, ovário, outros órgãos genitais femininos e os não especificados e placenta), a estratégia de blocagem relacionou 270.609 registros com sensibilidade 97,5% e especificidade de 99,3%. Verifica-se que, acima do escore 18 foram identificados como pares verdadeiros 7.537 casos em 9.382 registros. Por outro lado, nos 261.227 registros restantes abaixo deste ponto foram identificados, apenas, 191 pares verdadeiros (Tabela 14). Por isso, considera-se que, a partir do escore 18, há uma otimização no processo de leitura dos pares verdadeiros.

O mesmo ponto de corte (escore =18) foi o escolhido no relacionamento entre as 24 sub-bases de dados do RCBP-SP *versus* APAC-SIA/SUS (Tabela 15). Obteve-se valores de sensibilidade variando de 56,7% a 99,0% e de especificidade entre 93,0% a 98,4%. Por exemplo, no *linkage* da sub-base C44.0 a C44.9, morfologia 8720 a 8780 (pele melanoma), onde verificou-se o menor valor de sensibilidade (56,7%) e de especificidade (98,2%), dos 150 pares verdadeiros, 85 foram identificados acima do escore 18. Os demais 65 pares verdadeiros foram localizados nos restantes 4.810 registros.

Tabela 14 – Descrição dos pontos de corte, valores de sensibilidade e especificidade para o *linkage* RCBP-SP versus PRO-AIM. RCBP-SP, 1997 a 2005.

Topografia	Pontos de corte	Sensibilidade (%)	Especificidade (%)	Área da curva ROC	Nº de pacientes acima deste ponto de corte	Nº de pares verdadeiros neste ponto de corte	Nº de pacientes da base de dados total	Nº de pares verdadeiros da base de dado total
C00.0 a C09.9	≥ 18	97,1	98,9	97,3	2398	2378	11219	2447
C10.0 a 11.9	≥ 18	95,9	99,2	98,9	790	780	5056	813
C12.0 a 14.8	≥ 18	93,8	98,9	98,2	649	604	4906	644
C15.0 a C15.9	≥ 18	96,7	98,7	98,7	2716	2597	11568	2687
C16.0 a C17.9	≥ 18	93,8	98,7	98,5	7080	6915	19829	7370
C18.0 a C18.9	≥ 18	89,0	99,3	97,7	5581	5451	16819	6126
C19.0 a C21.8	≥ 18	93,9	98,4	97,5	3251	3111	11842	3312
C22.0 a C24.9	≥ 18	92,8	98,6	98,5	2228	2140	8547	2307
C25.0 a C26.9	≥ 18	95,5	98,7	98,4	3766	3658	12072	3831
C30.0 a C33.9	≥ 18	95,3	99,0	98,1	2038	1964	9577	2060
C34.0 a C34.9	≥ 18	92,7	98,6	98,0	8709	8535	21869	9204
C37.0 a C39.9	≥ 18	94,9	98,7	98,0	430	394	3168	415
C40.0 a C42.4	≥ 18	93,9	98,6	98,1	4677	4577	13580	4874
C44.0 a C44.9 (8720 a 8780)	≥ 18	73,7	99,4	94,0	624	592	5745	803
C44.0 a C44.9	≥ 18	81,7	98,7	94,7	3581	3362	21157	4113
C47.0 a C49.9	≥ 18	94,8	98,7	98,3	856	797	5220	841
C50.0 a C50.9	≥ 18	90,0	99,0	97,7	6344	6291	13316	6988
C51.0 a C58.9	≥ 18	97,5	99,3	99,0	9382	7537	270609	7728
C60.0 a C60.9/ C62.0 a C63.9	≥ 18	92,8	99,2	98,4	468	359	13246	387
C61.9	≥ 18	92,0	99,0	98,3	9169	7110	221695	7731
C64.0 a C68.9	≥ 18	94,2	98,7	97,4	3336	3218	12639	3417
C69.0 a C72.9	≥ 18	94,2	98,5	98,6	2925	2808	10974	2890
C73.0 a C75.9	≥ 18	85,7	99,1	96,2	672	635	5077	741
C76.0 a C80.9	≥ 18	89,9	99,0	97,8	7201	7104	17881	7902

Tabela 15 – Descrição dos pontos de corte, valores de sensibilidade e especificidade para o *linkage* RCBP-SP versus APAC-SIA/SUS. RCBP-SP, 1997 a 2005.

Topografia	Pontos de corte	Sensibilidade (%)	Especificidade (%)	Área da curva ROC	Nº de pacientes acima deste ponto de corte	Nº de pares verdadeiros neste ponto de corte	Nº de pacientes da base de dados total	Nº de pares verdadeiros da base de dados total
C00.0 a C09.9	≥ 18	96,0	97,6	98,9	1125	557	23481	580
C10.0 a 11.9	≥ 18	96,9	97,4	99,3	392	84	11339	87
C12.0 a 14.8	≥ 18	87,7	97,5	94,6	128	50	3131	57
C15.0 a C15.9	≥ 18	66,9	98,3	94,5	214	97	7011	145
C16.0 a C17.9	≥ 18	99,0	97,5	99,3	1046	389	26226	393
C18.0 a C18.9	≥ 18	81,2	98,4	97,9	648	371	16963	457
C19.0 a C21.8	≥ 18	83,2	96,2	92,4	554	297	7076	357
C22.0 a C24.9	≥ 18	83,6	98,7	99,2	227	51	12704	61
C25.0 a C26.9	≥ 18	80,5	98,6	92,2	106	33	5215	41
C30.0 a C33.9	≥ 18	98,9	96,8	99,1	962	356	18826	360
C34.0 a C34.9	≥ 18	80,8	98,0	92,9	603	449	8245	556
C37.0 a C39.9	≥ 18	81,0	98,2	96,4	53	17	1976	21
C40.0 a C42.4	≥ 18	98,4	97,3	99,3	1220	547	24723	556
C44.0 a C44.9 (8720 a 8780)	≥ 18	56,7	98,2	89,7	170	85	4981	150
C44.0 a C44.9	≥ 18	80,1	96,2	93,1	1457	827	17652	1032
C47.0 a C49.9	≥ 18	98,7	98,2	99,6	482	223	14757	226
C50.0 a C50.9	≥ 18	69,6	80,2	81,3	4890	2380	16069	3419
C51.0 a C58.9	≥ 18	88,1	96,4	98,2	3269	1616	48192	1834
C60.0 a C60.9/ C62.0 a C63.9	≥ 18	98,8	98,4	99,6	691	165	32242	167
C61.9	≥ 18	86,5	98,5	96,5	3268	1357	127734	1569
C64.0 a C68.9	≥ 18	98,7	96,6	99,0	1107	304	23729	308
C69.0 a C72.9	≥ 18	72,8	96,2	89,3	453	201	6873	276
C73.0 a C75.9	≥ 18	92,7	97,8	99,1	730	253	21537	273
C76.0 a C80.9	≥ 18	87,4	93,0	93,0	1291	536	11350	613

5. DISCUSSÃO

A vigilância do câncer é feita, fundamentalmente, pelos RCBP's e pelos RHC. Estes serviços são fundamentais para conhecer a magnitude desta doença, bem como analisar as taxas de sobrevida. No caso do RHC's também é possível fazer a análise da efetividade de esquemas terapêuticos. Estas análises orientam o planejamento de estratégias de prevenção, detecção precoce e a avaliação de novos tratamentos e métodos diagnósticos (INCA, 2007).

Contudo, a vigilância de uma doença ou o monitoramento de qualquer situação de saúde, administrativa ou financeira, depende da organização dos serviços, da cobertura do evento a ser medido, da qualidade da informação e de recursos humanos e financeiros. O RCBP-SP, em sua última publicação intitulada “Câncer em São Paulo 1997-2008: incidência, mortalidade e tendência de câncer no Município de São Paulo”, apresenta parâmetros que mostraram que sua cobertura e validade estão dentro de padrões internacionais, de acordo com os critérios da *International Agency for Research on Cancer - IARC* (IARC, 2007). Apesar disso, outras informações relevantes que influenciam o seguimento e a identificação de pacientes como endereço completo, data da última informação e status (vivo ou morto) apresentavam grande volume de casos ignorados. Sabe-se, também, que a coleta das variáveis referentes à data da última informação, assim como causa básica do óbito e nome da mãe não se fazia obrigatória na coleta do RCBP-SP até o ano de 2007. A necessidade de completar informações e de analisar a sobrevida de pacientes, tornou-se imprescindível e obrigatória.

Para melhorar a completitude das informações supracitadas, o RCBP-SP depende de recursos humanos e financeiros que, pelo tamanho da população do Município e a quantidade de dados é uma tarefa complexa, difícil e onerosa. Neste sentido, esta pesquisa teve como objetivo completar e melhorar a qualidade da informação da coorte de pacientes registrados no RCBP-SP, no período de 1997 a 2005, a partir da estratégia de *linkage* entre diversas bases de dados.

No estudo de HOLMAN et al. (2008), que descreve a experiência de uma década de estudos com *linkage* pelo *Western Australian Data Linkage System - WALDS*, foram observados os seguintes benefícios: 1) estudos longitudinais mais baratos e com informações mais completas; 2) melhora da qualidade do dados por meio da limpeza das duplicidades e correções de registros; e 3) manutenção da privacidade do paciente. Destaca-se que, nesta pesquisa, estes três benefícios também foram alcançados pelo RCBP-SP após o *linkage*.

Para se obter tais vantagens, o primeiro passo é a escolha das bases de dados a serem utilizadas no processo de *linkage*. Neste estudo, foram utilizadas as bases de dados do PRO-AIM, da FSEADE e da APAC-SIA/SUS.

A base de dados do PRO-AIM foi utilizada para a identificação dos óbitos ocorridos no Município de São Paulo, entre os anos de 1997 e 2007. A escolha de dois anos além do seguimento do RCBP-SP foi para captar um número maior de óbitos, além deste ser o período disponível pela CEINFO naquele momento.

A base de dados da FSEADE (2000 a 2009) foi utilizada para identificar os óbitos ocorridos fora do Município de São Paulo, pois sabe-se que existe uma taxa de evasão de óbito neste Município de 4,3% (TANIGUSHI et al., 2006). Destaca-se que, dos pacientes que vieram a óbito registrados na base do RCBP-SP identificados

pela vinculação com a base de dados da FSEADE, 21% faleceram fora do Município de São Paulo, mostrando que se faz necessário vincular a base de dados do RCBP-SP com bases de abrangência estadual.

Por fim, também se fazia necessário identificar os pacientes vivos. Para tanto, optou-se por escolher a base de dados da APAC-SIA/SUS, pois, nesta, encontram-se os pacientes que receberam ou realizaram qualquer procedimento de alta complexidade e custo. Foram solicitadas à SES-SP os dados referentes as APAC-SIA/SUS para o período de 1997 a 2007, porém só foi possível obter as informações do período de agosto de 2003 a dezembro de 2007, disponibilizadas pela SMS-SP.

Ainda em relação à escolha das bases, para alcançar um bom resultado no processo de *linkage*, as bases de dados envolvidas deveriam apresentar uma alta cobertura do evento medido e uma boa qualidade no registro da informação (CAMARGO e COELI, 2000; NITSCH et al., 2006). Neste estudo, verificou-se que as bases de dados do PRO-AIM e FSEADE apresentavam uma cobertura superior a 90% para o Município de São Paulo (MS, 2009).

Quanto à base de dados da APAC-SIA/SUS, embora a cobertura desta seja referente a procedimentos e medicamentos de pacientes que utilizaram os serviços do SUS, CARVALHO (2009) relata que os procedimentos de alta complexidade registrados nesse sistema, devido ao seu custo elevado e a limitada cobertura pelos planos privados de saúde, representam 90% do total de procedimentos realizados na região Sudeste. Para o Município de São Paulo, verifica-se que a taxa de cobertura pelos planos de saúde é de 55,5% e, segundo a pesquisa realizada pelo Instituto de Estudos de Saúde Suplementar (IESS), 26% das pessoas que possuem assistência médica privada, também, fazem uso do SUS (IESS, 2010; ANS, 2011).

De posse das bases de dados do PRO-AIM e APAC-SIA/SUS, iniciou-se o processo de preparação dos arquivos para o relacionamento com as duas bases de dados que foram entregues em diversos arquivos. A primeira etapa realizada foi agregar as bases de dados, facilitando assim, todo processo de limpeza e padronização para o relacionamento. Após esta etapa, foi executada a limpeza das bases de dados, a saber, RCBP-SP, PRO-AIM, FSEADE e APAC-SIA/SUS, que correspondeu à identificação de duplicidades, valores ignorados, retirada de indivíduos sem identificação e a retirada de campos e caracteres excedentes.

Na base de dados do RCBP-SP foram encontrados 18 casos duplicados, pois continham o mesmo nome, data de nascimento, topografia e morfologia. Ressalta-se que, durante o ano de 2008 a equipe deste registro fez uma verificação manual de todos os casos, a fim de eliminar as possíveis duplicidades. Atualmente, para o controle de qualidade desta base, o RCBP-SP possui dupla digitação dos casos novos e antes de serem agregados, estes ficam em uma banco denominado "temporário" para que seja feita a revisão por uma supervisora. Somente após estas etapas, o novo caso é agregado ao banco definitivo (RCBP-SP, 2011).

Em relação a valores ignorados, não foi observado a ausência de registro de dados para as variáveis "nome do paciente" e "sexo". Entretanto, 35,9% da "data de nascimento" apresentava ausência de informação e muitas datas de nascimento estavam incompletas e apresentavam registro como 01 de janeiro de um ano referido.

Vale lembrar que um número elevado de valores ignorados pode dificultar o *linkage* dos bancos de dados, devido a não identificação de alguns pares verdadeiros, fato agravado pela grande quantidade de homônimos existentes nas bases de dados. As variáveis de confirmação do *linkage*, como nome da mãe e endereço de

residência, apresentavam baixa completude, respectivamente, 0,5% e 23,6%. Além disso, registros com a data de nascimento incompletos (01/01/ de qualquer ano) também acarretava na recusa de um par verdadeiro.

Na base de dados do PRO-AIM identificou-se que as variáveis "nome do paciente", "data de nascimento" e "sexo" apresentavam menos de 1% de valores ignorados. Com isso, um mínimo de registros foram perdidos. Foram excluídos desta base de dados somente os óbitos que não apresentavam a identificação do nome. Quanto à qualidade do registro, não foram identificados maiores problemas. Os registros que apresentavam no campo nome do paciente as palavras “indigente” ou “suspeito” ou outro caractere estranho, totalizaram 0,85% da base de dados. Para a data de nascimento, o preenchimento com datas incoerentes não ultrapassou a 5%.

Nos anos de 2002 e 2003, a FSEADE realizou pesquisa para avaliar a qualidade dos dados produzidos pelos sistemas municipais de saúde, analisando cada variável presente na base de dados e sua fidedignidade em relação à DNV e DO. Observou-se que campos em que a informação é pré-codificada a qualidade do dado é boa. Os campos como Município de ocorrência, idade da mãe, data do parto, data de nascimento e sexo apresentavam menos de 2% de divergência nas informações. Por outro lado, algumas variáveis que necessitavam de manuais de codificação e treinamento para os profissionais envolvidos, apresentavam problemas de preenchimento. Os autores ressaltaram que erros de codificação da variável causa básica do óbito variaram conforme o Município de ocorrência. Para o Estado de São Paulo este erro foi de 15% (WALDVOGEL et al., 2008).

Outro fato relevante foi a construção de uma base de dados unificada de nascimentos e óbitos, a partir da parceria realizada, em 2005, com a SES-SP, o que

contribui, ainda mais, para a melhora da cobertura e da qualidade do dado propiciando uma base de dados mais completa e precisa (WALDVOGEL et al., 2008).

Na base de dados da APAC-SIA/SUS foi verificado que 100% das variáveis "sexo" e "nome do paciente" não apresentavam valores ignorados. Para a variável "data de nascimento" o percentual de informações ignoradas foi de 0,1%. Embora, houvesse boa completitude das variáveis, a qualidade dos registros mostrou-se inferior em relação às bases de dados do RCBP-SP e do PRO-AIM, pois no nome do paciente, em 5.397 casos havia o caractere apóstrofo ('), 4.413 ponto-e-vírgula (;), 1.736 hífen (-) e 5.878 vírgula (,). Destaca-se que nestes casos, estes caracteres foram retirados e a informação resgatada.

Após esta etapa, foi necessário deixar a base de dados com um único registro para cada paciente e neste, deveria estar presente a data da última vez que um procedimento ou medicamento foi solicitado. Como os registros desta base são voltados para o faturamento de procedimentos e medicações, um único paciente pode apresentar vários registros.

Ao empregar a técnica de *linkage* probabilística, o procedimento de retirada dos múltiplos registros de um mesmo paciente fez-se essencial, dado que o produto de um relacionamento é o resultado da multiplicação dos registros nas bases. Neste sentido, os casos excedentes aumentariam o número de pares e, conseqüentemente, aumentariam o tempo de processamento e dificultariam a identificação de pares verdadeiros.

No entanto, erros de preenchimento da base de dados da APAC-SIA/SUS foram observados com frequência. Observou-se que a partir do primeiro registro, os

demais, para aquele mesmo paciente, apresentavam falhas na informação como nomes abreviados, endereço incompleto, CPF zerado ou em branco, ou ainda diferentes números deste documento para um mesmo paciente, nome da mãe em branco ou com as palavras "O MESMO", fazendo referência ao registro anterior.

Outros estudos também mostraram problemas no preenchimento ou ausência de informação nas variáveis. BRITO et al. (2005) observaram que muitas variáveis com campos obrigatórios não eram digitadas no sistema informatizado da APAC-SIA/SUS, destacando que se os campos fossem adequadamente preenchidos, a necessidade de recorrer a outras fontes de dados complementares seria reduzida.

CHERCHIGLIA et al. (2007) ressaltaram os mesmos problemas observados nesta pesquisa, mostrando preenchimento inadequado da variável CPF (na qual um mesmo paciente poderia apresentar mais de um número) e na identificação das palavras "A MESMA" no campo nome da mãe.

Desta maneira, a limpeza prévia do banco de dados foi prejudicada. Na retirada dos múltiplos registros para um mesmo paciente, este deveria apresentar as mesmas informações em quatro variáveis (nome do paciente, data de nascimento, CPF e nome da mãe). Logo, um registro com possíveis erros no preenchimento em alguma destas quatro variáveis não poderia ser reconhecido como um caso de registro excedente.

Quanto ao CPF, que seria o principal critério de retirada de mais de um registro para um mesmo paciente, verificou-se que, em mais de 60% destes, este campo estava preenchido com o número zero (0), ou, ainda, havia registros com dígitos faltantes.

Evidencia-se que, nesta etapa de retirada dos registros excedentes de um mesmo paciente, estes não puderam ser lidos manualmente, devido ao tamanho da base de dados da APAC-SIA/SUS (aproximadamente, 32 milhões de registros). Desta forma, todo o processo foi feito via programação no programa SPSS versão 17.0 para *Windows*, o que pode ter acarretado uma perda destes casos.

A falta de rigor no preenchimento dos dados da APAC-SIA/SUS também foi observada na variável que identifica o Município de residência do paciente. No registro desta variável, na qual o preenchimento deve ser feito com o código do Município estabelecido pelo IBGE, foram observados caracteres alfanuméricos, códigos supostamente errados e incompletos. Para o Município de São Paulo identificou-se a sigla da UF, o código do Município com seis dígitos (355030) ou sete dígitos (3550308). Observou-se, também, que, provavelmente, em alguns registros houve a inversão dos últimos dígitos e em outros erros de digitação. Desta forma, tornou-se complexa a identificação dos pacientes avaliados.

BRITO et al. (2005) avaliaram a concordância de dados clínicos e demográficos das APAC-SIA/SUS para pacientes de câncer de mama, confrontado com dados de prontuários em uma amostra de 310 pacientes do Estado do Rio de Janeiro. Os autores observaram boa concordância para as variáveis data de nascimento e data do diagnóstico (respectivamente, $Kappa=0,96$ e $Kappa=0,92$). Por outro lado, a variável Município de residência apresentou uma concordância razoável ($Kappa=0,52$), mostrando, também, dificuldades semelhantes às desta pesquisa.

Outra variável que apresentou problemas foi a "data do diagnóstico do tumor", na qual é possível identificar a data do caso novo. Destaca-se que esta variável, no banco de dados da APAC-SIA/SUS, não estava registrada em um campo

exclusivo. Onde deveria constar somente esta data, havia informações sobre topografia e morfologia para pacientes que receberam qualquer tratamento relativo ao câncer e, em muitos casos, o registro estava confuso, dificultando sua extração, utilizando um único campo.

Outra limitação desta variável, é que seu registro é feito pelo médico responsável pelo tratamento do paciente, no ato do preenchimento do laudo de solicitação de procedimentos oncológicos, e estima-se que haja um erro de, aproximadamente, 10% (GOMES Jr e ALMEIDA, 2004). Nesta pesquisa, foi utilizada esta data do diagnóstico do tumor para a identificar um possível caso subnotificado pelo RCBP-SP. Todavia, ressalta-se que a qualidade do dado pode estar comprometida, devido às possíveis falhas de registro e no processo de concatenação ou extração dos dados.

Após a limpeza das bases de dados (RCBP-SP, PRO-AIM, FSEADE e APAC-SIA/SUS), deu-se início à padronização. Esta fase objetivou deixar compatível os campos que foram utilizados no processo de vinculação das bases de dados.

Nesta etapa, mais de um *software* foi utilizado. Programas com boa capacidade de memória e de fácil manipulação facilitaram a aplicação de rotinas e a otimização do tempo nestas etapas. Nesta pesquisa, optou-se por utilizar o programa SPSS versão 17.0 para *Windows* na identificação de duplicidades, padronização, leitura manual e análise estatística. O programa Reclink III versão 3.1.6 foi utilizado para padronização, cálculo do escore e para o processamento do *linkage* probabilístico. Para o *linkage* determinístico, a FSEADE trabalhou numa plataforma denominada *Visual Basic*, hospedada em um servidor de alta capacidade, em *SQL*

Server, para o armazenamento e manipulação das bases de dados, o que otimizou a vinculação das bases.

