


Concordância e validade dos diagnósticos de cânceres associados ao asbesto no sistema de informação hospitalar do Sistema Único de Saúde

Agreement and validity of asbestos-related cancers in the hospital information system of the Brazilian Unified Health System

Diego Rodrigues Mendonça e Silva^{III} , Carolina Terra de Moraes Luizaga^{III} ,
Tatiana Natasha Toporcov^{IV} , Eduardo Algranti^V 

RESUMO: *Objetivo:* Estimar o grau de concordância e validade dos diagnósticos de neoplasias malignas relacionadas à exposição ao asbesto registrados no Sistema de Informação Hospitalar do Sistema Único de Saúde (SIH/SUS), em comparação aos Registros Hospitalares de Câncer do Estado de São Paulo (RHC/SP). *Métodos:* Óbitos com registros de neoplasias malignas associadas ao asbesto foram identificados e extraídos do SIH/SUS entre 2007 e 2014. Óbitos nos casos de câncer registrados na base do RHC/SP foram extraídos para o mesmo período. Essas bases foram unidas pelos mesmos indivíduos empregando-se o software *Link Plus*. Um único diagnóstico codificado pela CID-O3 ou CID-10 selecionado de cada sistema foi analisado. A proporção de concordância e a sensibilidade, especificidade e valores preditivos foram estimados. *Resultados:* Encontraram-se 19.458 pares com registros nas duas bases. A proporção de concordância foi elevada, variando de 92,4% para a localização primária desconhecida a 99,7% para o câncer de pleura. O índice Kappa variou de 0,05 (IC95% 0,04 – 0,07) para o câncer de pleura a 0,85 (IC95% 0,84 – 0,87) para o câncer de pulmão. A menor sensibilidade foi de 0,08 (IC95% 0,01 – 0,25), para o câncer de pleura, e a maior de 0,90 (IC95% 0,90 – 0,91), para o câncer de pulmão. *Conclusão:* Diagnósticos de neoplasias malignas associadas ao asbesto alcançaram maiores níveis de concordância e validade quando comuns. Os diagnósticos mais raros apresentaram baixa acurácia no SIH/SUS.

Palavras-chave: Interoperabilidade da informação em saúde. Asbestos. Neoplasias. Armazenamento e recuperação da informação.

^ICentro Internacional de Pesquisa, A. C. Camargo Cancer Center – São Paulo (SP), Brasil.

^{II}Programa de Pós-Graduação em Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo – São Paulo (SP), Brasil.

^{III}Diretoria de Informação e Epidemiologia, Fundação Oncocentro de São Paulo – São Paulo (SP), Brasil.

^{IV}Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo – São Paulo (SP), Brasil.

^VDiretoria de Pesquisa Aplicada, Fundação Jorge Duprat Figueiredo, Ministério da Economia – São Paulo (SP), Brasil.

Autor correspondente: Diego Rodrigues Mendonça e Silva. Centro Internacional de Pesquisa, A. C. Camargo Cancer Center. Rua Taguá, 440, Liberdade, CEP: 01508-010, São Paulo, SP, Brasil. E-mail: diego.rmesilva@gmail.com

Conflito de interesses: nada a declarar – **Fonte de financiamento:** Ministério Público do Trabalho da 15ª Região, Escritório de Campinas, São Paulo (CODIN No. 45672/2014).

ABSTRACT: *Objective:* To estimate the degree of agreement and validity of diagnoses of asbestos-related malignant neoplasms registered in the Hospital Information System of the Brazilian Unified Health System (SIH/SUS), in comparison to the Hospital Cancer Registries of the State of São Paulo (HCR/SP). *Methods:* Deaths with records of malignant neoplasms associated with asbestos were identified and extracted from SIH/SUS between 2007 and 2014. Deaths in cases registered in the HCR/SP were extracted for the same period. The databases were linked using software *Link Plus*. A single ICD-10-coded diagnosis selected from each system was analyzed. The proportion of agreement, and the sensitivity, specificity and predictive values were estimated. *Results:* 19,458 pairs were found with records in both bases. The proportion of agreement was high, ranging from 92.4% for the unknown primary site, to 99.7% for cancer of the pleura. The Kappa Index ranged from 0.05 (95%CI 0.04 – 0.07) for cancer of the pleura to 0.85 (95%CI 0.84 – 0.87) for lung cancer. Sensitivity varied from 0.08 (95%CI 0.01 – 0.25) for cancer of the pleura, to 0.90 (95%CI 0.90 – 0.91) for lung cancer. *Conclusion:* Diagnosis of asbestos-related malignancies reached higher levels of agreement and validity when common. Rare diagnoses showed low accuracy in SIH/SUS. *Keywords:* Health information interoperability. Asbestos. Neoplasms. Information storage and retrieval.

INTRODUÇÃO

O monitoramento de indicadores epidemiológicos de doenças de relevância para a saúde pública é um dos pilares da vigilância em saúde, permitindo estimar a extensão da morbidade e mortalidade, a carga, as variações temporais e distribuições espaciais, entre outras, contribuindo para as respostas de controle, prevenção e melhoria da saúde da população. No Brasil, vários sistemas de informação em saúde (SIS) foram aprimorados e vêm consolidando-se desde a criação do Sistema Único de Saúde (SUS)¹⁻³, sendo disponibilizados por meio do DATASUS.

O Sistema de Informação Hospitalar (SIH/SUS) é um SIS que registra dados de admissões em serviços de internação e destina-se à gestão, especialmente ao reembolso de despesas em hospitais públicos, filantrópicos e privados, quando conveniados ou contratados pelo SUS. Por abranger 