A padronização das variáveis permitiu que nomes com grafias diferentes, porém com mesmo fonema fossem equiparados, assim como o tamanho dos campos e tipo de caracteres. A realização deste processo tornou-se essencial, pois se sabe que as etapas de coleta e digitação das informações são passíveis de interpretações diferentes por parte dos profissionais e de erros nas etapas de registro do dado.

Outro ponto importante da padronização, é que, nesta etapa, foi possível resgatar a informação do ano de nascimento dos pacientes do RCBP-SP, pois a partir da data do diagnóstico menos a idade, foi obtido o provável ano de nascimento. Com isso, a ausência de registro para o ano de nascimento foi reduzida de 35,9% para 9,8%.

Pelo exposto, conclui-se que, antes de escolher qual a técnica de *linkage* a ser utilizada, os primeiros cuidados estão na escolha das bases de dados e na preparação das mesmas. Para a localização dos óbitos, as bases de dados do PRO-AIM e da FSEADE foram extremamente úteis e contribuíram de forma relevante. Na localização dos pacientes vivos, a única opção disponível foi a base de dados da APAC-SIA/SUS que apresentou os problemas já descritos. Mesmo assim foi útil no processo de *linkage*, embora acarretasse um trabalho extra em sua preparação. Com isso, há a necessidade de se utilizar outras bases de dados para seguimento de pacientes vivos que, no Brasil, ainda não estão disponíveis para estudos epidemiológicos. Exemplo disso é a base de dados disponível no Tribunal Regional Eleitoral.

Um segundo passo é a escolha da técnica de *linkage* a ser empregada. As duas técnicas, determinística ou probabilística, já foram validadas no Brasil e em diversos países como Canadá, Áustria, Escócia, Holanda e França (WOSCOPS, 1995; CAMARGO JR. e COELI, 2000; WALDVOGEL, 2002; COELI et al., 2003; OBERAIGNER e STUHLINGER, 2005; COUTINHO e COELI, 2006; LI et al., 2006; NITSCH et al., 2006; MÉRAY et al., 2007; PACHECO et al., 2008; MIGOWSKI et al., 2011).

Vale lembrar que a opção do tipo de *linkage* empregado deve considerar os objetivos do relacionamento e suas implicações epidemiológicas. Por exemplo, em um hospital com um número alto de pacientes perdidos de seguimento, faz-se necessário identificar os óbitos para dar baixa nos prontuários e, nestes casos, sugere-se o uso do *linkage* determinístico, dado que esta técnica trabalha com a certeza na combinação das variáveis de relacionamento. Sendo assim, não há a possibilidade de serem excluídos prontuários de pacientes ativos. Esta técnica, também é usada preferencialmente para a identificação de sub-registros, pois não pode haver dúvidas na inclusão de um caso novo (WALDVOGEL, 2002).

Quanto ao *linkage* probabilístico, pesa a crítica deste ser menos transparente, dado que o poder de discriminação dos pares é baseado em situações de ponderação. No entanto, esta técnica torna-se uma solução viável quando a qualidade dos registros a serem vinculados é baixa, podendo as variáveis apresentarem erros e falhas na informação, ausência de um indicador unívoco ou poucas variáveis para a vinculação dos registros em bases de dados com grande número de casos (MACHADO, 2004; TROMP et al., 2011).

Nesta pesquisa, optou-se por utilizar o *linkage* probabilístico para os bancos de dados do PRO-AIM e da APAC-SIA/SUS, partindo do pressuposto que as bases de dados poderiam apresentar falhas no registro da informação não detectadas anteriormente, nos processos de limpeza e padronização. O *linkage* determinístico foi utilizado entre as bases de dados do RCBP-SP *versus* FSEADE, devido a característica da base de dados de óbito do Estado de São Paulo apresentar um conjunto de variáveis com alta qualidade de registro.

Desta forma, sabendo-se das possíveis limitações da técnica probabilística de *linkage*, fez-se necessário conhecer dentre todos os casos pareados, quais seriam os pares incorretos e identificar quantos pares poderiam ser formados a partir de um único registro, a fim de reduzir o grau de incerteza na vinculação de um par verdadeiro. Para o *linkage* determinístico, devido a sua metodologia de vinculação, não houve a necessidade de se identificar pontos de corte para a seleção de pares verdadeiros. Assim, este relacionamento se deu de forma mais otimizada, sendo todo processo, que inclui limpeza, padronização, vinculação dos registros, seleção automática e verificação visual, executado em três semanas.

No *linkage* probabilístico, foi realizada a leitura manual de todos os registros pareados, a partir do escore mínimo de concordância. Nesta leitura manual a cada par verdadeiro foi atribuído a nota um (1) e a cada par identificado como falso, atribuiu-se a nota zero (0), possibilitando confrontar com o escore calculado e obter os valores de sensibilidade e especificidade do método. A partir disso, foi definido mais um objetivo para esta pesquisa ao propor pontos de corte para facilitar a identificação e a revisão dos pares verdadeiros. Nesta etapa foram lidas 648 bases de dados com tamanho máximo de 1.309.283 registros pareados. O tempo gasto envolvendo o

processamento das bases de dados e a leitura manual, foi de, aproximadamente, 4.836 horas de trabalho. Neste estudo participaram três pesquisadores, trabalhando, em média, quatro horas/dia, durante dezoito meses. Ressalta-se que este tempo foi essencial, dado que houve a necessidade de identificar a melhor ou as melhores estratégias de blocagem.

Para todas as topografias, independente do tamanho da base de dados, no *linkage* entre o RCBP-SP e o PRO-AIM, a mesma estratégia de blocagem (PBLOCO + UBLOCO + SEXO) com os campos de pareamento nome do paciente e data de nascimento mostrou-se a mais efetiva. Somente nos casos em que o câncer é restrito a um único sexo, por exemplo, próstata, foi retirada da estratégia de blocagem o campo SEXO. Em todas as 24 sub-bases, esta blocagem identificou o maior número de pares verdadeiros com o tempo de execução entre os menores, assim como as duplicidades.

No *linkage* entre as 24 sub-bases de dados do RCBP-SP com a base de dados da APAC-SIA/SUS, as estratégias de blocagem também só apresentaram modificações para aquelas topografias restritas a um único sexo. Quanto aos campos de pareamento, em onze sub-bases o melhor desempenho foi a combinação dos campos de pareamento "nome do paciente" e "ano de nascimento".

Pelo exposto, sugere-se que a estratégia de blocagem seja PBLOCO + UBLOCO + SEXO e que, devido a um menor número de registros ignorados no campo de pareamento "ano de nascimento" em comparação ao campo "data de nascimento", respectivamente, 9,8% e 35,9%, para algumas topografias, seja utilizado o ano de nascimento. No estudo de KARMEL et al. (2010), que objetivou descrever os aspectos empíricos da técnica de *linkage*, também verificou-se que a

presença de valores ignorados diminuiu a probabilidade de identificar um par verdadeiro.

Também foram identificadas as estratégias de pior desempenho. Em todas as topografias e para ambos os *linkages*, (RCBP-SP versus PRO-AIM e versus APAC-SIA/SUS) quando havia quatro campos para uma chave de blocagem o número de pares relacionados ficava reduzido e não se obtinha o escore máximo. Verificou-se que nestes *linkages* o escore atingido não ultrapassou 16,69.

Nesta pesquisa, optou-se por utilizar somente uma estratégia de blocagem, pois todos os registros da base de dados foram lidos "um a um". Por outro lado, a literatura ressalta que para um melhor desempenho do método probabilístico, recomenda-se que mais de uma estratégia seja empregada, variando de três a cinco passos, pois aumentam a chance de um par verdadeiro ser capturado. Neste processo, os pares identificados com escores baixos na primeira estratégia de blocagem, são enviados a um segundo arquivo para ser aplicada uma nova estratégia de blocagem, esta menos restrita (COELI e CAMARGO Jr., 2002; NAKHAEI et al. 2007; SOUSA et al. 2008; FONSECA et al., 2010; MIGOWSKI et al. 2011).

Todavia, outras pesquisas mostraram que o número de pares verdadeiros encontrados a partir da segunda estratégia de blocagem representaram 4% a 10% do total de pares verdadeiros. Além disso, estas estratégias menos restritas aumentam o número de registros relacionados com maior tempo de execução (COELI e CAMARGO Jr., 2002; TANIGUSHI et al., 2006; MIGOWSKI et al., 2011). Evidencia-se que, em bancos com grande volume de dados, como os utilizados nesta pesquisa, o uso de mais de uma estratégia pode representar alto esforço com pequeno benefício.

Com a identificação da melhor estratégia de blocagem e campos de pareamento, foi sugerido um ponto de corte para o escore. Para todas as topografias, indica-se que seja utilizado o escore dezoito (18), dado que acima deste ponto de corte encontram-se a maioria dos pares verdadeiros e se otimiza a revisão manual. Nosso achado corrobora com o estudo de TROMP et al. (2011) no qual foi observado que a partir do escore 18 estavam localizados 94% dos pares verdadeiros.

Neste estudo, o ponto de corte sugerido (escore = 18) apresentou valores de sensibilidade acima de 73,3% e especificidade de 98,5% entre o *linkage* do RCBP-SP versus PRO-AIM. Para o *linkage* entre o RCBP-SP versus APAC-SIA/SUS a sensibilidade variou de 56,7% a 99,0% e a especificidade 80,2% a 98,7%.

Destaca-se que o menor valor de sensibilidade observado nesta pesquisa foi atribuído à sub-base de dados pele melanoma (topografias C44.0 a C44.9 M8720 a 8780). Observou-se que nesta sub-base muitos casos registrados possuíam nomes que apresentavam grafias complexas e fonemas diferentes dos identificados na população brasileira como SILVA, SANTOS, PEREIRA, entre outros. Neste contexto, os sobrenomes que apresentavam muitas consoantes e diferentes fonemas, possivelmente eram de origem européia, tornando difíceis de serem compreendidos e, conseqüentemente, digitados, aumentando, assim a probabilidade de erro. Sabendo que a qualidade do registro está diretamente ligada ao poder discriminante da variável, acredita-se que, por apresentar grafias pouco habituais, o erro de preenchimento da ficha de coleta e o erro de digitação foram acima do observado para outros tipos de topografias, acarretando valor menor no escore. Para avaliação desses pares, foi necessária a leitura em voz alta por mais de um pesquisador envolvido no estudo, pois esta possibilitou verificar a similaridade dos fonemas.

O elevado número de registros na base de câncer de pele do tipo melanoma com supostos problemas de grafias e fonemas revela que as características das bases de dados devem ser levadas em consideração. Nesta pesquisa, sabendo-se que o risco para o melanoma é ter ascendência ou origem européia, a probabilidade de se observar nomes pouco comuns foi maior que nas outras topografias (BRADFORD, 2009; GONÇALVES et al., 2011). Ressalta-se que o mesmo fato foi observado para a topografia de câncer de mama (C50.0 a C50.9) no *linkage* entre a base do RCBP-SP e APAC-SAI/SUS.

A partir da identificação dos pares verdadeiros, foram analisados os resultados, segundo a completitude e qualidade das bases de dados. Para os óbitos, o processo de *linkage* entre a base de dados do RCBP-SP *versus* PRO-AIM detectou 89.461 pares verdadeiros, dos quais 81% dos registros já constavam com a data do óbito. Destaca-se que o PRO-AIM, anualmente, realiza o processo de *linkage* probabilístico com o RCBP-SP. Contudo, este relacionamento era feito somente com os casos que apresentaram óbito por câncer, com a finalidade de fechar o seguimento e resgatar possíveis casos incidentes não detectados pelo RCBP-SP. Por isso, a partir de agora, o *linkage* passará a ser feito com a base contendo todos os óbitos.

Além das informações complementadas, foram agregadas as informações do nome da mãe (2.398,1%) e endereço (69,2%). Ressalta-se que a completitude da variável endereço feita via os registro de óbito e pelos registros da APAC-SIA/SUS podem apresentar vieses, pois o endereço do óbito ou do endereço atual podem não ser o mesmo do momento do diagnóstico, implicando na distribuição espacial errônea dos casos de câncer. Indica-se que, nestes casos, esta informação não seja utilizada para realizar análises de georreferenciamento. Porém, em situações que haja

necessidade de realizar a busca de pacientes vivos, tal informação mostra-se relevante.

A causa básica do óbito, sendo ela câncer ou não câncer, entrou como uma nova variável e, dos óbitos encontrados, 87,6% tiveram como causa básica o câncer. Esta proporção de causas de óbito por câncer deve ser analisada com cautela por dois motivos. Primeiramente, pode haver a ocorrência de viés de informação, que está atrelado ao preenchimento da DO. Acredita-se que em óbitos de pacientes com câncer há uma tendência do médico em preencher o atestado, dando como causa da morte causas existentes que não foram as que levaram a este desfecho (LAURENTI et al., 2008). Em segundo lugar, no processo de *linkage*, a causa básica do óbito foi uma das variáveis utilizadas como parâmetro de confirmação na identificação de um par verdadeiro. Porém, conforme descrito na metodologia, a causa básica do óbito não foi utilizada isoladamente para identificar um registro vinculado. Para ser um par verdadeiro, fazia-se necessário, ao menos, mais duas condições das variáveis de relacionamento e de confirmação; por exemplo, endereço compatível e ano do óbito.

Em relação à completude dos dados para os pacientes vivos, o *linkage* com a base de dados da APAC-SIA/SUS detectou 13.571 pares verdadeiros. A partir da eliminação das duplicidades, foram identificados 12.926 pacientes e, em 1.145 (8,9%) casos que já apresentavam a data da última informação, o dado foi atualizado. No total, o ganho desta variável foi de 80,0%, passando de 11.462 para 20.628 casos com data da última informação.

Revela-se que, no *linkage* com a base de dados da APAC-SIA/SUS, houve maior proporção de pares duplicados (4,8%) em relação ao *linkage* com a base de dados do PRO-AIM (0,3%). Isto ocorre porque a base de dados da APAC-SIA/SUS é

um banco que a cada novo procedimento ou renovação de medicamento o paciente é registrado novamente, aumentando a probabilidade deste registro ser relacionado mais de uma vez, configurando os casos de pares falso-positivos.

Quanto ao *linkage* com os óbitos da base de dados da FSEADE, dos 30.480 pares vinculados, 20.463 (67,1%) dos registros foram considerados pares verdadeiro. Obteve-se com este *linkage*, um ganho de 80,9% da informação na data do óbito. Destaca-se que a completitude das variáveis nome da mãe e endereço foram, respectivamente, 1.136,2% e 24,9%.

No processo de verificação visual do *linkage* determinístico, a causa básica do óbito, sendo esta câncer, também foi utilizada para a identificação de um par verdadeiro. Contudo, esta informação nunca foi utilizada como a única variável de confirmação. Nesta etapa, a aceitação de um par verdadeiro levou em consideração a similaridade das informações presentes nas variáveis endereço, combinada com data do óbito ou data de nascimento.

Para avaliar o impacto que o processo de *linkage* ocasionou nas estatísticas do RCBP-SP, estas foram analisadas antes e após o processo. Foram observadas mudanças em todos os parâmetros.

Nesta pesquisa, o CBI estava subestimado em 4,3%. Da mesma forma, outros estudos que fizeram uso da técnica de *linkage* para a completitude da informação, mostraram as diferenças no número de eventos registrados. O estudo de CAVALCANTE et al. (2005), realizado no Município de Fortaleza (Ceará), fez uso da estratégia de *linkage*, relacionando as bases de dados do SINASC, Laboratório Central do Ceará (Lacen-CE), SINAN e uma base de dados gerada a partir da busca ativa em prontuários de maternidades públicas. O objetivo dos pesquisadores era

melhorar a vigilância das gestantes infectadas pelo HIV e das crianças expostas ao vírus. Após este processo, identificou-se uma subnotificação de 70,3% de mulheres infectadas pelo HIV.

A pesquisa de DRUMOND et. al (2008) observou, após o *linkage* entre a base de dados do SINASC e SIH, uma subnotificação de 8,4% dos nascimentos no Estado de Minas Gerais, para o ano de 2001. DUNBAR et al. (2011), em pesquisa realizada com duas comunidades urbanas na África do Sul, com o objetivo de avaliar a acurácia e a completude do registro de tratamento por tuberculose, por meio da técnica de *linkage*, verificaram um sub-registro de 50% nos casos da doença.

Nesta pesquisa, houve um ganho de 25,8% no CBM após o processo de *linkage*. No estudo de PEREIRA et al. (2007) realizado no Rio de Janeiro, entre 1999 e 2001, foi verificado que havia uma subestimação de 20% no coeficiente de mortalidade neonatal. Da mesma forma, a pesquisa de RAFAEL et al. (2011), que fez uso das bases de dados do SIM e SIH no Estado do Maranhão, observou uma subestimação nos coeficientes de mortalidade neonatal e infantil de 24,9% e 19,8%, respectivamente.

Destaca-se que a técnica de *linkage* é utilizada para diversos estudos, apresentando eficiência na identificação de casos subnotificados. Nesta pesquisa, os óbitos por câncer identificados pelo local de residência serão reavaliados por profissionais do RCBP-SP e, a partir da confirmação, passarão a integrar a base de dados definitiva.

Na atual pesquisa, também foi realizada a análise sobrevida global dos pacientes do RCBP-SP. Verificou-se que, após o *linkage*, as probabilidades de sobrevida aumentaram ao longo dos anos analisados para ambos os sexos, todas as

faixas etárias e para as topografias de câncer. Pode-se destacar, neste caso, a importância de se relacionar não só bases de dados de óbitos e, sim, de pacientes vivos. Pois, os pacientes vivos identificados via APAC-SIA/SUS, proporcionaram um ganho na sobrevida ao longo dos anos.

No Brasil, pouco se trabalha com análises de sobrevida utilizando bases de dados populacionais. Pode-se citar o estudo pioneiro de sobrevida em pessoas vivendo com HIV/Aids realizado pelo Programa Estadual de DST/Aids do São Paulo e a FSEADE. Esta pesquisa, utilizou a base de dado BIP-Aids gerada a partir do *linkage* determinístico, mostrando não só a eficácia da metodologia como sua contribuição epidemiológica, que permitiu conhecer o comportamento desta epidemia ao longo das décadas, podendo ser observado o comportamento segundo sexo, faixa etária, categoria de exposição e critério de definição de caso (TANCREDI et al. 2010).

Neste contexto, a vinculação entre bases de dados de perfil administrativo como APAC-SIA/SUS, SIAB, Tribunal Regional Eleitoral com bases que trazem informações epidemiológicas como óbitos, nascimentos e notificações de doenças, enriquecem e ampliam as possibilidades de análises, além de gerarem informações de melhor qualidade e serem menos onerosas, buscando sempre o aprimoramento dos indicadores de saúde.

Pelo exposto, conclui-se que a estratégia de *linkage* foi efetiva, propiciando uma melhora na qualidade do dado nas análises realizadas pelo RCBP-SP. Enfatiza-se, também, que os resultados encontrados nesta pesquisa, refletem a capacidade das bases de dados, quando trabalhadas de maneira conjunta, em fornecerem informações de qualidade. Contudo, recomenda-se que para a completude das

informações de pacientes vivos, sejam utilizadas outras fontes de dados e que as autoridades reconheçam a importância e seriedade deste trabalho, disponibilizando suas bases de dados para atualizar o seguimento de pacientes vivos.

6. CONCLUSÕES

1. As técnicas de *linkage* tanto probabilística quanto determinística foram efetivas para completar/aperfeiçoar as informações da base de dados do RCBP-SP. Houve um ganho informações de 101,5% para o endereço, 80,0% para a data da última informação e 31,5% para a data do óbito.
2. Após o processo de *linkage*, identificou-se uma subestimação de 4,3% no coeficiente bruto de incidência e de 25,8% no coeficiente bruto de mortalidade.
3. Após o uso da técnica de *linkage*, foi verificado que os valores para a sobrevida global acumulada estavam subestimados para homens e mulheres, todas as faixas etárias e topografias.
4. A melhor estratégia de blocagem foi PBLOCO + UBLOCO + SEXO, com os campos de pareamento nome do paciente e data de nascimento ou ano de nascimento e o melhor ponto de corte o escore 18. Neste ponto de corte, os valores para a sensibilidade variaram de 56,7% a 99,0% e especificidade de 80,2 % a 98,7%.

7. REFERÊNCIAS

1. Almeida MF. Descentralização de sistemas de informação e o uso das informações a nível municipal. IESUS. 1998;7;27-33.
2. Almeida MF, Alencar GP; França-Jr I, Novaes HMD, Siqueira AAF, Schoeps D et al. Validade das informações das declarações de nascidos vivos com base em estudo caso-controle. Cad Saúde Pública. 2006;22:643- 652.
3. Almeida MF, Mello-Jorge MH. O uso de *linkage* de sistemas de informação em estudos de coorte sobre mortalidade neonatal. Rev Saúde Pública. 1996;30:141-147.
4. ANS - Agência Nacional de Saúde; 2011 [acesso em 20 out 2011]. <http://www.ans.gov.br/index.php/materiais-para-pesquisas/perfil-setor/dados-e-indicadores-do-setor>
5. Baldijão MF. Os sistemas de informação em saúde. São Paulo Perspect. 1992; 6:21-28.
6. Barfield WD, Clements KM, Lee KG, Kotelchuck M, Wilber N, Wise PH. Using linked data to assess patterns of Early Intervention (EI) referral among Very Low Birth Weight Infants. Matern Child Health J. 2008;12:24-33.
7. Bradford PT. Skin cancer in skin of color. Dermatol Nurs. 2009;21(4):170-178.
8. Brito C, Portela MC, Vasconcellos MTL. Avaliação da concordância de dados clínicos e demográficos entre autorizações de procedimentos de alta complexidade oncológica e prontuários de mulheres atendidas pelo sistema

- único de saúde no Estado do Rio de Janeiro, Brasil. *Cad Saúde Pública*. 2005;21(6):1829-1835.
9. Camargo-Jr KR, Coeli CM. Reclink III: relacionamento probabilístico de registros. Versão 3.1.6.3160. Rio de Janeiro; 2007.
 10. Camargo-Jr KR, Coeli CM. *Reclink*: aplicativo para o relacionamento de bases de dados, implementando o método *probabilistic record linkage*. *Cad Saúde Pública*. 2000;16:439-447.
 11. Carvalho DMT. Sistema de Infomação Hospitalares do SUS-SUS-SUS. In: MS-Ministério da Saúde. A experiência brasileira em sistemas de informação em saúde/ Ministério da Saúde, Organização Pan-Americana da Saúde, Fundação Oswaldo Cruz. – Brasília: Editora do Ministério da Saúde, 2009,49-70.
 12. Cavalcante MS, Ramos Jr AN, Pontes LRSK. Relacionamento de sistemas de informação em saúde: uma estratégia para otimizar a vigilância das gestantes infectadas pelo HIV. *Epidemiol Serv Saúde*. 2005;14(2):127-133.
 13. Cherchiglia ML, Guerra-Jr AA, Andrade ELG; Machado CJ, Acúrcio FA, Meira-Jr W et al. A construção da base de dados nacional em terapia Renal Substitutiva (TRS) centrada no indivíduo: aplicação do método de *linkage* determinístico-probabilístico. *Rev Bras Est Pop São Paulo*. 2007;24:163-167.
 14. Coeli CM, Camargo Jr KR. Avaliação de diferentes estratégias de blocagem no relacionamento probabilístico de registros. *Rev Bras Epidemiol*. 2002;5(2):185-196.
 15. Coeli CM, Blais R, Costa MCE, Almeida LM. Probabilistic linkage in household survey on hospital care usage. *Rev Saúde Pública*. 2003;37:91-99.

16. CEINFO- Coordenação de Epidemiologia e Informação. Sistema de monitoramento e avaliação da qualidade das bases de dados - SUS: histórico, conceitos, indicadores e métodos. São Paulo; 2011.
17. Coutinho ESF, Coeli CM. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevivência. *Cad Saúde Pública*. 2006;22:2249-2252.
18. DATASUS. Sistemas de Informação; 2008a [acesso em 14 mai 2008]. Disponível em: <http://portal.saude.gov.br/portal/svs/area>.
19. DATASUS. Sistemas e aplicativos. CNES - Cadastro Nacional de Estabelecimentos de Saúde; 2011a [acesso em 15 fev 2011]. <http://cnes.datasus.gov.br/>
20. DATASUS. Sistemas de Informação. Sistemas de Informação sobre Mortalidade; 2008b [acesso em 14 mai 2008]. Disponível em: <http://portal.saude.gov.br/portal/svs/area>.
21. DATASUS. Sistemas de Informação. Sistemas de Informação sobre Nascidos Vivos; 2008c [acesso em 14 mai 2008]. Disponível em: <http://portal.saude.gov.br/portal/svs/area>.
22. DATASUS. Sistemas de Informação. Sistemas de Informação da Atenção Básica; 2011b [acesso em 15 fev 2011]. Disponível em: <http://datasus.gov.br/siab/siab.htm>
23. Deshaies JC, Seidman DR. Health information systems. *Socio-Econ Plan Sci*. 1971;515-533.

24. Drumond EF, Machado CJ, Franca E. Underreporting of live births: measurement procedures using the Hospital Information System. *Rev Saúde Pública*. 2008;42(1):55-63.
25. Dunbar R, Lawrence K, Verver S, Enarson DA, Lombard C, Hargrove J et al. Accuracy and completeness of recording of confirmed tuberculosis in two South African communities. *Int J Tuberc Lung Dis*. 2011;15(3):337-343.
26. Fellegi IP, Sunter AB. A theory for record linkage. *JASA*. 1969;64:1183-1210.
27. Fischbacher CM, Bhopal R; Povey C, Steiner M, Chalmers J, Mueller G et al. Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians. *BMC Public Health*. 2007;7:142-152.
28. 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(7):1431-1438.
29. Fournel I, Schwarzinger 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;150:91-95.
30. FSEADE -Fundação Sistema Estadual de Análise de Dados. Dados para repensar a Aids no Estado de São Paulo: resultados da parceria entre Programa Estadual DST/Aids e Fundação Seade. São Paulo: DST/Aids, Fundação Seade; 2010.

31. FSEADE -Fundação Sistema Estadual de Análise de Dados. 2011 [acesso em 15 set 2011]. Disponível em: <http://www.seade.gov.br/master.php>.
32. FSEADE -Fundação Sistema Estadual de Análise de Dados. Nascimentos e óbitos. Integração das bases de dados do Ministério da Saúde e das bases de Dados do IBGE para aprimorar os indicadores epidemiológicos. Experiência do Espírito Santo: síntese do projeto. Espírito Santo; 2009.
33. Gomes Jr SCS, Almeida RT. Identificação de um caso novo de câncer no sistema de informação ambulatorial do Sistema Único de Saúde. *Cad Saúde Coletiva*. 2004;12(1):57-68.
34. Gonçalves FT, Francisco G, Souza SP de, Luiz OC, Festa-Neto C, Sanches JA et al. European ancestry and polymorphisms in DNA repair genes modify the risk of melanoma: a case-control study in a high UV index region in Brazil. *J Dermatol Sci*. 2011, doi:10.1016/j.jdermsci.2011.06.003
35. Herk-Sukel MPP van, Poll-Franse LV van de, Lemmens VEPP, Vreugdenhil G, Pruijt JFM, Coebergh JWW et al. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven cancer registry and the PHARMO record linkage system. *Eur J Cancer*. 2010;46(2):395-404.
36. Holman CDJ, Bass JA, Rosman DL, Smith MB, Semmens JB, Glasson EJ, Brook EL, Trutwein B, Rouse IL, Watson CR, Klerk NH, Stanley FJ. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aus Health Rev*. 2008;32(4):766-777.
37. Howe GR. Use of computerized record Linkage in cohort studies. *Epidemiol Rev*. 1998;20(1):112-121.

38. IARC-International Agency for Research on Cancer. Cancer incidence in five continents Volume- IX. IARC Scientific Publications No. 160. Lyon:IARCPress; 2007.
39. IBGE-Instituto Brasileiro de Geografia e Estatística. 2011 [acesso em 16 jun 2011]. Disponível em: <http://www.ibge.gov.br/cidadesat/default.php>.
40. IESS - Instituto de Estudos de Saúde Suplementar. Os custos do reassarcimento ao SUS. Saúde Suplementar em Foco. Informativo Eletrônico. 2010;1(8).
41. INCA-Instituto Nacional de Câncer. Câncer no Brasil: dados dos registros de base populacional- volume 4. Rio de Janeiro: INCA, 2010.
42. Jaro MA. Advances in Record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida. JASA. 1989;84:414-420.
43. Karmel R, Anderson P, Gibson D, Peut A, Duckett S, Wells Y. Empirical aspects of record linkage across multiple data sets using statistical linkage keys: the experience of the PIAC cohort study. BMC Health Serv Res. 2010;10:41.
44. Latorre MRDO. Câncer em Goiânia: análise da incidência e da mortalidade no período de 1988 a 1997. [Tese de livre docência]. São Paulo: Faculdade de Saúde Pública da USP; 2001.
45. Laudon KC; Laudon JP. Administração da empresa digital. In: _____ Trad. A S Marques. Sistemas de Informação gerenciais: administrando a empresa digital. São Paulo: Printice-Hall; 2004. cap 1 p. 5.

46. Laurenti R, Mello Jorge MHP, Gotlieb SLD. Mortalidade segundo causas: considerações sobre a fidedignidade dos dados. *Rev Panam Salud Publica*. 2008;23(5):349-356.
47. Laurenti R, Mello Jorge MHP, Lebrão ML, Gotlieb SLD. *Estatísticas de Saúde*. 2ed. São Paulo: EPU; 2005.
48. Li B, Quan H, Fong A, Lu M. Assessing record linkage between health care and vital statistics databases using deterministic methods. *BMC Health Serv Res*. 2006;6:48.
49. Machado CJ, Hill K. Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. *Cad Saúde Pública*. 2004;20(4):915-925.
50. Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad Saúde Pública*. 2004;20(2):362-371.
51. Machado JP, Silveira DP, Santos IS, Piovesan MF, Albuquerque C. Aplicação da metodologia de relacionamento probabilístico de base dados para a identificação de óbitos em estudos epidemiológicos. *Rev Bras Epidemiol*. 2008;11(1):43-54.
52. McGlynn EA, Damberg CL, Kerr EA, Brook RH. *Health information systems: design issues and analytic applications*. Rand Health; 1998.
53. Méray N, Reitsma JB, Ravelli ACJ, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol*. 2007;60:833-891.
54. Migowski A, Chaves RBM, Coeli CM, Ribeiro ALP, Tura BR, Kuschnir MCC et al. Accuracy of probabilistic record linkage in the assessment of

- high-complexity cardiology procedures. *Rev Saúde Pública*. 2011;45(2):269-275.
55. Mirra AP. Registro de Câncer no Brasil e sua história. Ministério da Saúde/ Instituto Nacional do Câncer/CONPREV -/Secretaria de Estado da Saúde/Fundação Oncocentro de São Paulo -/ Secretaria Municipal da Saúde/PRO-AIM -/Faculdade de Saúde Pública da Universidade de São Paulo / Departamento de Epidemiologia. São Paulo, BR; 2005.
56. Mirra AP, Latorre MRDO, Veneziano DB. Incidência de câncer no Município de São Paulo, Brasil: 1997-2003. Tendência no período 1969-2003. Mortalidade por câncer no Município de São Paulo, Brasil: 1997-2003. Ministério da Saúde/ Instituto Nacional do Câncer/CONPREV -/Secretaria de Estado da Saúde/Fundação Oncocentro de São Paulo -/ Secretaria Municipal da Saúde/PRO-AIM -/Faculdade de Saúde Pública da Universidade de São Paulo / Departamento de Epidemiologia. São Paulo, BR; 2007.
57. Moraes GH, Duarte EC. Análise da concordância dos dados de mortalidade por dengue em dois sistemas nacionais de informação em saúde, Brasil, 2000-2005. *Cad Saúde Pública*. 2009;25(11):2354-2364.
58. Morais LCC, Aranha V, Teixeira MLP, Waldvogel BC, Tayra A, Guibu IA et al. A construção e o tratamento da Base Integrada Paulista de Aids. In: XVII Encontro Nacional de Estudos Populacionais, ABEP; 2010; Caxambu, BR. Minas Gerais: Associação Brasileira de Estudos Populacionais; 2010.p. 1-21.
59. MS-Ministério da Saúde. Sistema de informação de agravos de notificação (SINAN). Brasília/DF; 2007.

60. MS-Ministério da Saúde. Sistema de informação hospitalar (SIH/SUS). Manual de orientações técnicas. Brasília/DF; 2006.
61. MS/SAS/DRAC/CGSI-Ministério da Saúde. Sistema de informações ambulatoriais do SUS (SIA/SUS): Manual de orientação técnicas. Brasília/DF; 2008.
62. Nakhaee F, McDonald A, Black D, Law M. A feasible method for linkage studies avoiding clerical review: linkage of the national HIV/AIDS surveillance databases with the national death index in Australia. *Aus N Z J Public Health*. 2007;31(4):308-312.
63. Nitsch D, Morton S, DeStavola BL, Clark H, Leon DA. How good is probabilistic record linkage to reconstruct reproductive histories? Results from the Aberdeen children of the 1950s study. *BMC Med Res Methodol*. 2006, 6:15.
64. Oberaigner W, Stuhlinger W. Record linkage in the Câncer Registry of Tyrol, Austria. *Methods Inf Med*. 2005;44:626-630.
65. Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC et al. Validation of the hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency vírus-infected persons and mortality databases in Brazil. *Am J Epidemiol*. 2008;168:1326-1332.
66. Pereira APE, Gama SGN, Leal MC. Mortalidade infantil em uma amostra de nascimentos do Município do Rio de Janeiro, 1999-2001; “*linkage*” com o Sistema de Informação de Mortalidade. *Rev Bras Saúde Matern Infant*. 2007;7:83-88.

67. Rafael RAA, Ribeiro VS, Cavalcante MCV, Santos AM, Simões VMF. Relacionamento probabilístico: recuperação de informações de óbitos infantis e natimortos em localidades no Maranhão, Brasil. *Cad Saúde de Pública*. 2011;7:1371-1379.
68. RCBP-SP - Registro de Câncer de Base Populacional de São Paulo. Câncer em São Paulo 1997-2008: incidência, mortalidade e tendência de câncer no Município de São Paulo. São Paulo, S.P.-Registro de Câncer de São Paulo; 2011.
69. Rezende DA. Planejamento de informações públicas municipais: sistemas de informação e de conhecimento, informática e governo eletrônico integrados aos planejamentos das prefeituras e Municípios. *RAP Rio de Janeiro*. 2007;41:505-536.
70. Silva CF, Leite AJM, Almeida NMGS. Linkage entre bancos de dados de nascidos vivos e óbitos infantis em município do Nordeste do Brasil: qualidade dos sistemas de informação. *Cad Saúde Pública*. 2009;25(7):1552-1558.
71. Silva ZP, Almeida MF, Ortiz LP, Alencar GP, Alencar AP Schoeps D et al. Morte neonatal precoce segundo complexidade hospitalar e rede SUS e não-SUS na Região Metropolitana de São Paulo, Brasil. *Cad Saúde Pública*. 2010;26(1):123-134.
72. Silveira DP, Artmann E. Accuracy of probabilistic record linkage applied to health databases: systematic review. *Rev Saúde Pública*. 2009;43(5):875-882.
73. Silveira MH. Considerações sobre o sistema de informação no setor saúde. *Rev Saúde Pública*. 1974;8:119-128.

74. Solomon DJ, Henry RC, Hogan JG, Amburg GHV, Taylor J. Evaluation and implementation of public health registries. *Public Health Rep.* 1991;105(2):142-150.
75. Sousa LMO, Pinheiro RS. Unnotified deaths and hospital admissions for tuberculosis in the municipality of Rio de Janeiro. *Rev Saúde Pública.* 2011;45(1):31-39.
76. Sousa MH, Cecatti JG, Hardy E, Serruya SJ. Relacionamento probabilístico de registros: uma aplicação na área de morbidade materna grave (near miss) e mortalidade materna. *Cad Saúde Pública.* 2008;24(3):653-662.
77. Stevens A. Pareamento de registros/SUS [CD - ROM]. São Paulo: Secretaria Municipal de Saúde; 2007.
78. Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D et al. Development and implementation of a high-quality clinical database: the Australian and New Zealand intensive care society adult patient database. *J Crit Care.* 2006;21(2):133-141.
79. Tancredi MV, Holcman MM, Teixeira Jr AE, Farias NSO. Análise da sobrevivência de pacientes com Aids no Estado de São Paulo. In: *Dados para repensar a Aids no Estado de São Paulo: resultados da parceria entre Programa Estadual DST/Aids e Fundação Seade.* São Paulo: DST/Aids, Fundação Seade; 2010:83-110.
80. Taniguchi MT, Pelaquin MHH, Latorre MRDO. Relacionamento probabilístico entre as bases de dados do registro de câncer de São Paulo e do sistema de informações de mortalidade municipal. [Trabalho de conclusão de

- curso de Especialização em Epidemiologia para Serviços de Saúde]. São Paulo: Faculdade de Saúde Pública da USP; 2006.
81. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol*. 2011;64(5):565-572.
 82. Vidal EIO, Coeli CM, Pinheiro RS, Camargo-Jr KR. Mortality within 1 year after hip fracture surgical repair in the elderly according to postoperative period: a probabilistic record linkage study in Brazil. *Osteoporos Int*. 2006;17:1569-1576.
 83. Wagner G. History of cancer registration. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. *Cancer registration principles and methods*. International agency for research on cancer (WHO)/International association of cancer registries. Lyon: IARC Scientific Publications; 1991:3-7.
 84. Waldvogel BC, Ferreira CEC, Camargo ABM, Jordani MS, Ortiz LP. Base unificada de nascimentos e óbitos no Estado de São Paulo: instrumento para aprimorar os indicadores de saúde. *São Paulo Perspect*. 2008;22(1):5-18.
 85. Waldvogel BC. *Acidentes do trabalho: os casos fatais a questão da identificação e da mensuração*. Belo Horizonte: Segrac; 2002.
 86. Waldvogel BC, Moraes LCC, Guibu IA. Base integrada Paulista de Aids: recuperação, gerência e vinculação de dados para melhoria da qualidade da informação. In: *Dados para repensar a Aids no Estado de São Paulo: resultados da parceria entre Programa Estadual DST/Aids e Fundação Seade*. São Paulo: DST/Aids, Fundação Seade; 2010:9-24.

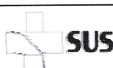
87. White KL. Information for health care: an epidemiological perspective. *Inquiry-The Journal of Health Care Organization, providing and financing.* 1980;17:296-312.
88. WHO-World Health Organization. PAHO-Pan American Health Organization. Collection and use of core health data. Report. Washington; 1997.
89. WOSCOPS-The West of Scotland Coronary Prevention Study Group. Computerised record linkage: compared with traditional patient follow-up methods in clinical trials and illustrated in a prospective epidemiological study. *J Clin Epidemiol.* 1995;48:1441-1452.
90. Wotton CJ, Yeates DGR, Goldacre MJ. Cancer in patients admitted to hospital with diabetes mellitus aged 30 years and over: record linkage studies. *Diabetologia.* 2011;54(3):527-534.

Anexos

Anexo 1 - Declaração de Óbito (DO)

República Federativa do Brasil Ministério da Saúde 1ª VIA - SECRETARIA DE SAUDE		Declaração de Óbito				
I	Cartório	1 Cartório	Código	2 Registro	3 Data	
		4 Município	5 UF	6 Cemitério		
II	Identificação	7 Tipo de Óbito <input type="checkbox"/> 1 - Fetal <input type="checkbox"/> 2 - Não Fetal	8 Óbito Data _____ Hora _____	9 RIC	10 Naturalidade	
		11 Nome do falecido				
		12 Nome do pai	13 Nome da mãe			
III	Residência	14 Data de nascimento	15 Idade Anos completos _____ Meses _____ Dias _____ Horas _____ Minutos _____ Ignorado <input type="checkbox"/>	16 Sexo <input type="checkbox"/> M - Masc. <input type="checkbox"/> F - Fem. <input type="checkbox"/> 1 - Ignorado	17 Raça/cor <input type="checkbox"/> 1 - Branca <input type="checkbox"/> 2 - Preta <input type="checkbox"/> 3 - Amarela <input type="checkbox"/> 4 - Parda <input type="checkbox"/> 5 - Indígena	
		18 Estado Civil <input type="checkbox"/> 1 - Solteiro <input type="checkbox"/> 2 - Casado <input type="checkbox"/> 3 - Viúvo <input type="checkbox"/> 4 - Separado Judicialmente <input type="checkbox"/> 5 - União consensual <input type="checkbox"/> 9 - Ignorado	19 Escolaridade (Em anos de estudo concluídos) <input type="checkbox"/> 1 - Nenhuma <input type="checkbox"/> 2 - De 1 a 3 <input type="checkbox"/> 3 - De 4 a 7 <input type="checkbox"/> 4 - De 8 a 11 <input type="checkbox"/> 5 - 12 e mais <input type="checkbox"/> 9 - Ignorado		20 Ocupação habitual e ramo de atividade (Se aposentado, colocar a ocupação habitual anterior) _____ Código _____	
		21 Logradouro (Rua, praça, avenida etc.)	Código	Número	Complemento	22 CEP
IV	Ocorrência	23 Bairro/Distrito	Código	24 Município de Residência	Código	
		25 UF				
		26 Local de ocorrência do óbito <input type="checkbox"/> 1 - Hospital <input type="checkbox"/> 2 - Outros estab. Saúde <input type="checkbox"/> 3 - Domicílio <input type="checkbox"/> 4 - Via pública <input type="checkbox"/> 5 - Outros <input type="checkbox"/> 9 - Ignorado	27 Estabelecimento Código _____			
V	Fetal ou menor que 1 ano	28 Endereço da ocorrência, se fora do estabelecimento ou da residência (Rua, praça, avenida, etc.)	Número	Complemento	29 CEP	
		30 Bairro/Distrito	Código	31 Município de ocorrência	Código	
		32 UF				
VI	Condições e causas do óbito	PREENCHIMENTO EXCLUSIVO PARA ÓBITOS FETAIS E DE MENORES DE 1 ANO				
		INFORMAÇÕES SOBRE A MÃE				
		33 Idade Anos _____	34 Escolaridade (Em anos de estudo concluídos) <input type="checkbox"/> 1 - Nenhuma <input type="checkbox"/> 2 - De 1 a 3 <input type="checkbox"/> 3 - De 4 a 7 <input type="checkbox"/> 4 - De 8 a 11 <input type="checkbox"/> 5 - 12 e mais <input type="checkbox"/> 9 - Ignorado	35 Ocupação habitual e ramo de atividade da mãe Código _____	36 Número de filhos tidos Obs. Utilizar 99 para ignorados) Nascidos vivos _____ Nascidos mortos _____	
VII	Médico	37 Duração da gestação (Em semanas) <input type="checkbox"/> 1 - Menos de 22 <input type="checkbox"/> 2 - De 22 a 27 <input type="checkbox"/> 3 - De 28 a 31 <input type="checkbox"/> 4 - De 32 a 36 <input type="checkbox"/> 5 - De 37 a 41 <input type="checkbox"/> 6 - 42 e mais <input type="checkbox"/> 9 - Ignorado	38 Tipo de Gravidez <input type="checkbox"/> 1 - Única <input type="checkbox"/> 2 - Dupla <input type="checkbox"/> 3 - Tripla e mais <input type="checkbox"/> 9 - Ignorada	39 Tipo de Parto <input type="checkbox"/> 1 - Vaginal <input type="checkbox"/> 2 - Cesáreo <input type="checkbox"/> 9 - Ignorado	40 Morte em relação ao parto <input type="checkbox"/> 1 - Antes <input type="checkbox"/> 2 - Durante <input type="checkbox"/> 3 - Depois <input type="checkbox"/> 9 - Ignorado	
		41 Peso ao nascer Gramas _____	42 Num. da Declar. de Nascidos Vivos			
		OBITOS EM MULHERES 43 A morte ocorreu durante a gravidez, parto ou aborto? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado		44 A morte ocorreu durante o puerpério? <input type="checkbox"/> 1 - Sim até 42 dias <input type="checkbox"/> 2 - Sim de 43 dias a 1 ano <input type="checkbox"/> 3 - Não <input type="checkbox"/> 9 - Ignorado		ASSISTÊNCIA MÉDICA 45 Recebeu assist. médica durante a doença que ocasionou a morte? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado
VIII	Causas externas	DIAGNÓSTICO CONFIRMADO POR: 46 Exame complementar? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado				
		47 Cirurgia? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado				
		48 Necropsia? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado				
IX	Localidade Médico	49 CAUSAS DA MORTE PARTE I ANOTE SOMENTE UM DIAGNÓSTICO POR LINHA				
		Doença ou estado mórbido que causou diretamente a morte				
		CAUSAS ANTECEDENTES Estados mórbidos, se existirem, que produziram a causa acima registrada, mencionando-se em último lugar a causa básica.				
50 Nome do médico		51 CRM	52 O médico que assina atendeu ao falecido? <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Substituto <input type="checkbox"/> 3 - IML <input type="checkbox"/> 4 - SVO <input type="checkbox"/> 5 - Outros			
53 Meio de contato (Telefone, fax, e-mail etc.)		54 Data do atestado	55 Assinatura			
56 Tipo <input type="checkbox"/> 1 - Acidente <input type="checkbox"/> 2 - Suicídio <input type="checkbox"/> 3 - Homicídio <input type="checkbox"/> 4 - Outros <input type="checkbox"/> 9 - Ignorado		57 Acidente do trabalho <input type="checkbox"/> 1 - Sim <input type="checkbox"/> 2 - Não <input type="checkbox"/> 9 - Ignorado		58 Fonte de informação <input type="checkbox"/> 1 - Boletim de Ocorrência <input type="checkbox"/> 2 - Hospital <input type="checkbox"/> 3 - Família <input type="checkbox"/> 4 - Outra <input type="checkbox"/> 9 - Ignorado		
59 Descrição sumária do evento, incluindo o tipo de local de ocorrência						
SE A OCORRÊNCIA FOR EM VIA PÚBLICA, ANOTAR O ENDEREÇO						
60 Logradouro (Rua, praça, avenida etc.)		Código				
61 Declarante		62 Testemunhas				

Anexo 2 - Ficha de Solicitação da APAC-SIA/SUS

 Sistema Único de Saúde		Ministério da Saúde		LAUDO PARA SOLICITAÇÃO/AUTORIZAÇÃO DE PROCEDIMENTO AMBULATORIAL		<i>fls. 1/2</i>		
IDENTIFICAÇÃO DO ESTABELECIMENTO DE SAÚDE (SOLICITANTE)								
1 - NOME DO ESTABELECIMENTO DE SAÚDE SOLICITANTE						2 - CNES		
IDENTIFICAÇÃO DO PACIENTE								
3 - NOME DO PACIENTE						4 - Nº DO PRONTUÁRIO		
5 - CARTÃO NACIONAL DE SAÚDE (CNS)				6 - DATA DE NASCIMENTO		7 - SEXO		
						Masc. <input type="checkbox"/> Fem. <input type="checkbox"/>		
8 - NOME DA MÃE OU RESPONSÁVEL						9 - TELEFONE DE CONTATO Nº DO TELEFONE		
						DDD		
10 - ENDEREÇO (RUA, Nº, BAIRRO)								
11 - MUNICÍPIO DE RESIDÊNCIA						12 - Cód. IBGE MUNICÍPIO	13 - UF	14 - CEP
PROCEDIMENTO SOLICITADO								
15 - CÓDIGO DO PROCEDIMENTO PRINCIPAL				16 - NOME DO PROCEDIMENTO PRINCIPAL			17 - QTDE	
PROCEDIMENTO(S) SECUNDÁRIO(S)								
18 - CÓDIGO DO PROCEDIMENTO SECUNDÁRIO		19 - NOME DO PROCEDIMENTO SECUNDÁRIO					20 - QTDE	
21 - CÓDIGO DO PROCEDIMENTO SECUNDÁRIO		22 - NOME DO PROCEDIMENTO SECUNDÁRIO					23 - QTDE	
24 - CÓDIGO DO PROCEDIMENTO SECUNDÁRIO		25 - NOME DO PROCEDIMENTO SECUNDÁRIO					26 - QTDE	
27 - CÓDIGO DO PROCEDIMENTO SECUNDÁRIO		28 - NOME DO PROCEDIMENTO SECUNDÁRIO					29 - QTDE	
30 - CÓDIGO DO PROCEDIMENTO SECUNDÁRIO		31 - NOME DO PROCEDIMENTO SECUNDÁRIO					32 - QTDE	
JUSTIFICATIVA DO(S) PROCEDIMENTO(S) SOLICITADO(S)								
33 - DESCRIÇÃO DO DIAGNÓSTICO				34 - CID10 PRINCIPAL, 35 - CID10 SECUNDÁRIO, 36 - CID10 CAUSAS ASSOCIADAS				
37 - OBSERVAÇÕES								
SOLICITAÇÃO								
38 - NOME DO PROFISSIONAL SOLICITANTE				39 - DATA DA SOLICITAÇÃO		42 - ASSINATURA E CARIMBO (Nº REGISTRO DO CONSELHO)		
40 - DOCUMENTO		41 - Nº DOCUMENTO (CNS/CPF) DO PROFISSIONAL SOLICITANTE						
() CNS () CPF								
AUTORIZAÇÃO								
43 - NOME DO PROFISSIONAL AUTORIZADOR				44 - Cód. ÓRGÃO EMISSOR		49 - Nº DA AUTORIZAÇÃO (APAC)		
45 - DOCUMENTO		46 - Nº DOCUMENTO (CNS/CPF) DO PROFISSIONAL AUTORIZADOR						
() CNS () CPF								
47 - DATA DA AUTORIZAÇÃO		48 - ASSINATURA E CARIMBO (Nº DO REGISTRO DO CONSELHO)				50 - PERÍODO DE VALIDADE DA APAC		
						a		
IDENTIFICAÇÃO DO ESTABELECIMENTO DE SAÚDE (EXECUTANTE)								
51 - NOME FANTASIA DO ESTABELECIMENTO DE SAÚDE EXECUTANTE						52 - CNES		

Anexo 3 - Ficha de Notificação do RCBP-SP

Ministério da Saúde / Instituto Nacional de Câncer / CONPREV Secretaria de Estado da Saúde de São Paulo / Fundação Oncocentro de São Paulo Secretaria Municipal de Saúde / PROAIM Faculdade de Saúde Pública da Universidade de São Paulo / Departamento de Epidemiologia			
Registro de Câncer de Base Populacional de São Paulo			
			Nº do registro hospital
			<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Nome completo do paciente _____			

Nome da mãe _____			

Data Nascimento	____/____/____	Idade	____
Sexo	<input type="text"/> <input type="text"/>	Masc.	<input type="text"/> <input type="text"/>
		Fem.	<input type="text"/> <input type="text"/>
Cor			
<input type="text"/> <input type="text"/>	Branca	<input type="text"/> <input type="text"/>	Negra
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Parda
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Amarela
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Ignorado
Estado Civil			
<input type="text"/> <input type="text"/>	Solt.	<input type="text"/> <input type="text"/>	Cas.
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Viúva
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Div.
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	Ignorado
Nacionalidade	_____		<input type="text"/> <input type="text"/> <input type="text"/>
Naturalidade	_____		<input type="text"/> <input type="text"/> <input type="text"/>
Residência	_____		<input type="text"/> <input type="text"/> <input type="text"/>
Profissão	_____		<input type="text"/> <input type="text"/> <input type="text"/>
Fonte de notificação	_____		<input type="text"/> <input type="text"/> <input type="text"/>
Topografia (localização) _____			<input type="text"/> <input type="text"/> <input type="text"/>
Morfologia (tipo histológico) _____			<input type="text"/> <input type="text"/> <input type="text"/>
Meio de Diagnóstico		Estadiamento	
<input type="text"/> <input type="text"/>	Histológico	<input type="text"/> <input type="text"/>	Clínico
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	
<input type="text"/> <input type="text"/>	Citológico	<input type="text"/> <input type="text"/>	N
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	
<input type="text"/> <input type="text"/>	Cirúrgico	<input type="text"/> <input type="text"/>	M
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	
<input type="text"/> <input type="text"/>	Raio X	<input type="text"/> <input type="text"/>	FIGO
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	
<input type="text"/> <input type="text"/>	Ignorado	<input type="text"/> <input type="text"/>	
Informação			
<input type="text"/> <input type="text"/>	Notificação	<input type="text"/> <input type="text"/>	Atestado de Óbito
<input type="text"/> <input type="text"/>		<input type="text"/> <input type="text"/>	
Data do Primeiro Diagnóstico	Data Última Informação	Data do Óbito	Coletador
____/____/____	____/____/____	____/____/____	_____

Anexo 4 - Tabela das 24 sub-bases de dados

Quadro das sub-bases utilizadas no processo de *linkage*. RCBP-SP, 1997 a 2005.

Sub-bases de dados	Código CID-O 3ª edição	Descrição	Tamanho (n)
1	C00.0 a C09.9	Partes do lábio, base da língua, outras partes e partes não especificadas da língua, gengiva, assoalho da boca, palato, outras partes e partes não especificada da bica, glândula parótida, outras glândulas salivares maiores e salivares maiores não especificadas e amígdala	9.321
2	C10.0 a 11.9	Orofaringe e nasofaringe	1.613
3	C12.0 a 14.8	Seio piriforme, hipofaringe e outras localizações e localizações mal definidas, do lábio, cavidade oral e faringe	1.381
4	C15.0 a C15.9	Esôfago	5.579
5	C16.0 a C17.9	Estômago e Intestino Delgado	18.511
6	C18.0 a C18.9	Cólon	15.684
7	C19.0 a C21.8	Junção retossigmóide, reto, ânus e canal anal	10.615
8	C22.0 a C24.9	Fígado e via biliares intra-hepáticas, outras partes e partes não especificadas das vias biliares	4.021
9	C25.0 a C26.9	Pâncreas e outros órgãos digestivos e localizações mal definidas do aparelho digestivo	5.896
10	C30.0 a C33.9	Cavidade nasal e ouvido médio, seio da face, laringe e traquéia	6.260
11	C34.0 a C34.9	Brônquios e pulmões	17.455
12	C37.0 a C39.9	Timo, coração, mediastino e pleura, outras localizações e localizações mal definidas do aparelho respiratório e dos órgãos intratorácicos	854
13	C40.0 a C42.4	Ossos, articulações e cartilagens articulares dos membros e de outras localizações e de localizações não especificadas, sistema hematopoiético e reticuloendotelial	11.358
14	C44.0 a C44.9 (8720 a 8780)	Pele melanoma	5.034

Continua

Sub-bases de dados	Código CID-O 3ª edição	Descrição	Tamanho (n)
15	C44.0 a C44.9	Pele não melanoma	70.662
16	C47.0 a C49.9	Nervos periféricos e sistema nervoso, retroperitônio e peritônio, tecido conjuntivo, subcutâneo e outros tecidos moles	2.830
17	C50.0 a C50.9	Mama	41.307
18	C51.0 a C58.9	Vulva, vagina, colo do útero, corpo do útero, útero, ovário, outros órgãos genitais femininos e os não especificados, e placenta	28.359
19	C60.0 a C60.9 e C62.0 a C63.9	Pênis, testículo e outros órgãos genitais masculinos e os não especificados	2.153
20	C61.9	Próstata	29.455
21	C64.0 a C68.9	Rim, pelve renal, ureter, bexiga, outros órgãos urinários	12.343
22	C69.0 a C72.9	Olho e anexos, meninges, encéfalo, medula espinhal, nervos cranianos e outras partes do sistema nervoso central	7.199
23	C73.0 a C75.9	Glândula tireóide e supra-renal, outras glândulas endócrinas e estruturas relacionadas	11.696
24	C76.0 a C80.9	Outras localizações e localizações mal definidas, linfonodos e localização primária desconhecida	23.720

Continuação

Anexo 5 - Termo de Cooperação Técnica entre a FSP/USP e a FSEADE

Termo de Cooperação Técnica que entre si celebram a Fundação Sistema Estadual de Análise de Dados – SEADE e a Faculdade de Saúde Pública da Universidade de São Paulo por meio de seu Departamento de Epidemiologia, objetivando promover pesquisas e estudos no campo das análises demográficas e epidemiológicas.

Por este instrumento particular, de um lado, a FUNDAÇÃO SISTEMA ESTADUAL DE ANÁLISE DE DADOS – SEADE, autorizadas pela Lei nº 1.866/78 e Decreto nº 13.161/79, que aprovou seus estatutos, com sede na Avenida Cásper Líbero, nº 464, Luz, na Cidade de São Paulo, Estado de São Paulo, inscrita no Cadastro Nacional da Pessoa Jurídica – CNPJ sob o nº 51.169.555/0001-00, inscrita na Fazenda Estadual sob o nº 110.169.886.110, neste ato representada por sua Diretora Executiva, Dra. Felícia Reicher Madeira, doravante denominada SEADE e, de outro lado, a UNIVERSIDADE DE SÃO PAULO, sediada na Rua da Reitoria, nº 109 na Cidade de São Paulo, Estado de São Paulo, CEP 05509-900, inscrita no CNPJ sob nº 63.025.530/0001-04, neste ato representada por seu Reitor João Grandino Rodas, RG 3304770-4, CPF nº 243200218-00, com interveniência da FACULDADE DE SAÚDE PÚBLICA DA UNIVERSIDADE DE SÃO PAULO – FSP/USP, sediada na Avenida Doutor Arnaldo nº 715, Cerqueira César, na Cidade de São Paulo, Estado de São Paulo, CEP 01246-904 representada pela sua diretora Helena Ribeiro, portadora do RG nº 3572554 e inscrita no CPF/MF nº 028402688-33, doravante denominada INTERVENIENTE, resolvem, de comum acordo celebrar o presente Termo de Cooperação Técnica, que se regerá pelas cláusulas e condições a seguir descritas:

CLÁUSULA PRIMEIRA – DO OBJETO

Constitui objeto da presente cooperação técnica, a produção de estatísticas complementares sobre os óbitos por câncer, no Estado de São Paulo, que serão atualizadas anualmente a partir da vinculação, entre a base de dados relativa aos pacientes que recebem tratamento para as citadas doenças, do INTERVENIENTE, e a base de óbitos produzida pelo SEADE.

CLÁUSULA SEGUNDA – DAS CONDIÇÕES DE EXECUÇÃO



09

Para consecução do objeto desta cooperação caberá aos partícipes executar, acompanhar e avaliar as atividades de pesquisa a serem desenvolvidas, elaborando a documentação técnica a ela relacionada.

CLÁUSULA TERCEIRA – DOS COMPROMISSOS DOS PARTÍCIPES

1 – Compete especificamente ao INTERVENIENTE:

- a) Disponibilizar anualmente a base atualizada de dados de pacientes de seu registro hospitalar contendo, principalmente, as seguintes variáveis: nome do paciente, nome da mãe, data de nascimento e município de residência;
- b) Realizar reuniões técnicas para analisar as bases de dados resultantes do presente Convênio;
- c) Nos casos de divulgação de estudos e pesquisas, explicitar a autoria citando o presente termo de cooperação e mencionar a Fundação Seade como fonte de dados;
- d) Encaminhar, à Fundação Seade, cópia de publicações e demais produtos resultantes deste convênio;
- e) Responder pelos custos decorrentes do desenvolvimento do objeto deste termo, tais como envolvimento os deslocamento de seus técnicos e pesquisadores.

2 – Compete especificamente ao SEADE:

- a) Realizar anualmente tratamento e aplicação do processo de vinculação das bases de dados do SEADE e do INTERVENIENTE, para atingir os objetivos desta cooperação;
- b) Elaborar anualmente a base de óbitos com os pacientes identificados no processo de vinculação realizada no item anterior, contendo informações sobre as causas de morte;
- c) Realizar reuniões técnicas para analisar as bases de dados resultantes do presente Convênio;
- d) Responder pelos custos decorrentes do desenvolvimento do objeto deste termo, tais como envolvimento os deslocamento de seus técnicos e pesquisadores.

CLÁUSULA QUARTA – DOS RECURSOS FINANCEIROS

A presente cooperação técnica não implica em qualquer ônus, encargos ou repasse de recursos entre os partícipes, sendo que as despesas de cada um onerarão as dotações próprias.



Handwritten signatures and initials, including a large signature on the left and initials 'JCM' and 'VPR' on the right.



108

Parágrafo único: Caso haja a necessidade de recursos financeiros na execução do objeto desta Cooperação Técnica, os mesmos serão definidos em termos aditivos próprios.

CLÁUSULA QUINTA – DO ACOMPANHAMENTO E DA EXECUÇÃO

Os partícipes definirão os coordenadores que serão responsáveis pela comunicação entre as instituições, dando apoio nos procedimentos técnico-administrativos diretamente relacionados à execução do presente instrumento.

CLÁUSULA SEXTA – DA CONFIDENCIALIDADE

A qualquer tempo, as partes manterão estrito sigilo acerca dos documentos e informações de caráter confidencial, colocados à disposição destas, em decorrência deste termo.

CLÁUSULA SÉTIMA – DO PRAZO DE VIGÊNCIA

O presente ajuste terá duração de 24 (vinte e quatro) meses, a partir de sua assinatura, podendo ser prorrogado mediante lavratura de termo aditivo.

CLÁUSULA OITAVA – DA RESCISÃO OU DENÚNCIA

O presente ajuste será rescindido por infração legal, descumprimento de quaisquer de suas cláusulas ou divergências que não forem dirimidas administrativamente, e poderá ser denunciado, a qualquer tempo e por qualquer dos partícipes, mediante comunicação expressa, com antecedência mínima de 60 (sessenta) dias.

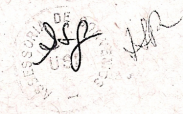
Parágrafo único: Havendo pendências, os partícipes definirão, mediante Termo de Encerramento da Cooperação Técnica, as responsabilidades pela conclusão ou encerramento de cada um dos trabalhos e todas as demais pendências, respeitando as atividades em curso.

CLÁUSULA NONA – DAS DISPOSIÇÕES GERAIS

1 – As dúvidas que eventualmente surgirem, assim como as divergências e os casos omissos, serão dirimidos por via de entendimento entre os partícipes, ouvidos os órgãos envolvidos e levadas a termo.



Handwritten signature: JRY



2 – Os termos do presente instrumento poderão ser alterados em qualquer de suas cláusulas e disposições, mediante concordância expressa e por escrito, através de termo aditivo, salvo sue objeto.

3 – Uma vez cessada a vigência deste instrumento, cessarão automaticamente os aditamentos e contratos firmados com terceiros em razão desta Cooperação Técnica, celebrado entre as partes.

CLÁUSULA DÉCIMA – DO FORO

Para dirimir dúvidas que possam ser suscitadas na execução e interpretação do presente ajuste, fica eleito o foro da Capital do Estado de São Paulo, com exclusão de qualquer outro, mesmo privilegiado.

E, por estarem assim justos e convencionados, os partícipes assinam o presente instrumento em 03 (três) vias de igual teor e forma, para um só efeito, na presença de 02 (duas) testemunhas.

São Paulo, 23 de fevereiro de 20 11.

Felícia Reicher Madeira

FUNDAÇÃO SISTEMA ESTADUAL DE ANÁLISE DE DADOS – SEADE

Dra. Felícia Reicher Madeira

Diretora Executiva

João Grandino Rodas

UNIVERSIDADE DE SÃO PAULO – USP

Prof. Dr. João Grandino Rodas.

Reitor

Helena Ribeiro

FACULDADE DE SAÚDE PÚBLICA DA UNIVERSIDADE DE SÃO PAULO – FSP/USP

Profa. Dra. Helena Ribeiro

Diretora

TESTEMUNHAS:

1. _____

Nome:

RG:

2. _____

Nome:

RG:



Anexo 6 - Rotinas

Descrição dos procedimentos para a formatação da base de dados do RCBP-SP.

Proposta	Expressão Numérica (Syntax)
Gerar a idade a partir da data do diagnóstico e data de nascimento	COMPUTE idade=CTIME.DAYS(dtdiag – dtnasc) / 365.25. EXECUTE.
Extrair o ano de nascimento a partir da data de nascimento	COMPUTE anonasc=XDATE.YEAR(dtnasc). EXECUTE.
Extrair o ano do diagnóstico a partir da data do diagnóstico	COMPUTE anodiag=XDATE.YEAR(dtdiag). EXECUTE.
Gerar o ano de nascimento a partir da Idade e Ano do diagnóstico	COMPUTE anonasc=ano do diagnóstico - idade. EXECUTE.
Transformar a variável Sexo numérica em alfanumérica	STRING sexo1 (A8). COMPUTE sexo1=STRING(sexo,f1). EXECUTE.
Transformar a variável Ano de nascimento numérica em alfanumérica	STRING anonasc1 (A8). COMPUTE anonasc1=STRING(anonasc,f4). EXECUTE.
Extrair o dia, mês e o ano de nascimento da variável Data de Nascimento para transformar esta variável em alfanumérica	COMPUTE dianasc=XDATE.MDAY(dtnasc). EXECUTE.
	COMPUTE mesnasc=XDATE.MONTH(dtnasc). EXECUTE.
	COMPUTE anonasc3=XDATE.YEAR(dtnasc). EXECUTE.
Após a variável Data de nascimento ser desmembrada foi realizado o agrupamento para finalizar o processo de transformação em caractere alfanumérico	COMPUTE diames=CONCAT(RTRIM(dianasc),RTRIM(mesnasc)). EXECUTE.
Data de nascimento (as variáveis devem estar em string)	COMPUTE dtcompl=CONCAT(RTRIM(diames),RTRIM(anonasc3)). EXECUTE.

Descrição das funções aplicadas na base de dados do PRO-AIM e Apac-SIA/SUS.

Proposta	Bases	Expressão Numérica (Syntax)
Unir as bases de dados – as variáveis de cada banco deverão estar com as mesmas configurações	PRO-AIM APAC	DATASET ACTIVATE DataSet1. ADD FILES /FILE=* /FILE='C:\linkage\obito1998.sav'. EXECUTE.
Eliminar os óbitos fetais – os óbitos não fetais estavam classificados como 1 na base de dados	PRO-AIM	USE ALL. COMPUTE filter_\$=(obito = "1"). VARIABLE LABEL filter_\$ 'obito = "1" (FILTER)'. VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'. FORMAT filter_\$ (f1.0). FILTER BY filter_\$. EXECUTE. FILTER OFF. USE ALL. SELECT IF (obito = "1"). EXECUTE.
Retirar os espaços excedentes no início do nome e entre o nome e sobrenomes	PRO-AIM APAC	STRING nome1 (A160). COMPUTE nome1=LTRIM(nome). EXECUTE.
Organizar a variável em ascendente ou descendente para identificar os espaços vazios	PRO-AIM APAC	SORT CASES BY id(A) SORT CASES BY id(D)
Recodificação da variável sexo	PRO-AIM APAC	STRING SEXO1 (A8). RECODE SEXO ('M'='1') ('F'='2') INTO SEXO1. EXECUTE.

Anexo 7 - Quadros dos *linkages* entre o RCBP-SP versus PRO-AIM e Apac-SIA/SUS

RCBP-SP versus PRO-AIM

Topografia									
C00.0 a C09.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 9..321	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:50	0:06:50	0:14:40	11219	1654	2447	14.74	21.81	3
P+S/N+DT	0:10:38	0:06:49	0:17:27	7305	243	377	3.33	5.16	2
U+S/N+DT	0:10:36	0:06:48	0:17:24	6150	336	566	5.46	9.20	7
Fp+Fu+S/N+DT	0:07:24	0:06:49	0:14:13	9609	1770	2031	18.42	21.14	15
P+U+S+A/N+DT	0:07:16	0:06:44	0:14:00	3747	2212	52	59.03	1.39	0
P+S+A/N+DT	0:10:26	0:06:31	0:16:57	433	141	13	32.56	3.00	0
U+S+A/N+DT	0:10:22	0:06:33	0:16:55	388	168	28	43.30	7.22	0
Fp+Fu+S+A/N+DT	0:07:23	0:06:44	0:14:07	3747	2212	59	59.03	1.57	0
Fp+S+A/N+DT	0:10:00	0:06:30	0:16:30	494	217	30	43.93	6.07	0
Fu+S+A/N+DT	0:09:54	0:06:30	0:16:24	448	227	39	50.67	8.71	0
P+U+S/N+A	0:07:56	0:06:48	0:14:44	8411	1842	2418	21.90	28.75	22
P+S/N+A	0:11:20	0:06:32	0:17:52	787	261	379	33.16	48.16	1
U+S/N+A	0:11:15	0:06:33	0:17:48	912	359	560	39.36	61.40	2
Fp+Fu+S/N+A	0:07:28	0:06:47	0:14:15	7534	1972	2152	26.17	28.56	37

Topografia									
C10.0 a C11.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 1.613	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:06:12	0:06:17	0:12:29	5056	546	813	10.80	16.08	16
P+S/N+DT	0:09:26	0:06:21	0:15:47	8007	175	261	2.19	3.26	0
U+S/N+DT	0:09:21	0:06:21	0:15:42	7522	207	320	2.75	4.25	0
Fp+Fu+S/N+DT	0:06:13	0:06:20	0:12:33	4317	551	660	12.76	15.29	14
P+U+S+A/N+DT	0:05:44	0:06:04	0:11:48	1084	583	31	53.78	2.86	0
P+S+A/N+DT	0:08:02	0:06:04	0:14:06	261	148	29	56.70	11.11	0
U+S+A/N+DT	0:08:13	0:06:04	0:14:17	273	99	32	36.26	11.72	0
Fp+Fu+S+A/N+DT	0:05:54	0:06:04	0:11:58	957	583	31	60.92	3.24	0
Fp+S+A/N+DT	0:07:35	0:06:03	0:13:38	271	168	26	61.99	9.59	0
Fu+S+A/N+DT	0:07:47	0:06:04	0:13:51	225	103	33	45.78	14.67	0
P+U+S/N+A	0:06:10	0:06:16	0:12:26	2936	590	792	20.10	26.98	16
P+S/N+A	0:09:22	0:06:03	0:15:25	621	189	251	30.43	40.42	1
U+S/N+A	0:09:16	0:06:04	0:15:20	788	213	309	27.03	39.21	0
Fp+Fu+S/N+A	0:06:16	0:06:14	0:12:30	2608	596	654	22.85	25.08	13

Topografia									
C12.0 a C14.8	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 1.381	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:16:18	0:20:58	0:37:16	4906	460	644	9.38	13.13	9
P+S/N+DT	0:17:37	0:21:30	0:39:07	10263	156	241	1.52	2.35	3
U+S/N+DT	0:23:30	0:20:49	0:44:19	7823	192	298	2.45	3.81	11
Fp+Fu+S/N+DT	0:16:14	0:20:25	0:36:39	4319	470	548	10.88	12.69	0
P+U+S+A/N+DT	0:15:05	0:20:37	0:35:42	797	489	43	61.36	5.40	2
P+S+A/N+DT	0:14:02	0:20:38	0:34:40	3120	550	29	17.63	0.93	1
U+S+A/N+DT	0:20:16	0:19:40	0:39:56	275	153	23	55.64	8.36	0
Fp+Fu+S+A/N+DT	0:15:54	0:19:38	0:35:32	713	490	19	68.72	2.66	0
Fp+S+A/N+DT	0:19:09	0:19:30	0:38:39	207	73	27	35.27	13.04	0
Fu+S+A/N+DT	0:19:15	0:19:28	0:38:43	287	166	17	57.84	5.92	0
P+U+S/N+A	0:15:50	0:20:18	0:36:08	3355	502	639	14.96	19.05	23
P+S/N+A	0:11:04	0:06:53	0:17:57	627	164	239	26.16	38.12	8
U+S/N+A	0:11:07	0:06:52	0:17:59	567	198	286	34.92	50.44	0
Fp+Fu+S/N+A	0:16:17	0:20:12	0:36:29	3026	513	566	16.95	18.70	29

Topografia									
C15.0 a C15.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.579	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:01	0:06:22	0:13:23	11568	2000	2687	17.29	23.23	10
P+S/N+DT	0:10:19	0:06:24	0:16:43	10759	465	633	4.32	5.88	3
U+S/N+DT	0:10:05	0:06:23	0:16:28	10597	517	775	4.88	7.31	1
Fp+Fu+S/N+DT	0:06:43	0:06:23	0:13:06	10011	2103	2306	21.01	23.03	1
P+U+S+A/N+DT	0:06:25	0:06:17	0:12:42	3061	1650	63	53.90	2.06	0
P+S+A/N+DT	0:09:17	0:05:57	0:15:14	389	182	39	46.79	10.03	0
U+S+A/N+DT	0:09:20	0:06:06	0:15:26	550	229	41	41.64	7.45	0
Fp+Fu+S+A/N+DT	0:06:19	0:06:15	0:12:34	2841	1733	58	61.00	2.04	0
Fp+S+A/N+DT	0:08:33	0:06:03	0:14:36	381	220	28	57.74	7.35	0
Fu+S+A/N+DT	0:08:40	0:06:05	0:14:45	552	293	44	53.08	7.97	0
P+U+S/N+A	0:07:02	0:06:20	0:13:22	8424	2148	2673	25.50	31.73	26
P+S/N+A	0:16:41	0:09:50	0:26:31	1171	468	649	39.97	55.42	7
U+S/N+A	0:12:41	0:06:07	0:18:48	1284	533	771	41.51	60.05	3
Fp+Fu+S/N+A	0:06:38	0:06:22	0:13:00	7584	2204	2359	29.06	31.10	45

Topografia									
C16.0 a C17.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 18.511	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:47	0:06:39	0:14:26	19829	4896	7370	24.69	37.17	20
P+S/N+DT	0:11:01	0:06:36	0:17:37	11693	744	1078	6.36	9.22	1
U+S/N+DT	0:11:03	0:06:33	0:17:36	11441	978	1609	8.55	14.06	1
Fp+Fu+S/N+DT	0:07:23	0:06:33	0:13:56	16877	5309	6032	31.46	35.74	9
P+U+S+A/N+DT	0:07:00	0:06:27	0:13:27	5247	2922	238	55.69	4.54	4
P+S+A/N+DT	0:10:32	0:06:13	0:16:45	567	292	58	51.50	10.23	0
U+S+A/N+DT	0:10:29	0:06:13	0:16:42	720	372	63	51.67	8.75	0
Fp+Fu+S+A/N+DT	0:06:55	0:06:24	0:13:19	4627	3006	130	64.97	2.81	0
Fp+S+A/N+DT	0:09:33	0:06:13	0:15:46	691	431	69	62.37	9.99	0
Fu+S+A/N+DT	0:09:20	0:06:12	0:15:32	898	554	107	61.69	11.92	0
P+U+S/N+A	0:07:44	0:06:32	0:14:16	15532	5233	7262	33.69	46.76	78
P+S/N+A	0:10:56	0:06:13	0:17:09	1728	766	1063	44.33	61.52	4
U+S/N+A	0:10:51	0:06:24	0:17:15	2103	1009	1612	47.98	76.65	8
Fp+Fu+S/N+A	0:07:11	0:06:32	0:13:43	13774	5694	6193	41.34	44.96	77

Topografia									
C18.0 a C18.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 15.684	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:19:58	0:20:58	0:40:56	16819	3501	6132	20.82	36.46	85
P+S/N+DT	0:27:01	0:21:06	0:48:07	15250	689	1131	4.52	7.42	5
U+S/N+DT	0:27:04	0:20:06	0:47:10	14156	957	1826	6.76	12.90	16
Fp+Fu+S/N+DT	0:18:10	0:20:43	0:38:53	13643	3651	4547	26.76	33.33	9
P+U+S+A/N+DT	0:17:47	0:21:12	0:38:59	4192	2587	355	61.71	8.47	1
P+S+A/N+DT	0:26:06	0:19:36	0:45:42	493	244	158	49.49	32.05	5
U+S+A/N+DT	0:26:49	0:19:52	0:46:41	681	362	263	53.16	38.62	8
Fp+Fu+S+A/N+DT	0:17:23	0:20:47	0:38:10	3765	2629	325	69.83	8.63	1
Fp+S+A/N+DT	0:25:14	0:20:15	0:45:29	593	322	195	54.30	32.88	1
Fu+S+A/N+DT	0:23:39	0:20:26	0:44:05	789	470	339	59.57	42.97	6
P+U+S/N+A	0:19:40	0:21:26	0:41:06	12663	3656	5903	28.87	46.62	127
P+S/N+A	0:27:01	0:20:12	0:47:13	1492	710	1090	47.59	73.06	12
U+S/N+A	0:27:01	0:20:45	0:47:46	2307	976	1749	42.31	75.81	17
Fp+Fu+S/N+A	0:18:25	0:21:22	0:39:47	10758	3814	4570	35.45	42.48	84

Topografia									
C19.0 a C21.8	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 10.615	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:22	0:06:21	0:13:43	11842	2043	3312	17.25	27.97	5
P+S/N+DT	0:10:46	0:06:16	0:17:02	4837	318	569	6.57	11.76	4
U+S/N+DT	0:10:43	0:06:16	0:16:59	5425	427	921	7.87	16.98	3
Fp+Fu+S/N+DT	0:06:53	0:06:19	0:13:12	9913	2162	2600	21.81	26.23	7
P+U+S+A/N+DT	0:06:47	0:06:15	0:13:02	3212	1843	125	57.38	3.89	0
P+S+A/N+DT	0:10:17	0:06:00	0:16:17	357	182	43	50.98	12.04	0
U+S+A/N+DT	0:10:41	0:06:12	0:16:53	717	291	65	40.59	9.07	0
Fp+Fu+S+A/N+DT	0:06:32	0:06:23	0:12:55	2990	1998	94	66.82	3.14	0
Fp+S+A/N+DT	0:09:19	0:06:09	0:15:28	430	256	57	59.53	13.26	0
Fu+S+A/N+DT	0:09:09	0:06:13	0:15:22	768	383	64	49.87	8.33	0
P+U+S/N+A	0:07:19	0:06:30	0:13:49	9471	1752	3308	18.50	34.93	22
P+S/N+A	0:10:49	0:06:13	0:17:02	899	266	576	29.59	64.07	1
U+S/N+A	0:10:41	0:06:14	0:16:55	1565	371	945	23.71	60.38	6
Fp+Fu+S/N+A	0:06:57	0:06:28	0:13:25	8224	1866	2735	22.69	33.26	57

Topografia									
C22.0 a C24.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 4.021	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:07	0:07:13	0:15:20	8547	1525	2307	17.84	26.99	4
P+S/N+DT	0:12:40	0:07:18	0:19:58	11139	446	680	4.00	6.10	2
U+S/N+DT	0:12:22	0:07:15	0:19:37	8772	501	857	5.71	9.77	3
Fp+Fu+S/N+DT	0:07:55	0:07:12	0:15:07	6956	1562	1808	22.46	25.99	2
P+U+S+A/N+DT	0:07:35	0:06:56	0:14:31	1224	721	129	58.91	10.54	0
P+S+A/N+DT	0:11:18	0:07:01	0:18:19	260	108	88	41.54	33.85	0
U+S+A/N+DT	0:11:07	0:06:56	0:18:03	429	165	108	38.46	25.17	1
Fp+Fu+S+A/N+DT	0:07:33	0:06:58	0:14:31	1092	727	95	66.58	8.70	0
Fp+S+A/N+DT	0:10:30	0:06:58	0:17:28	259	132	75	50.97	28.96	0
Fu+S+A/N+DT	0:10:29	0:06:58	0:17:27	383	176	95	45.95	24.80	0
P+U+S/N+A	0:08:06	0:07:14	0:15:20	5892	1279	2304	21.71	39.10	37
P+S/N+A	0:12:30	0:06:56	0:19:26	1002	361	669	36.03	66.77	4
U+S/N+A	0:12:20	0:06:57	0:19:17	1216	407	840	33.47	69.08	9
Fp+Fu+S/N+A	0:07:54	0:07:13	0:15:07	4895	1307	1821	26.70	37.20	33

Topografia									
C25.0 a C26.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.896	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:28	0:07:13	0:15:41	12072	2533	3831	20.98	31.73	6
P+S/N+DT	0:12:44	0:07:14	0:19:58	11018	571	829	5.18	7.52	1
U+S/N+DT	0:12:38	0:07:11	0:19:49	8873	759	1302	8.55	14.67	3
Fp+Fu+S/N+DT	0:08:04	0:07:12	0:15:16	10019	2626	2995	26.21	29.89	1
P+U+S+A/N+DT	0:07:29	0:06:54	0:14:23	1337	788	89	58.94	6.66	0
P+S+A/N+DT	0:11:02	0:06:51	0:17:53	631	418	73	66.24	11.57	0
U+S+A/N+DT	0:10:45	0:06:50	0:17:35	500	353	73	70.60	14.60	0
Fp+Fu+S+A/N+DT	0:07:26	0:06:53	0:14:19	1230	792	71	64.39	5.77	0
Fp+S+A/N+DT	0:10:14	0:06:52	0:17:06	651	437	60	67.13	9.22	0
Fu+S+A/N+DT	0:10:15	0:06:52	0:17:07	489	364	68	74.44	13.91	0
P+U+S/N+A	0:08:36	0:07:12	0:15:48	9157	2032	3699	22.19	40.40	42
P+S/N+A	0:12:41	0:06:54	0:19:35	1690	444	866	26.27	51.24	14
U+S/N+A	0:12:34	0:06:53	0:19:27	1722	588	1294	34.15	75.15	10
Fp+Fu+S/N+A	0:08:05	0:07:11	0:15:16	7859	2113	3065	26.89	39.00	57

Topografia									
C30.0 a C33.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 6.260	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:29	0:07:15	0:15:44	9577	1423	2060	14.86	21.51	5
P+S/N+DT	0:12:18	0:07:13	0:19:31	7679	244	372	3.18	4.84	1
U+S/N+DT	0:12:11	0:07:11	0:19:22	6151	316	557	5.14	9.06	1
Fp+Fu+S/N+DT	0:08:03	0:07:15	0:15:18	8103	1498	1707	18.49	21.07	1
P+U+S+A/N+DT	0:07:48	0:07:08	0:14:56	2482	1537	60	61.93	2.42	0
P+S+A/N+DT	0:11:48	0:06:57	0:18:45	382	207	31	54.19	8.12	0
U+S+A/N+DT	0:11:46	0:06:56	0:18:42	434	184	35	42.40	8.06	0
Fp+Fu+S+A/N+DT	0:07:46	0:07:12	0:14:58	2228	1564	49	70.20	2.20	1
Fp+S+A/N+DT	0:10:49	0:06:57	0:17:46	418	245	31	58.61	7.42	0
Fu+S+A/N+DT	0:10:54	0:07:02	0:17:56	447	228	42	51.01	9.40	0
P+U+S/N+A	0:08:39	0:07:12	0:15:51	7131	1245	2093	17.46	29.35	20
P+S/N+A	0:12:30	0:06:36	0:19:06	849	202	386	23.79	45.47	12
U+S/N+A	0:12:24	0:06:54	0:19:18	914	268	568	29.32	62.14	5
Fp+Fu+S/N+A	0:08:17	0:07:12	0:15:29	6241	1312	1812	21.02	29.03	35

Topografia									
C34.0 a C34.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 17.455	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:16	0:06:59	0:15:15	21869	5842	9204	26.71	42.08	18
P+S/N+DT	0:11:40	0:06:58	0:18:38	15814	825	1298	5.21	8.20	0
U+S/N+DT	0:11:38	0:06:56	0:18:34	14136	1091	2072	7.71	14.65	11
Fp+Fu+S/N+DT	0:06:01	0:06:57	0:12:58	18042	6276	7360	34.78	40.79	6
P+U+S+A/N+DT	0:07:10	0:06:49	0:13:59	3723	2275	406	61.10	10.90	0
P+S+A/N+DT	0:11:00	0:07:41	0:18:41	498	261	136	52.40	27.30	0
U+S+A/N+DT	0:10:36	0:06:37	0:17:13	760	397	238	52.23	31.31	1
Fp+Fu+S+A/N+DT	0:07:05	0:06:50	0:13:55	3151	2322	394	73.69	12.50	3
Fp+S+A/N+DT	0:09:52	0:06:37	0:16:29	579	362	204	62.52	35.23	0
Fu+S+A/N+DT	0:09:51	0:06:37	0:16:28	802	491	292	61.22	36.40	3
P+U+S/N+A	0:08:23	0:06:58	0:15:21	17389	6038	9030	34.72	51.92	92
P+S/N+A	0:11:10	0:06:39	0:17:49	1632	833	838	51.04	51.34	90
U+S/N+A	0:11:29	0:06:47	0:18:16	2406	1107	1984	46.00	82.46	0
Fp+Fu+S/N+A	0:07:48	0:06:59	0:14:47	14869	6502	7348	43.72	49.41	6

Topografia									
C37.0 a C39.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 854	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:17	0:07:03	0:14:20	3168	268	415	8.46	13.10	0
P+S/N+DT	0:11:14	0:07:07	0:18:21	7156	138	213	1.93	2.98	0
U+S/N+DT	0:11:06	0:07:07	0:18:13	6348	142	250	2.24	3.94	5
Fp+Fu+S/N+DT	0:07:24	0:07:02	0:14:26	2732	270	305	9.88	11.16	4
P+U+S+A/N+DT	0:06:58	0:06:48	0:13:46	631	329	6	52.14	0.95	0
P+S+A/N+DT	0:09:27	0:06:50	0:16:17	300	112	7	37.33	2.33	0
U+S+A/N+DT	0:09:33	0:06:48	0:16:21	271	115	13	42.44	4.80	0
Fp+Fu+S+A/N+DT	0:07:10	0:06:51	0:14:01	587	329	8	56.05	1.36	0
Fp+S+A/N+DT	0:08:56	0:06:49	0:15:45	279	114	6	40.86	2.15	0
Fu+S+A/N+DT	0:09:04	0:06:49	0:15:53	254	115	7	45.28	2.76	0
P+U+S/N+A	0:07:17	0:06:50	0:14:07	1933	222	412	11.48	21.31	10
P+S/N+A	0:11:03	0:06:50	0:17:53	567	108	210	19.05	37.04	1
U+S/N+A	0:11:01	0:06:49	0:17:50	696	105	265	15.09	38.07	11
Fp+Fu+S/N+A	0:07:25	0:06:48	0:14:13	1667	225	338	13.50	20.28	17

Topografia									
C40.0 a C42.4	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 11.358	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:21	0:07:08	0:15:29	13580	3023	4874	22.26	35.89	8
P+S/N+DT	0:12:05	0:07:08	0:19:13	12624	441	769	3.49	6.09	3
U+S/N+DT	0:10:43	0:07:05	0:17:48	8533	611	1188	7.16	13.92	2
Fp+Fu+S/N+DT	0:07:43	0:07:51	0:15:34	11205	3226	3850	28.79	34.36	6
P+U+S+A/N+DT	0:07:05	0:06:48	0:13:53	1858	1066	148	57.37	7.97	2
P+S+A/N+DT	0:10:17	0:07:51	0:18:08	498	289	79	58.03	15.86	2
U+S+A/N+DT	0:10:17	0:06:35	0:16:52	278	81	89	29.14	32.01	1
Fp+Fu+S+A/N+DT	0:06:53	0:06:36	0:13:29	1610	1079	126	67.02	7.83	2
Fp+S+A/N+DT	0:09:39	0:06:36	0:16:15	557	350	94	62.84	16.88	3
Fu+S+A/N+DT	0:09:31	0:06:35	0:16:06	266	138	89	51.88	33.46	0
P+U+S/N+A	0:08:02	0:06:53	0:14:55	9942	2503	4848	25.18	48.76	61
P+S/N+A	0:10:42	0:06:36	0:17:18	1225	374	750	30.53	61.22	7
U+S/N+A	0:10:50	0:06:36	0:17:26	1331	497	1157	37.34	86.93	9
Fp+Fu+S/N+A	0:07:35	0:06:53	0:14:28	8605	2657	3896	30.88	45.28	41

Topografia									
C44.0 a C44.9 (8720 a 8780)	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.034	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:16	0:07:14	0:15:30	5745	398	803	6.93	13.98	3
P+S/N+DT	0:12:44	0:07:17	0:20:01	9392	167	288	1.78	3.07	3
U+S/N+DT	0:12:35	0:07:16	0:19:51	9039	203	437	2.25	4.83	2
Fp+Fu+S/N+DT	0:07:47	0:07:11	0:14:58	4747	404	608	8.51	12.81	4
P+U+S+A/N+DT	0:08:02	0:07:10	0:15:12	3001	1595	190	53.15	6.33	0
P+S+A/N+DT	0:12:25	0:06:53	0:19:18	402	219	75	54.48	18.66	0
U+S+A/N+DT	0:12:24	0:06:51	0:19:15	507	181	101	35.70	19.92	0
Fp+Fu+S+A/N+DT	0:07:38	0:07:05	0:14:43	2562	1634	98	63.78	3.83	0
Fp+S+A/N+DT	0:11:03	0:06:53	0:17:56	434	267	99	61.52	22.81	0
Fu+S+A/N+DT	0:10:53	0:06:52	0:17:45	568	318	152	55.99	26.76	0
P+U+S/N+A	0:08:13	0:07:08	0:15:21	4097	523	803	12.77	19.60	3
P+S/N+A	0:12:31	0:07:04	0:19:35	648	163	290	25.15	44.75	2
U+S/N+A	0:12:30	0:06:56	0:19:26	954	211	458	22.12	48.01	2
Fp+Fu+S/N+A	0:07:45	0:07:14	0:14:59	3441	540	606	15.69	17.61	6

Topografia									
C44.0 a C44.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 70.662	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:42	0:08:01	0:17:43	21157	1983	4113	9.37	19.44	4
P+S/N+DT	0:12:36	0:07:52	0:20:28	10479	175	371	1.67	3.54	0
U+S/N+DT	0:12:37	0:07:52	0:20:29	9978	265	671	2.66	6.72	0
Fp+Fu+S/N+DT	0:08:44	0:07:57	0:16:41	17545	2200	2816	12.54	16.05	2
P+U+S+A/N+DT	0:09:16	0:07:52	0:17:08	11024	5800	1608	52.61	14.59	2
P+S+A/N+DT	0:12:31	0:07:14	0:19:45	736	386	52	52.45	7.07	0
U+S+A/N+DT	0:10:49	0:07:11	0:18:00	627	201	87	32.06	13.88	0
Fp+Fu+S+A/N+DT	0:08:12	0:07:31	0:15:43	9718	6699	61	68.93	0.00	0
Fp+S+A/N+DT	0:10:28	0:07:13	0:17:41	981	560	59	57.08	6.01	0
Fu+S+A/N+DT	0:10:46	0:07:13	0:17:59	1541	862	189	55.94	12.26	0
P+U+S/N+A	0:09:07	0:07:36	0:16:43	16214	2036	4174	12.56	25.74	8
P+S/N+A	0:11:49	0:07:14	0:19:03	775	153	313	19.74	40.39	2
U+S/N+A	0:11:59	0:07:36	0:19:35	1055	229	601	21.71	56.97	4
Fp+Fu+S/N+A	0:08:21	0:07:34	0:15:55	14116	2297	2987	16.27	21.16	9

Topografia									
C47.0 a C49.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 2.830	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:00	0:06:47	0:13:47	5220	551	841	10.56	16.11	2
P+S/N+DT	0:10:28	0:06:50	0:17:18	8522	180	310	2.11	3.64	0
U+S/N+DT	0:10:20	0:06:46	0:17:06	5856	214	357	3.65	6.10	0
Fp+Fu+S/N+DT	0:06:29	0:06:46	0:13:15	4450	574	678	12.90	15.24	1
P+U+S+A/N+DT	0:06:49	0:06:30	0:13:19	1044	617	29	59.10	2.78	0
P+S+A/N+DT	0:09:52	0:06:27	0:16:19	210	104	9	49.52	4.29	0
U+S+A/N+DT	0:09:44	0:06:29	0:16:13	308	158	14	51.30	4.55	0
Fp+Fu+S+A/N+DT	0:06:39	0:06:31	0:13:10	920	618	14	67.17	1.52	0
Fp+S+A/N+DT	0:09:14	0:06:27	0:15:41	200	108	11	54.00	5.50	0
Fu+S+A/N+DT	0:09:13	0:06:29	0:15:42	317	171	11	53.94	3.47	0
P+U+S/N+A	0:06:57	0:06:42	0:13:39	3385	474	833	14.00	24.61	7
P+S/N+A	0:10:23	0:06:30	0:16:53	654	137	305	20.95	46.64	3
U+S/N+A	0:10:20	0:06:29	0:16:49	503	171	351	34.00	69.78	0
Fp+Fu+S/N+A	0:06:54	0:06:41	0:13:35	2999	497	698	16.57	23.27	5

Topografia									
C50.0 a C50.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 41.307	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:52	0:07:18	0:15:10	13316	3677	6989	27.61	52.49	16
P+S/N+DT	0:10:58	0:07:04	0:18:02	10175	320	551	3.14	5.42	0
U+S/N+DT	0:10:32	0:07:04	0:17:36	10355	4130	873	39.88	8.43	1
Fp+Fu+S/N+DT	0:07:30	0:07:04	0:14:34	10800	4256	5448	39.41	50.44	3
P+U+S+A/N+DT	0:07:18	0:06:58	0:14:16	3078	1704	451	55.36	14.65	1
P+S+A/N+DT	0:10:04	0:06:48	0:16:52	337	200	109	59.35	32.34	0
U+S+A/N+DT	0:10:06	0:06:47	0:16:53	426	216	152	50.70	35.68	2
Fp+Fu+S+A/N+DT	0:07:17	0:06:56	0:14:13	2771	1824	147	65.82	0.00	3
Fp+S+A/N+DT	0:09:32	0:06:46	0:16:18	415	290	178	69.88	42.89	2
Fu+S+A/N+DT	0:09:32	0:06:47	0:16:19	638	339	286	53.13	44.83	1
P+U+S/N+A	0:07:48	0:07:02	0:14:50	10119	3805	6748	37.60	66.69	28
P+S/N+A	0:09:17	0:06:52	0:16:09	743	328	525	44.15	70.66	0
U+S/N+A	0:10:29	0:06:58	0:17:27	1220	424	852	34.75	69.84	1
Fp+Fu+S/N+A	0:07:28	0:07:02	0:14:30	8666	4396	5415	50.73	62.49	20

Topografia									
C51.0 a C58.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 28.359	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U/N+DT	0:25:06	0:13:19	0:38:25	270609	5049	7728	1.87	2.86	13
Fp+Fu/N+DT	0:20:50	0:13:29	0:34:19	228058	5049	6002	2.21	2.63	5
P+U+A/N+DT	0:07:39	0:08:19	0:15:58	47460	5049	7348	10.64	15.48	5
P+A/N+DT*	0:22:16	0:21:32	0:43:48	493688	5049	5048	1.02	1.02	10
U+A/N+DT**	0:22:32	0:37:02	0:59:34	1309283	5049	7822	0.39	0.60	12
Fp+Fu+A/N+DT	0:07:40	0:08:06	0:15:46	38495	5049	5887	13.12	15.29	4
Fp+A/N+DT	0:18:05	0:34:31	0:52:36	1188694	5049	6595	0.42	0.55	5
Fu+A/N+DT	0:20:34	0:35:02	0:55:36	1233726	5049	7079	0.41	0.57	7
P+U/N+A	0:23:28	0:10:15	0:33:43	123098	5039	7808	4.09	6.34	105
Fp+Fu/N+A	0:19:18	0:09:52	0:29:10	107114	5039	6988	4.70	6.52	612

* 96 horas

** 96 horas

Topografia									
C60.0 a C60.9 e C62.0 a C63.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 2.153	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U/N+DT	0:07:58	0:07:24	0:15:22	13246	280	387	2.11	2.92	0
Fp+Fu/N+DT	0:07:47	0:07:24	0:15:11	11406	280	328	2.45	2.88	1
P+U+A/N+DT	0:06:33	0:06:50	0:13:23	2032	280	371	13.78	18.26	0
P+A/N+DT	0:07:40	0:08:05	0:15:45	54420	280	375	0.51	0.69	1
U+A/N+DT	0:08:17	0:08:42	0:16:59	74124	280	374	0.38	0.50	0
Fp+Fu+A/N+DT	0:06:45	0:06:59	0:13:44	1683	280	311	16.64	18.48	0
Fp+A/N+DT	0:07:43	0:07:52	0:15:35	43248	280	349	0.65	0.81	0
Fu+A/N+DT	0:08:28	0:09:13	0:17:41	69776	280	349	0.40	0.50	0
P+U/N+A	0:07:55	0:07:16	0:15:11	7094	295	383	4.16	5.40	5
Fp+Fu/N+A	0:07:41	0:07:12	0:14:53	6310	295	380	4.68	6.02	28

Topografia									
C61.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 29.455	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U/N+DT	0:16:53	0:11:04	0:27:57	221695	5270	7731	2.38	3.49	10
Fp+Fu/N+DT	0:14:41	0:11:18	0:25:59	202910	5270	6670	2.60	3.29	131
P+U+A/N+DT	0:06:43	0:07:17	0:14:00	32182	5270	7265	16.38	22.57	23
P+A/N+DT*	0:20:42	0:27:47	0:48:29	604343	5270	7414	0.50	1.23	6
U+A/N+DT**	0:25:19	0:23:00	0:48:19	1056482	5270	7298	18.56	0.69	281
Fp+Fu+A/N+DT	0:06:46	0:29:04	0:35:50	28390	5270	5701	0.54	20.08	2
Fp+A/N+DT	0:18:40	0:17:00	0:35:40	978165	5270	6516	0.54	0.67	4
Fu+A/N+DT	0:20:25	0:40:33	1:00:58	1013810	5270	6583	0.52	0.65	7
P+U/N+A	0:16:17	0:09:04	0:25:21	135646	6561	8840	4.84	6.52	791

Topografia									
C64.0 a C68.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 12.343	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:07	0:07:22	0:16:29	12639	2160	3417	17.09	27.04	2
P+S/N+DT	0:13:17	0:07:18	0:20:35	7716	293	466	3.80	6.04	0
U+S/N+DT	0:13:16	0:07:20	0:20:36	6825	392	754	5.74	11.05	0
Fp+Fu+S/N+DT	0:08:28	0:07:19	0:15:47	10369	2314	2671	22.32	25.76	4
P+U+S+A/N+DT	0:08:18	0:07:14	0:15:32	3662	2160	66	58.98	1.80	0
P+S+A/N+DT	0:12:42	0:07:00	0:19:42	508	202	27	39.76	5.31	0
U+S+A/N+DT	0:12:42	0:06:59	0:19:41	607	311	36	51.24	5.93	0
Fp+Fu+S+A/N+DT	0:07:56	0:07:13	0:15:09	3155	2271	72	71.98	2.28	0
Fp+S+A/N+DT	0:11:26	0:06:59	0:18:25	530	253	48	47.74	9.06	0
Fu+S+A/N+DT	0:11:27	0:06:58	0:18:25	687	442	62	64.34	9.02	0
P+U+S/N+A	0:09:13	0:07:18	0:16:31	9673	1869	3477	19.32	35.95	44
P+S/N+A	0:13:26	0:06:56	0:20:22	1025	224	478	21.85	46.63	1
U+S/N+A	0:13:12	0:06:56	0:20:08	1105	335	756	30.32	68.42	2
Fp+Fu+S/N+A	0:08:29	0:07:16	0:15:45	8303	2004	2876	24.14	34.64	23

Topografia									
C69.0 a C72.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 7.199	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:41	0:07:17	0:15:58	10974	1970	2980	17.95	27.16	11
P+S/N+DT	0:12:58	0:07:24	0:20:22	13477	421	665	3.12	4.93	1
U+S/N+DT	0:12:57	0:07:17	0:20:14	10519	494	921	4.70	8.76	5
Fp+Fu+S/N+DT	0:08:15	0:07:14	0:15:29	9098	2062	2417	22.66	26.57	7
P+U+S+A/N+DT	0:07:51	0:06:57	0:14:48	1394	768	119	55.09	8.54	0
P+S+A/N+DT	0:11:48	0:06:55	0:18:43	318	151	44	47.48	13.84	0
U+S+A/N+DT	0:11:33	0:06:51	0:18:24	286	113	78	39.51	27.27	1
Fp+Fu+S+A/N+DT	0:07:38	0:06:56	0:14:34	1246	799	93	64.13	7.46	0
Fp+S+A/N+DT	0:10:44	0:06:56	0:17:40	329	174	48	52.89	14.59	0
Fu+S+A/N+DT	0:10:51	0:06:51	0:17:42	327	137	89	41.90	27.22	1
P+U+S/N+A	0:08:46	0:07:15	0:16:01	7529	1636	2957	21.73	39.27	27
P+S/N+A	0:12:55	0:06:55	0:19:50	956	336	650	35.15	67.99	5
U+S/N+A	0:12:44	0:06:52	0:19:36	1181	397	901	33.62	76.29	9
Fp+Fu+S/N+A	0:08:17	0:07:09	0:15:26	6457	1715	2463	26.56	38.14	40

Topografia									
C73.0 a C75.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 11.696	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:45	0:07:17	0:16:02	5077	417	741	8.21	14.60	3
P+S/N+DT	0:13:07	0:07:19	0:20:26	4966	90	159	1.81	3.20	1
U+S/N+DT	0:13:07	0:07:15	0:20:22	4394	92	185	2.09	4.21	0
Fp+Fu+S/N+DT	0:08:21	0:07:17	0:15:38	4244	439	572	10.34	13.48	2
P+U+S+A/N+DT	0:08:21	0:07:06	0:15:27	1444	645	71	44.67	4.92	0
P+S+A/N+DT	0:12:43	0:07:06	0:19:49	288	71	19	24.65	6.60	1
U+S+A/N+DT	0:12:42	0:07:03	0:19:45	187	60	34	32.09	18.18	0
Fp+Fu+S+A/N+DT	0:07:55	0:06:57	0:14:52	1307	699	60	53.48	4.59	0
Fp+S+A/N+DT	0:11:19	0:06:55	0:18:14	320	101	36	31.56	11.25	0
Fu+S+A/N+DT	0:11:01	0:06:58	0:17:59	218	83	49	38.07	22.48	0
P+U+S/N+A	0:08:37	0:07:12	0:15:49	3456	357	736	10.33	21.30	10
P+S/N+A	0:12:52	0:06:57	0:19:49	384	68	148	17.71	38.54	0
U+S/N+A	0:12:51	0:06:55	0:19:46	270	77	181	28.52	67.04	1
Fp+Fu+S/N+A	0:08:10	0:07:10	0:15:20	3030	385	600	12.71	19.80	10

Topografia									
C76.0 a C80.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 23.720	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:44	0:07:06	0:15:50	17881	4873	7902	27.25	44.19	7
P+S/N+DT	0:11:58	0:07:00	0:18:58	7529	599	913	7.96	12.13	0
U+S/N+DT	0:11:42	0:06:59	0:18:41	6815	668	1319	9.80	19.35	2
Fp+Fu+S/N+DT	0:08:00	0:07:03	0:15:03	14987	5339	6556	35.62	43.74	5
P+U+S+A/N+DT	0:07:41	0:07:04	0:14:45	5483	2966	720	54.09	13.13	0
P+S+A/N+DT	0:10:43	0:06:44	0:17:27	557	302	192	54.22	34.47	0
U+S+A/N+DT	0:10:39	0:06:45	0:17:24	555	259	292	46.67	52.61	1
Fp+Fu+S+A/N+DT	0:07:41	0:07:07	0:14:48	4846	3133	603	64.65	12.44	3
Fp+S+A/N+DT	0:10:25	0:06:42	0:17:07	687	431	261	62.74	37.99	0
Fu+S+A/N+DT	0:10:08	0:06:42	0:16:50	836	457	422	54.67	50.48	2
P+U+S/N+A	0:08:39	0:07:02	0:15:41	14128	4030	7881	28.52	55.78	55

RCBP-SP versus Apac-SIA/SUS

Topografia									
C00.0 a C09.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 9.321	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:18	0:07:10	0:15:28	8034	320	517	3.98	6.44	93
P+S/N+DT	0:11:55	0:07:10	0:19:05	9042	62	117	0.69	1.29	14
U+S/N+DT	0:10:59	0:07:08	0:18:07	7298	69	155	0.95	2.12	33
Fp+Fu+S/N+DT	0:07:49	0:07:09	0:14:58	7160	349	433	4.87	6.05	75
P+U+S+A/N+DT	0:07:52	0:07:05	0:14:57	3618	1823	76	50.39	2.10	5
P+S+A/N+DT	0:11:42	0:06:51	0:18:33	493	193	17	39.15	3.45	2
U+S+A/N+DT	0:12:05	0:06:50	0:18:55	422	163	38	38.63	9.00	5
Fp+Fu+S+A/N+DT	0:07:41	0:07:03	0:14:44	3288	1858	47	56.51	1.43	2
Fp+S+A/N+DT	0:10:25	0:06:50	0:17:15	561	262	24	46.70	4.28	1
Fu+S+A/N+DT	0:10:59	0:06:50	0:17:49	450	202	47	44.89	10.44	3
P+U+S/N+A	0:08:28	0:07:22	0:15:50	23481	478	580	2.04	2.47	83
P+S/N+A	0:11:30	0:07:52	0:19:22	58723	97	157	0.17	0.27	41
U+S/N+A	0:11:05	0:07:39	0:18:44	43982	96	178	0.22	0.40	55
Fp+Fu+S/N+A	0:07:57	0:07:42	0:15:39	18938	512	551	2.70	2.91	126

Topografia									
C10.0 a C11.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 1.613	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:06:26	0:06:31	0:12:57	4135	96	72	2.32	1.74	5
P+S/N+DT	0:09:28	0:06:35	0:16:03	8735	64	53	0.73	0.61	2
U+S/N+DT	0:09:35	0:06:35	0:16:10	8486	37	42	0.44	0.49	0
Fp+Fu+S/N+DT	0:07:07	0:06:32	0:13:39	3711	96	60	2.59	1.62	1
P+U+S+A/N+DT	0:06:11	0:06:17	0:12:28	936	443	11	47.33	1.18	0
P+S+A/N+DT	0:08:26	0:06:17	0:14:43	312	130	9	41.67	2.88	0
U+S+A/N+DT	0:09:04	0:06:16	0:15:20	202	63	4	31.19	1.98	0
Fp+Fu+S+A/N+DT	0:06:30	0:06:23	0:12:53	833	443	9	53.18	1.08	0
Fp+S+A/N+DT	0:08:02	0:06:21	0:14:23	326	147	9	45.09	2.76	0
Fu+S+A/N+DT	0:08:45	0:06:22	0:15:07	202	69	9	34.16	4.46	0
P+U+S/N+A	0:06:31	0:06:43	0:13:14	11339	143	87	1.26	0.77	0
P+S/N+A	0:09:37	0:06:57	0:16:34	38771	72	50	0.19	0.13	0
U+S/N+A	0:10:01	0:06:58	0:16:59	33425	56	34	0.17	0.10	0
Fp+Fu+S/N+A	0:06:32	0:06:38	0:13:10	9798	143	69	1.46	0.70	0

Topografia									
C12.0 a C14.8	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 1.381	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:44	0:07:29	0:15:13	3131	67	57	2.14	1.82	1
P+S/N+DT	0:11:05	0:07:30	0:18:35	5918	26	28	0.44	0.47	1
U+S/N+DT	0:11:38	0:07:21	0:18:59	4922	25	31	0.51	0.63	2
Fp+Fu+S/N+DT	0:07:37	0:07:21	0:14:58	2988	72	54	2.41	1.81	2
P+U+S+A/N+DT	0:07:30	0:07:07	0:14:37	709	412	4	58.11	0.56	0
P+S+A/N+DT	0:10:14	0:07:08	0:17:22	225	76	2	33.78	0.89	0
U+S+A/N+DT	0:10:38	0:07:08	0:17:46	213	97	3	45.54	1.41	0
Fp+Fu+S+A/N+DT	0:07:42	0:07:06	0:14:48	654	412	3	63.00	0.46	0
Fp+S+A/N+DT	0:09:24	0:07:06	0:16:30	209	77	2	36.84	0.96	0
Fu+S+A/N+DT	0:10:08	0:07:10	0:17:18	226	103	56	45.58	24.78	0
P+U+S/N+A	0:07:42	0:07:28	0:15:10	6944	84	53	1.21	0.76	0
P+S/N+A	0:10:52	0:07:37	0:18:29	21538	2	16	0.01	0.07	0
U+S/N+A	0:11:30	0:07:37	0:19:07	18152	25	32	0.14	0.18	0
Fp+Fu+S/N+A	0:07:08	0:07:27	0:14:35	6348	89	53	1.40	0.83	0

Topografia									
C15.0 a C15.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.579	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:29	0:07:29	0:15:58	7011	83	145	1.18	2.07	2
P+S/N+DT	0:12:31	0:07:34	0:20:05	9959	31	45	0.31	0.45	0
U+S/N+DT	0:12:56	0:07:35	0:20:31	10673	29	47	0.27	0.44	4
Fp+Fu+S/N+DT	0:08:31	0:07:30	0:16:01	6405	95	111	1.48	1.73	18
P+U+S+A/N+DT	0:11:34	0:07:20	0:18:54	470	228	16	48.51	3.40	1
P+S+A/N+DT	0:12:14	0:07:14	0:19:28	414	122	13	29.47	3.14	1
U+S+A/N+DT	0:08:20	0:07:29	0:15:49	2866	1188	35	41.45	1.22	0
Fp+Fu+S+A/N+DT	0:08:16	0:07:32	0:15:48	2717	1232	21	45.34	0.77	0
Fp+S+A/N+DT	0:10:52	0:07:12	0:18:04	473	252	20	53.28	4.23	0
Fu+S+A/N+DT	0:12:01	0:07:12	0:19:13	466	165	11	35.41	2.36	0
P+U+S/N+A	0:10:00	0:07:41	0:17:41	18173	152	105	0.84	0.58	1
P+S/N+A	0:12:41	0:08:08	0:20:49	44015	36	38	0.08	0.09	0
U+S/N+A	0:12:59	0:08:04	0:21:03	42980	37	34	0.09	0.08	0
Fp+Fu+S/N+A	0:08:32	0:07:35	0:16:07	15755	168	84	1.07	0.53	0

Topografia									
C16.0 a C17.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 18.511	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:31	0:07:42	0:17:13	11269	268	331	2.38	2.94	7
P+S/N+DT	0:13:59	0:07:42	0:21:41	15289	61	79	0.40	0.52	1
U+S/N+DT	0:14:03	0:07:41	0:21:44	12215	66	131	0.54	1.07	0
Fp+Fu+S/N+DT	0:08:57	0:07:36	0:16:33	10111	299	334	2.96	3.30	70
P+U+S+A/N+DT	0:09:06	0:07:34	0:16:40	4842	2188	64	45.19	1.32	2
P+S+A/N+DT	0:13:30	0:07:16	0:20:46	479	160	21	33.40	4.38	0
U+S+A/N+DT	0:13:35	0:07:16	0:20:51	457	198	22	43.33	4.81	0
Fp+Fu+S+A/N+DT	0:08:41	0:07:36	0:16:17	4421	2252	54	50.94	1.22	0
Fp+S+A/N+DT	0:12:04	0:07:18	0:19:22	576	254	37	44.10	6.42	0
Fu+S+A/N+DT	0:12:03	0:07:20	0:19:23	523	260	33	49.71	6.31	0
P+U+S/N+A	0:09:39	0:07:52	0:17:31	26225	402	393	1.53	1.50	8
P+S/N+A	0:14:10	0:08:23	0:22:33	54343	67	85	0.12	0.16	2
U+S/N+A	0:14:07	0:08:14	0:22:21	46669	88	116	0.19	0.25	6
Fp+Fu+S/N+A	0:09:02	0:07:51	0:16:53	21769	450	365	2.07	1.68	6

Topografia									
C18.0 a C18.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 15.684	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:30	0:07:13	0:15:43	8379	235	274	2.80	3.27	53
P+S/N+DT	0:11:07	0:07:14	0:18:21	10407	58	118	0.56	1.13	32
U+S/N+DT	0:12:30	0:07:12	0:19:42	7888	59	120	0.75	1.52	21
Fp+Fu+S/N+DT	0:07:57	0:07:12	0:15:09	7398	265	387	3.58	5.23	74
P+U+S+A/N+DT	0:08:11	0:07:02	0:15:13	3657	1653	90	45.20	2.46	10
P+S+A/N+DT	0:11:53	0:06:49	0:18:42	650	242	33	37.23	5.08	4
U+S+A/N+DT	0:11:01	0:06:50	0:17:51	507	182	39	35.90	7.69	3
Fp+Fu+S+A/N+DT	0:07:41	0:07:03	0:14:44	3370	1679	61	49.82	1.81	2
Fp+S+A/N+DT	0:10:48	0:06:51	0:17:39	685	276	31	40.29	4.53	4
Fu+S+A/N+DT	0:10:30	0:06:50	0:17:20	569	244	35	42.88	6.15	0
P+U+S/N+A	0:08:20	0:07:13	0:15:33	16963	288	457	1.70	2.69	28
P+S/N+A	0:11:38	0:07:30	0:19:08	42568	65	105	0.15	0.25	28
U+S/N+A	0:12:26	0:07:20	0:19:46	27786	65	130	0.23	0.47	22
Fp+Fu+S/N+A	0:07:55	0:07:11	0:15:06	14169	324	415	2.29	2.93	94

Topografia									
C19.0 a C21.8	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 10.615	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:45	0:06:37	0:14:22	7076	288	357	4.07	5.05	1
P+S/N+DT	0:11:28	0:06:34	0:18:02	2738	58	83	2.12	3.03	1
U+S/N+DT	0:11:36	0:06:34	0:18:10	2679	50	110	1.87	4.11	0
Fp+Fu+S/N+DT	0:07:19	0:06:37	0:13:56	6341	341	397	5.38	6.26	3
P+U+S+A/N+DT	0:07:23	0:06:34	0:13:57	3437	1500	115	43.64	3.35	1
P+S+A/N+DT	0:11:02	0:06:18	0:17:20	327	131	28	40.06	8.56	0
U+S+A/N+DT	0:11:19	0:06:18	0:17:37	674	217	24	32.20	3.56	0
Fp+Fu+S+A/N+DT	0:07:07	0:06:34	0:13:41	3282	1632	38	49.73	1.16	0
Fp+S+A/N+DT	0:07:57	0:06:21	0:14:18	410	196	25	47.80	6.10	0
Fu+S+A/N+DT	0:10:27	0:06:20	0:16:47	757	293	34	38.71	4.49	0
P+U+S/N+A	0:07:46	0:06:44	0:14:30	14455	366	357	2.53	2.47	1
P+S/N+A	0:11:30	0:06:41	0:18:11	10748	56	72	0.52	0.67	0
U+S/N+A	0:11:40	0:06:39	0:18:19	9941	57	97	0.57	0.98	0
Fp+Fu+S/N+A	0:07:21	0:06:39	0:14:00	11678	406	325	3.48	2.78	1

Topografia									
C22.0 a C24.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 4.021	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:06:51	0:06:28	0:13:19	4935	61	57	1.24	1.16	0
P+S/N+DT	0:11:10	0:06:36	0:17:46	10228	33	33	0.32	0.32	1
U+S/N+DT	0:10:40	0:06:32	0:17:12	8979	26	30	0.29	0.33	0
Fp+Fu+S/N+DT	0:06:53	0:06:30	0:13:23	4434	62	51	1.40	1.15	0
P+U+S+A/N+DT	0:06:32	0:06:15	0:12:47	1184	553	11	46.71	0.93	0
P+S+A/N+DT	0:09:34	0:06:18	0:15:52	285	71	5	24.91	1.75	0
U+S+A/N+DT	0:09:58	0:06:15	0:16:13	391	92	4	23.53	1.02	1
Fp+Fu+S+A/N+DT	0:06:42	0:06:15	0:12:57	1080	560	4	51.85	0.37	0
Fp+S+A/N+DT	0:08:40	0:06:12	0:14:52	273	79	7	28.94	2.56	0
Fu+S+A/N+DT	0:09:21	0:06:14	0:15:35	341	95	4	27.86	1.17	0
P+U+S/N+A	0:06:49	0:06:35	0:13:24	12704	84	61	0.66	0.48	0
P+S/N+A	0:10:21	0:07:06	0:17:27	45405	34	35	0.07	0.08	0
U+S/N+A	0:10:36	0:07:01	0:17:37	40975	26	27	0.06	0.07	0
Fp+Fu+S/N+A	0:06:50	0:06:37	0:13:27	11538	86	51	0.75	0.44	0

Topografia									
C25.0 a C26.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.896	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:06:49	0:06:28	0:13:17	5215	33	41	0.63	0.79	0
P+S/N+DT	0:10:07	0:06:30	0:16:37	7836	15	24	0.19	0.31	0
U+S/N+DT	0:10:29	0:06:31	0:17:00	6873	89	19	1.29	0.28	0
Fp+Fu+S/N+DT	0:06:50	0:06:30	0:13:20	4696	33	35	0.70	0.75	0
P+U+S+A/N+DT	0:06:29	0:06:14	0:12:43	1159	542	8	46.76	0.69	0
P+S+A/N+DT	0:09:15	0:06:14	0:15:29	483	259	8	53.62	1.66	1
U+S+A/N+DT	0:09:43	0:06:14	0:15:57	353	202	7	57.22	1.98	0
Fp+Fu+S+A/N+DT	0:06:38	0:06:13	0:12:51	1105	543	6	49.14	0.54	0
Fp+S+A/N+DT	0:08:38	0:06:14	0:14:52	530	267	7	50.38	1.32	0
Fu+S+A/N+DT	0:09:22	0:06:15	0:15:37	348	207	5	59.48	1.44	0
P+U+S/N+A	0:06:50	0:06:36	0:13:26	15306	54	51	0.35	0.33	0
P+S/N+A	0:10:15	0:07:07	0:17:22	45940	17	19	0.04	0.04	0
U+S/N+A	0:10:32	0:07:01	0:17:33	42568	9	12	0.02	0.03	0
Fp+Fu+S/N+A	0:06:52	0:06:36	0:13:28	13194	57	42	0.43	0.32	0

Topografia									
C30.0 a C33.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 6.260	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:53	0:07:30	0:16:23	6558	251	315	3.83	4.80	6
P+S/N+DT	0:12:54	0:07:31	0:20:25	9222	85	128	0.92	1.39	5
U+S/N+DT	0:13:23	0:07:32	0:20:55	6079	81	118	1.33	1.94	1
Fp+Fu+S/N+DT	0:08:28	0:07:31	0:15:59	5657	277	274	4.90	4.84	7
P+U+S+A/N+DT	0:08:32	0:07:29	0:16:01	2310	1147	46	49.65	1.99	0
P+S+A/N+DT	0:12:30	0:07:13	0:19:43	424	199	23	46.93	5.42	1
U+S+A/N+DT	0:12:54	0:07:14	0:20:08	437	145	39	33.18	8.92	1
Fp+Fu+S+A/N+DT	0:08:20	0:07:32	0:15:52	2117	1157	31	54.65	1.46	0
Fp+S+A/N+DT	0:11:27	0:07:17	0:18:44	448	221	41	49.33	9.15	0
Fu+S+A/N+DT	0:12:10	0:07:16	0:19:26	437	162	64	37.07	14.65	2
P+U+S/N+A	0:09:06	0:07:48	0:16:54	18826	333	360	1.77	1.91	12
P+S/N+A	0:13:20	0:08:51	0:22:11	50258	101	101	0.20	0.20	1
U+S/N+A	0:13:21	0:08:14	0:21:35	43262	98	129	0.23	0.30	0
Fp+Fu+S/N+A	0:08:35	0:07:46	0:16:21	15327	361	327	2.36	2.13	8

Topografia									
C34.0 a C34.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 17.455	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:31	0:07:10	0:15:41	8245	339	556	4.11	6.74	107
P+S/N+DT	0:10:58	0:07:09	0:18:07	8775	42	73	0.48	0.83	15
U+S/N+DT	0:11:35	0:07:05	0:18:40	4914	52	113	1.06	2.30	29
Fp+Fu+S/N+DT	0:07:57	0:07:09	0:15:06	7406	415	561	5.60	7.57	191
P+U+S+A/N+DT	0:08:00	0:07:07	0:15:07	2972	1441	79	48.49	2.66	2
P+S+A/N+DT	0:11:26	0:06:53	0:18:19	403	163	21	40.45	5.21	1
U+S+A/N+DT	0:11:51	0:06:53	0:18:44	546	240	35	43.96	6.41	3
Fp+Fu+S+A/N+DT	0:07:42	0:07:10	0:14:52	2695	1476	66	54.77	2.45	1
Fp+S+A/N+DT	0:10:19	0:06:50	0:17:09	427	212	30	49.65	7.03	1
Fu+S+A/N+DT	0:10:54	0:06:52	0:17:46	615	292	54	47.48	8.78	4
P+U+S/N+A	0:08:24	0:07:15	0:15:39	16495	394	539	2.39	3.27	174
P+S/N+A	0:10:56	0:07:24	0:18:20	30511	46	58	0.15	0.19	12
U+S/N+A	0:12:12	0:07:15	0:19:27	18105	60	105	0.33	0.58	23
Fp+Fu+S/N+A	0:08:00	0:07:12	0:15:12	14036	478	473	3.41	3.37	104

Topografia									
C37.0 a C39.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 854	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:06:24	0:06:20	0:12:44	1976	22	21	1.11	1.06	0
P+S/N+DT	0:09:09	0:06:36	0:15:45	5541	17	17	0.31	0.31	0
U+S/N+DT	0:09:40	0:06:36	0:16:16	5679	12	12	0.21	0.21	0
Fp+Fu+S/N+DT	0:06:19	0:06:22	0:12:41	1788	22	20	1.23	1.12	0
P+U+S+A/N+DT	0:06:21	0:06:19	0:12:40	615	215	4	34.96	0.65	0
P+S+A/N+DT	0:08:27	0:06:19	0:14:46	342	102	3	29.82	0.88	0
U+S+A/N+DT	0:09:04	0:06:18	0:15:22	304	106	2	34.87	0.66	0
Fp+Fu+S+A/N+DT	0:06:19	0:06:23	0:12:42	580	215	3	37.07	0.52	0
Fp+S+A/N+DT	0:07:59	0:06:18	0:14:17	330	104	3	31.52	0.91	0
Fu+S+A/N+DT	0:08:42	0:06:20	0:15:02	296	108	3	36.49	1.01	0
P+U+S/N+A	0:06:26	0:06:35	0:13:01	7632	29	21	0.38	0.28	0
P+S/N+A	0:09:15	0:07:03	0:16:18	34052	17	16	0.05	0.05	0
U+S/N+A	0:09:44	0:06:56	0:16:40	29692	12	15	0.04	0.05	0
Fp+Fu+S/N+A	0:06:20	0:06:36	0:12:56	6827	29	19	0.42	0.28	0

Topografia									
C40.0 a C42.4	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 11.358	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:51	0:07:25	0:16:16	8578	425	574	4.95	6.69	81
P+S/N+DT	0:12:26	0:07:27	0:19:53	11600	89	162	0.77	1.40	56
U+S/N+DT	0:12:24	0:07:24	0:19:48	7040	99	157	1.41	2.23	50
Fp+Fu+S/N+DT	0:08:02	0:07:08	0:15:10	7555	500	552	6.62	7.31	154
P+U+S+A/N+DT	0:07:41	0:06:51	0:14:32	1936	979	56	50.57	2.89	9
P+S+A/N+DT	0:10:39	0:06:51	0:17:30	539	319	30	59.18	5.57	2
U+S+A/N+DT	0:10:40	0:06:50	0:17:30	260	64	22	24.62	8.46	0
Fp+Fu+S+A/N+DT	0:07:32	0:06:51	0:14:23	1728	1000	42	57.87	2.43	2
Fp+S+A/N+DT	0:10:27	0:06:50	0:17:17	628	384	38	61.15	6.05	3
Fu+S+A/N+DT	0:10:50	0:06:51	0:17:41	249	100	28	40.16	11.24	3
P+U+S/N+A	0:08:38	0:07:45	0:16:23	24723	426	556	1.72	2.25	146
P+S/N+A	0:12:31	0:08:10	0:20:41	66701	92	126	0.14	0.19	28
U+S/N+A	0:11:02	0:07:33	0:18:35	36610	82	173	0.22	0.47	66
Fp+Fu+S/N+A	0:08:11	0:07:20	0:15:31	19566	485	547	2.48	2.80	141

Topografia									
C44.0 a C42.4 (8720 a 8780)	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 5.034	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:10:31	0:07:35	0:18:06	4981	91	150	1.83	3.01	1
P+S/N+DT	0:15:46	0:07:57	0:23:43	11324	49	66	0.43	0.58	1
U+S/N+DT	0:05:16	0:08:41	0:13:57	9992	43	80	0.43	0.80	0
Fp+Fu+S/N+DT	0:10:40	0:07:42	0:18:22	4372	91	137	2.08	3.13	1
P+U+S+A/N+DT	0:11:33	0:07:41	0:19:14	2686	1195	65	44.49	2.42	0
P+S+A/N+DT	0:14:04	0:07:28	0:21:32	354	141	20	39.83	5.65	0
U+S+A/N+DT	0:14:31	0:07:21	0:21:52	377	128	20	33.95	5.31	0
Fp+Fu+S+A/N+DT	0:09:16	0:07:41	0:16:57	2348	1224	35	52.13	1.49	0
Fp+S+A/N+DT	0:13:16	0:07:15	0:20:31	372	172	19	46.24	5.11	0
Fu+S+A/N+DT	0:13:15	0:07:29	0:20:44	400	187	30	46.75	7.50	0
P+U+S/N+A	0:10:40	0:07:54	0:18:34	10069	111	142	1.10	1.41	0
P+S/N+A	0:15:11	0:08:12	0:23:23	42383	58	53	0.14	0.13	0
U+S/N+A	0:16:12	0:08:12	0:24:24	35148	52	81	0.15	0.23	0
Fp+Fu+S/N+A	0:10:05	0:07:40	0:17:45	8703	111	109	1.28	1.25	0

Topografia									
C44.0 a C44.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 70.662	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:14	0:07:18	0:16:32	17652	715	1032	4.05	5.85	18
P+S/N+DT	0:12:05	0:07:20	0:19:25	13424	95	123	0.71	0.92	5
U+S/N+DT	0:12:12	0:07:19	0:19:31	13702	151	222	1.10	1.62	1
Fp+Fu+S/N+DT	0:08:26	0:07:18	0:15:44	15670	808	835	5.16	5.33	7
P+U+S+A/N+DT	0:09:02	0:07:12	0:16:14	9897	4178	142	42.21	1.43	0
P+S+A/N+DT	0:11:59	0:06:52	0:18:51	775	397	22	51.23	2.84	0
U+S+A/N+DT	0:12:11	0:06:52	0:19:03	445	144	44	32.36	9.89	0
Fp+Fu+S+A/N+DT	0:08:47	0:07:18	0:16:05	9291	4821	103	51.89	1.11	0
Fp+S+A/N+DT	0:11:10	0:06:58	0:18:08	977	517	43	52.92	4.40	0
Fu+S+A/N+DT	0:11:20	0:06:59	0:18:19	893	464	277	51.96	31.02	5
P+U+S/N+A	0:09:21	0:07:42	0:17:03	41125	907	1153	2.21	2.80	204
P+S/N+A	0:12:11	0:08:16	0:20:27	50056	89	342	0.18	0.68	67
U+S/N+A	0:12:17	0:07:52	0:20:09	47029	150	496	0.32	1.05	18
Fp+Fu+S/N+A	0:08:35	0:07:00	0:15:35	31923	1045	957	3.27	3.00	154

Topografia									
C47.0 a C49.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 2.830	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:36	0:07:03	0:14:39	4167	126	174	3.02	4.18	8
P+S/N+DT	0:11:40	0:07:07	0:18:47	6867	42	65	0.61	0.95	1
U+S/N+DT	0:11:10	0:07:05	0:18:15	4210	44	62	1.05	1.47	2
Fp+Fu+S/N+DT	0:07:20	0:07:05	0:14:25	3557	128	139	3.60	3.91	3
P+U+S+A/N+DT	0:07:07	0:06:50	0:13:57	1518	797	37	52.50	2.44	0
P+S+A/N+DT	0:10:42	0:06:50	0:17:32	173	52	19	30.06	10.98	0
U+S+A/N+DT	0:10:57	0:06:49	0:17:46	338	116	29	34.32	8.58	2
Fp+Fu+S+A/N+DT	0:07:15	0:06:48	0:14:03	1432	815	28	56.91	1.96	1
Fp+S+A/N+DT	0:10:02	0:06:47	0:16:49	171	63	24	36.84	14.04	2
Fu+S+A/N+DT	0:10:33	0:06:46	0:17:19	385	143	23	37.14	5.97	1
P+U+S/N+A	0:07:28	0:07:12	0:14:40	14757	185	231	1.25	1.57	32
P+S/N+A	0:10:46	0:07:45	0:18:31	51530	59	97	0.11	0.19	31
U+S/N+A	0:10:59	0:07:27	0:18:26	30633	54	62	0.18	0.20	0
Fp+Fu+S/N+A	0:07:22	0:07:09	0:14:31	12069	188	205	1.56	1.70	60

Topografia									
C50.0 a C50.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 41.307	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:36	0:07:46	0:17:22	16069	2683	3410	16.70	21.28	138
P+S/N+DT	0:12:58	0:07:44	0:20:42	11426	276	279	2.42	2.44	107
U+S/N+DT	0:13:16	0:07:47	0:21:03	12702	354	415	2.79	3.27	9
Fp+Fu+S/N+DT	0:08:58	0:07:44	0:16:42	13819	3145	2817	22.76	20.38	57
P+U+S+A/N+DT	0:09:02	0:07:40	0:16:42	5347	2611	167	48.83	3.12	1
P+S+A/N+DT	0:12:21	0:07:26	0:19:47	533	280	49	52.53	9.19	1
U+S+A/N+DT	0:12:52	0:07:27	0:20:19	723	337	83	46.61	11.48	0
Fp+Fu+S+A/N+DT	0:08:43	0:07:41	0:16:24	4761	2863	222	60.13	4.66	17
Fp+S+A/N+DT	0:11:11	0:07:28	0:18:39	772	46	98	5.96	12.69	1
Fu+S+A/N+DT	0:12:06	0:07:26	0:19:32	1074	551	162	51.30	15.08	0
P+U+S/N+A	0:12:07	0:08:00	0:20:07	31882	2962	2695	9.29	8.45	342
P+S/N+A	0:13:04	0:08:06	0:21:10	44139	314	281	0.71	0.64	5
U+S/N+A	0:13:25	0:08:06	0:21:31	48192	389	388	0.81	0.81	4
Fp+Fu+S/N+A	0:09:04	0:07:50	0:16:54	25199	3466	2668	13.75	10.59	49

Topografia									
C51.0 a C58.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 28.359	relaciona	combina	minutos	linkado	no escore máxim o	par verdade	no escore máximo	sucesso	duplicidades
P+U/N+DT	0:09:17	6:13:35	6:22:52	297460	1742	1843	0.59	0.62	104
Fp+Fu/N+DT	0:21:43	8:42:59	9:04:42	235095	1742	1637	0.74	0.70	29
P+U+A/N+DT	0:08:01	7:49:00	7:57:01	48192	1742	1834	3.61	3.81	101
P+A/N+DT*	0:20:37	12:15:20	12:35:57	1292276	1742	1582	0.13	0.12	60
U+A/N+DT*	0:20:19	16:12:48	16:33:07	1127284	1742	1559	0.15	0.14	68
Fp+Fu+A/N+DT	0:08:15	0:09:06	0:17:21	234875	1742	1445	0.74	0.62	24
Fp+A/N+DT	0:18:31	11:14:36	11:33:07	953949	1742	1561	0.18	0.16	32
Fu+A/N+DT	0:19:34	6:14:36	6:34:10	895252	1742	1534	0.19	0.17	52
P+U/N+A**	0:26:24	0:34:47	1:01:11	1221647	2912	1376	0.24	0.11	61
Fp+Fu/N+A*	0:26:02	12:23:50	12:49:52	1291497	2912	1536	0.23	0.12	18

* 24 horas

** 48 horas

Topografia									
C64.0 a C68.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 12.343	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:43	0:06:35	0:14:18	8344	288	309	3.45	3.70	2
P+S/N+DT	0:11:34	0:06:38	0:18:12	9479	77	89	0.81	0.94	1
U+S/N+DT	0:11:39	0:06:36	0:18:15	7829	69	103	0.88	1.32	1
Fp+Fu+S/N+DT	0:07:17	0:06:34	0:13:51	7139	305	256	4.27	3.59	3
P+U+S+A/N+DT	0:07:19	0:06:36	0:13:55	3234	1476	31	45.64	0.96	0
P+S+A/N+DT	0:11:16	0:06:17	0:17:33	477	129	14	27.04	2.94	2
U+S+A/N+DT	0:11:26	0:06:18	0:17:44	439	198	20	45.10	4.56	0
Fp+Fu+S+A/N+DT	0:07:02	0:06:31	0:13:33	2870	1568	26	54.63	0.91	0
Fp+S+A/N+DT	0:10:05	0:06:19	0:16:24	473	168	20	35.52	4.23	0
Fu+S+A/N+DT	0:10:26	0:06:17	0:16:43	451	259	28	57.43	6.21	0
P+U+S/N+A	0:07:50	0:06:49	0:14:39	23845	383	308	1.61	1.29	0
P+S/N+A	0:11:55	0:07:30	0:19:25	64318	89	90	0.14	0.14	0
U+S/N+A	0:11:51	0:07:15	0:19:06	45684	80	117	0.18	0.26	1
Fp+Fu+S/N+A	0:07:23	0:06:52	0:14:15	18490	402	227	2.17	1.23	0

Topografia									
C69.0 a C72.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 7.199	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:09:04	0:07:35	0:16:39	6873	204	276	2.97	4.02	8
P+S/N+DT	0:13:31	0:07:38	0:21:09	10508	40	50	0.38	0.48	0
U+S/N+DT	0:13:26	0:07:38	0:21:04	7758	50	67	0.64	0.86	1
Fp+Fu+S/N+DT	0:08:50	0:07:35	0:16:25	5836	221	223	3.79	3.82	6
P+U+S+A/N+DT	0:08:32	0:07:18	0:15:50	1668	687	55	41.19	3.30	0
P+S+A/N+DT	0:12:44	0:07:17	0:20:01	414	166	22	40.10	5.31	0
U+S+A/N+DT	0:12:59	0:07:17	0:20:16	259	70	21	27.03	8.11	0
Fp+Fu+S+A/N+DT	0:08:15	0:07:19	0:15:34	1576	715	43	45.37	2.73	0
Fp+S+A/N+DT	0:11:33	0:07:18	0:18:51	463	194	24	41.90	5.18	0
Fu+S+A/N+DT	0:12:00	0:07:17	0:19:17	281	87	23	30.96	8.19	0
P+U+S/N+A	0:09:10	0:07:46	0:16:56	21205	230	232	1.08	1.09	4
P+S/N+A	0:13:46	0:08:29	0:22:15	60376	50	50	0.08	0.08	1
U+S/N+A	0:13:38	0:08:06	0:21:44	40056	52	71	0.13	0.18	1
Fp+Fu+S/N+A	0:08:46	0:07:45	0:16:31	17189	247	187	1.44	1.09	1

Topografia									
C73.0 a C75.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 11.696	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:07:51	0:06:41	0:14:32	6219	182	210	2.93	3.38	1
P+S/N+DT	0:11:32	0:06:42	0:18:14	7992	28	52	0.35	0.65	0
U+S/N+DT	0:11:40	0:06:42	0:18:22	6669	40	55	0.60	0.82	0
Fp+Fu+S/N+DT	0:07:20	0:06:43	0:14:03	5272	211	183	4.00	3.47	0
P+U+S+A/N+DT	0:07:33	0:06:40	0:14:13	2181	861	32	39.48	1.47	0
P+S+A/N+DT	0:11:16	0:06:26	0:17:42	387	85	16	21.96	4.13	0
U+S+A/N+DT	0:11:23	0:06:24	0:17:47	217	61	19	28.11	8.76	0
Fp+Fu+S+A/N+DT	0:07:13	0:06:22	0:13:35	2042	951	65	46.57	3.18	0
Fp+S+A/N+DT	0:10:19	0:06:22	0:16:41	470	139	25	29.57	5.32	0
Fu+S+A/N+DT	0:10:23	0:06:22	0:16:45	249	83	23	33.33	9.24	0
P+U+S/N+A	0:07:58	0:06:50	0:14:48	21537	253	273	1.17	1.27	2
P+S/N+A	0:11:42	0:07:20	0:19:02	52082	37	52	0.07	0.10	0
U+S/N+A	0:11:50	0:07:09	0:18:59	38821	54	56	0.14	0.14	0
Fp+Fu+S/N+A	0:07:26	0:06:47	0:14:13	15792	287	238	1.82	1.51	1

Topografia									
C76.0 a C80.9	tempo (hh:mm:ss)		Total	n°	n de pares		% de pares	%	
n= 23.720	relaciona	combina	minutos	linkado	no escore máximo	par verdade	no escore máximo	sucesso	duplicidades
P+U+S/N+DT	0:08:36	0:06:42	0:15:18	10350	726	613	7.01	5.92	4
P+S/N+DT	0:11:52	0:06:45	0:18:37	10160	125	136	1.23	1.34	1
U+S/N+DT	0:11:50	0:06:45	0:18:35	8386	119	175	1.42	2.09	3
Fp+Fu+S/N+DT	0:07:48	0:06:46	0:14:34	9896	805	599	8.13	6.05	4
P+U+S+A/N+DT	0:07:50	0:06:41	0:14:31	5586	2479	95	44.38	1.70	1
P+S+A/N+DT	0:11:35	0:06:23	0:17:58	621	241	39	38.81	6.28	7
U+S+A/N+DT	0:11:41	0:06:27	0:18:08	427	137	41	32.08	9.60	0
Fp+Fu+S+A/N+DT	0:07:29	0:06:43	0:14:12	5050	2616	67	51.80	1.33	0
Fp+S+A/N+DT	0:10:40	0:06:27	0:17:07	706	320	47	45.33	6.66	0
Fu+S+A/N+DT	0:10:49	0:06:27	0:17:16	542	203	55	37.45	10.15	0
P+U+S/N+A	0:08:30	0:07:04	0:15:34	30783	729	579	2.37	1.88	2
P+S/N+A	0:12:30	0:07:33	0:20:03	56274	119	110	0.21	0.20	0
U+S/N+A	0:12:00	0:07:10	0:19:10	40985	126	152	0.31	0.37	2
Fp+Fu+S/N+A	0:07:53	0:06:59	0:14:52	24050	804	586	3.34	2.44	6

Anexo 8 - Aprovação do Comitê de Ética em Pesquisa (Faculdade de Saúde Pública/USP)



COMITÊ DE ÉTICA EM PESQUISA – COEP/FSP
 Universidade de São Paulo
 Faculdade de Saúde Pública

Of.COEP/ 274 / 08

Protocolo	1864
Projeto de Pesquisa	USO DA TÉCNICA DE LINKAGE NOS SISTEMAS DE INFORMAÇÕES EM SAÚDE: APLICAÇÃO NA BASE DE DADOS DO REGISTRO DE CÂNCER DE BASE POPULACIONAL DO MUNICÍPIO DE SÃO PAULO
Pesquisador(a)	Stela Verzinhasse Peres

24 de OUTUBRO de 2008.

Prezado(a) Orientador(a),

O Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da Universidade de São Paulo - COEP analisou, em sua 9.ª/08 Sessão ORDINÁRIA, realizada em 17/10/2008, de acordo com os requisitos da Resolução CNS/196/96 e suas complementares, o protocolo de pesquisa acima intitulada e o considerou **APROVADO**.

Cabe lembrar que conforme Resolução CNS/196/96 são deveres do (a) pesquisador (a):

1. Comunicar, de imediato, qualquer alteração no projeto e aguardar manifestação deste CEP (Comitê de Ética em Pesquisa), para dar continuidade à pesquisa;
2. Manter sob sua guarda e em local seguro, pelo prazo de 5 (cinco) anos, os dados da pesquisa, contendo fichas individuais e todos os demais documentos recomendados pelo CEP, no caso eventual auditoria;
3. Comunicar, formalmente a este Comitê, quando do encerramento deste projeto;
4. Elaborar e apresentar relatórios parciais e final;
5. Justificar, perante o CEP, interrupção do projeto ou a não publicação dos resultados.

Atenciosamente,

Cláudio Leone

Professor Associado
 Coordenador do Comitê de Ética em Pesquisa – FSP/COEP

Ilm.ª Sr.ª
 Prof.ª Dr.ª MARIA DO ROSÁRIO DIAS DE OLIVEIRA LATORRE
 DEPARTAMENTO DE EPIDEMIOLOGIA
 FACULDADE DE SAÚDE PÚBLICA
 UNIVERSIDADE DE SÃO PAULO

Anexo 9 - Aprovação do Comitê de Ética em Pesquisa (Secretaria Municipal da Saúde de São Paulo)



Secretaria Municipal da Saúde
Comitê de Ética em Pesquisa - CEP/SMS

1

São Paulo, 26 de março de 2009.
PARECER Nº 111/09 – CEP/SMS
CAAE: 0064.0.162.000-09

Para
STELA VERZINHASSE PERES
Sr(a) Pesquisadora

- **Projeto de Pesquisa: USO DA TÉCNICA DE LINKAGE NOS SISTEMAS DE INFORMAÇÃO SM SAÚDE: APLICAÇÃO NA BASE DE DADOS DO REGISTRO DE CANCER DE BASE POPULACIONAL DO MUNICÍPIO DE SÃO PAULO**
- **Pesquisador Responsável: STELA VERZINHASSE PERES**
- **Instituição: FACULDADE DE SAÚDE PÚBLICA DA UNIVERSIDADE DE SÃO PAULO**
- **Local onde os dados serão coletados: CEINFO/SMS.G**

I - Sumário Geral do Protocolo

Objetivo Geral - Avaliar as vantagens e limitações da utilização do linkage da base de dados gerada pelo RCBP de São Paulo com algumas das bases existentes, a saber: SIM, Apac-SIA/SUS e TRE (Tribunal Regional Eleitoral), visando à qualidade da informação sobre o status vital dos pacientes;

Objetivos Específicos

- testar a sensibilidade e a especificidade quanto a capacidade de identificação de duplicidades entre os softwares Reclink versão 3.1.6, SPSS para Windows versão 15.0 e STATA intercooled versão 9.0;
- calcular as taxas de incidência e de mortalidade por câncer, segundo tipo de tumor dos pacientes registrados na base de dados do RCBP de São Paulo, antes e depois do processo de relacionamento;
- analisar a sobrevivência global e específica por câncer dos pacientes registrados na base de dados do RCBP de São Paulo, antes e depois do processo de relacionamento.

As bases de dados utilizadas neste estudo serão fornecidas pelo RCBP de São Paulo, pela Fundação SEADE, PRO-AIM, Apac-SIA/SUS e Tribunal Regional Eleitoral (TRE).

Registro de Câncer de Base Populacional de São Paulo - a base de dados fornecida pelo RCBP de São Paulo compreenderá os anos de 1997 a 2007, totalizando, aproximadamente, 325000 casos incidentes de câncer notificados no município de São Paulo.

A primeira etapa desta pesquisa consistirá em verificar a existência de entradas duplas para um mesmo paciente na base de dados, ou seja, registro duplicado, ou a notificação de um segundo tumor primário, ou a ocorrência de metástase. Para as duplicidades, serão eliminados os casos excedentes. Serão excluídos também, os casos de segundo tumor primário. Após este procedimento, esta base será dividida em sub-bases, para otimização do relacionamento, de acordo com a topografia.

Sistema de Informação sobre Mortalidade - O SIM é um sistema de informação, implantado pelo MS, que gerencia as informações de forma descentralizada. No município de São Paulo a gerência dos óbitos é feita pelo PRO-AIM, esta instituição faz o processamento, a análise e divulgação dos dados de mortalidade no nível municipal. No Estado de São Paulo, o órgão responsável pela base de dados do SIM é a Fundação SEADE. Esta instituição recebe todos os dados dos sistemas municipais, controlando o envio, agregando e classificando as informações, segundo o município de residência (Fundação SEADE, 2008). Será solicitado ao PRO-AIM e a Fundação SEADE o banco de dados dos óbitos referente ao período de 1997 a 2008. A primeira etapa a ser realizada, será identificar as causas de óbitos e excluir as perdas fetais das bases de dados. O instrumento de coleta é a DO (Anexo 2).



Secretaria Municipal da Saúde
Comitê de Ética em Pesquisa - CEP/SMS

2

CAAE: 0064.0.162.000-09

Apac – SIA/SUS – a base de dados da Apac, implantada em 1996 pelo MS, é um aplicativo do SIA/SUS que dispõe do banco de dados do tipo relacional composto por um conjunto de tabelas que viabilizam o controle administrativo do atendimento ambulatorial. Desta base, serão fornecidos os dados sobre pacientes com câncer que estejam em tratamento contínuo (principais ou secundários) e que necessitam de procedimentos ambulatoriais de alta complexibilidade e/ou de alto custo (MS, 2008). Da base de dados da Apac-SIA/SUS.

Tribunal Regional Eleitoral – serão selecionados todos os cidadãos que participaram ou justificaram ausência do voto na última eleição.

Estudo Piloto

Para melhor execução do processo de linkage, será realizado um estudo piloto entre os casos da topografia C34, correspondente as localizações de brônquios e pulmões registrados na base do RCBP de São Paulo e as bases do SIM, Apac-SIA e TRE.

A finalidade do estudo piloto é definir de maneira clara a melhor estratégia de execução do projeto. Nesta etapa serão definidos os valores de escores e os pontos de corte para a identificação dos pares verdadeiros e falsos, e quantos processos de bloqueio serão necessários para uma melhor efetividade de relacionamento.

II. Considerações

A pesquisadora acessará as bases de dados do PRO-AIM e do SIA-APAC junto ao CEInfo e CIRS, em SMS.

Folha de rosto adequada e assinada pelo coordenador do PRO-AIM.

Os currículos apresentados, tanto da pesquisadora como das orientadoras, estão a contento.

No anexo 6, consta o orçamento e referência de que não há financiamento por agência de fomento e que os gastos descritos serão arcados pela pesquisadora.

O cronograma está adequado, com previsão de análise dos dados no primeiro semestre de 2009.

As bases em questão contêm dados de identificação dos pacientes, sendo os mesmos intrinsecamente necessários para a metodologia do estudo. A pesquisadora refere na página 34 do projeto de pesquisa que somente a pesquisadora principal e suas orientadoras terão acesso à totalidade dos dados do banco e a manipulação dos dados será feita somente pela pesquisadora principal. Após o relacionamento, para as análises, o nome, o nome da mãe, CPF e datas de nascimento, de óbito e da última eleição serão excluídos. São esperados, aproximadamente, 380.000 registros de pacientes com câncer e os pesquisadores se comprometem a manter em sigilo toda informação recebida.

No anexo 7, consta a seguinte DECLARAÇÃO DE ISENÇÃO DO TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO:

Declaro que a pesquisa "Uso de técnica de linkage nos sistemas de informação em saúde: aplicação na base de dados do Registro de Câncer de Base Populacional do município de São Paulo" fará uso de base de dados fornecida pelo Registro de Câncer de Base Populacional do município de São Paulo o que impossibilita a aplicação do Termo de Consentimento Livre e Esclarecido.

III. Parecer do CEP: Projeto APROVADO

Visto que o CEP teve alterações no seu endereço, solicitamos atenção para a necessidade de retificação do mesmo para: Comitê de Ética em Pesquisas da Secretaria Municipal da Saúde de São Paulo - Rua General Jardim, 36 - 8º andar - Telefone: 33972464 - e-mail: smscep@gmail.com.



Secretaria Municipal da Saúde
Comitê de Ética em Pesquisa - CEP/SMS

3

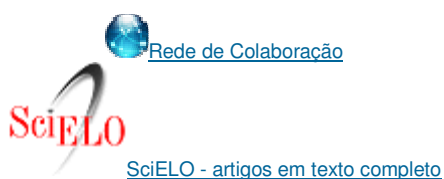
CAAE: 0064.0.162.000-09

Antes do início da coleta de dados, alertamos para a necessidade de contato com o gerente da unidade quando não foi ele quem autorizou a realização da pesquisa.

Salientamos que o pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado. Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. O relatório final deve ser apresentado ao CEP, logo que o estudo estiver concluído.

Jose Araújo Lima Filho
Coordenador
Comitê de Ética em Pesquisa da
Secretaria Municipal da Saúde – CEP/SMS

CURRÍCULO LATTES - ALUNO



Stela Verzinhasse Peres

Doutoranda em Saúde Pública, área de concentração em Epidemiologia pela Faculdade de Saúde Pública da Universidade de São Paulo, mestrado em Saúde Pública pela Faculdade de Saúde Pública da Universidade de São Paulo (2007), aperfeiçoamento técnico nível III em Nutrição pela Faculdade de Saúde Pública (2005) e Bacharel em Educação Física pela Universidade São Judas Tadeu - SP - (2004). Tem experiência na área de Epidemiologia, Bioestatística e Estatística de Saúde, atuando principalmente nos seguintes temas: Sistemas de Informação, Fatores Associados e Sobrevida de Pacientes com Câncer. **(Texto informado pelo autor)**

Última atualização do currículo em 09/05/2011

Endereço para acessar este CV:

<http://lattes.cnpq.br/7821332537772230>



Dados pessoais

Nome	Stela Verzinhasse Peres 
Nome em citações bibliográficas	PERES, Stela Verzinhasse; Rocha, Euclides Timóteo da
Sexo	Feminino
Endereço profissional	Faculdade de Saúde Pública da Universidade de São Paulo, Faculdade de Saúde Pública da Universidade de São Paulo. Av. Doutor Arnaldo, 175 2º andar Epidemiologia Cerqueira Cesar 01246-904 - Sao Paulo, SP - Brasil Telefone: (11) 30617744 Ramal: 118 URL da Homepage: http://

CURRÍCULO LATTES - ORIENTADOR



[SciELO - artigos em texto completo](#)



Maria do Rosario Dias de Oliveira Latorre

possui graduação em Bacharel em Estatística pela Universidade de São Paulo (1978) , especialização em Administração Hospitalar Para Graduados pela Faculdade de Saúde Pública Universidade de São Paulo (1983) , mestrado em Saúde Pública pela Universidade de São Paulo (1992) , doutorado em Saúde Pública pela Universidade de São Paulo (1996) , aperfeiçoamento em Conceitos Básicos Introdução Ao Sistema Ibm 360 Co pela Cursos Técnicos de Processamento de Dados (1974) , aperfeiçoamento em Treinamento Para a Preparação de Orientadores de Apr pelo Centro de Integração Empresa Escola (1977) , aperfeiçoamento em Elementos de Consultoria Estatística pelo Associação Profissional dos Estatísticos do Brasil (1978) , aperfeiçoamento em Métodos e Técnicas de Pesquisa Em Ciências Sociais pelo Associação dos Sociólogos do Estado de São Paulo (1979) , aperfeiçoamento em Planejamento e análise estatística em pesquisa pelo Associação Profissional dos Estatísticos do Brasil (1979) , aperfeiçoamento em Técnicas de Análise Demográfica pelo Instituto de Matemática e Estatística Usp (1980) , aperfeiçoamento em Relações Humanas do Trabalho pelo Serviço Nacional de Aprendizado Industrial (1980) , aperfeiçoamento em Elaboração e Prática de Estatística Hospitalar pela Faculdade de Administração Hospitalar do Iph (1980) , aperfeiçoamento em Curso de Treinamento no Uso da Classificação Inter pela Faculdade de Saúde Pública (1980) , aperfeiçoamento em Contabilidade Hospitalar pelo Associação Paulista de Hospitais (1981) , aperfeiçoamento em A Informática no Hospital pelo Associação Paulista de Hospitais (1984) , aperfeiçoamento em Métodos no Trabalho pelo Serviço Nacional de Aprendizado Industrial (1980) , aperfeiçoamento em Ensino Correto de Um Trabalho pelo Serviço Nacional de Aprendizado Industrial (1980) , aperfeiçoamento em Curso de Atuária pelo Associação Brasileira das Entidades Fechadas de Previdência Privada (1986) , aperfeiçoamento em Metodologia Epidemiológica II pela Faculdade de Saúde Pública (1990) , aperfeiçoamento em Classificação e codificação de neoplasias pela Faculdade de Saúde Pública (1991) , aperfeiçoamento em Planejamento de pesquisa clínico-epidemiológica pela Faculdade de Medicina da Universidade Federal do Rio Grande do Sul (1992) , aperfeiçoamento em Logistic regression methods for epidemiology pela Faculdade de Saúde Pública (1992) , aperfeiçoamento em Theory and practice of Epidemiology II pela Tufts University (1994) , aperfeiçoamento em Regression methods in Epidemiology pela Tufts University (1994) , aperfeiçoamento em Survival analysis in Epidemiology pela Tufts University (1994) , aperfeiçoamento em Conducting epidemiologic research pela Tufts University (1994) , aperfeiçoamento em Programa de Verão pela Division Of Epidemiology Faculty Of Medicine (1995) , aperfeiçoamento em Epidemiology of cancer pela Division Of Epidemiology Faculty Of Medicine (1995) , aperfeiçoamento em Ethics in scientific inference pela Division Of Epidemiology Faculty Of Medicine (1995) , aperfeiçoamento em Métodos Epidemiológicos Sobre Aids pelo Fundação Oswaldo Cruz (1989) , aperfeiçoamento em Molecular Biology For Genetic Epidemiology pela Johns Hopkins University (2002) , aperfeiçoamento em Family Based Genetic Epidemiology pela Johns Hopkins University (2002) e aperfeiçoamento em Sobrevidas Observada e Relativa Em Estudos Populac pelo Fundação Oswaldo Cruz (2003) . Atualmente é professor titular da Universidade de São Paulo, sem vínculo empregatício do Fundação Antônio Prudente e Membro de corpo editorial da Revista de Saúde Pública / Journal of Public Health. Tem experiência na área de Probabilidade e Estatística , com ênfase em

Probabilidade e Estatística Aplicadas. Atuando principalmente nos seguintes temas: cancer, incidência, mortalidade, tendência.
(Texto gerado automaticamente pela aplicação CVLattes)

Última atualização do currículo em 18/08/2011

Endereço para acessar este CV:
<http://lattes.cnpq.br/5543326971362158>



Dados pessoais

Nome Maria do Rosario Dias de Oliveira Latorre

Nome em citações bibliográficas LATORRE MRDO;Latorre, Maria do Rosário Dias de Oliveira;Latorre, Maria do Rosário Dias O

Sexo Feminino

Endereço profissional Universidade de São Paulo, Faculdade de Saúde Pública, Departamento de Epidemiologia.
Av. Dr. Arnaldo, 715
Cerqueira Cesar
01246904 - Sao Paulo, SP - Brasil
Telefone: (11) 30667744 Fax: (11) 32822920

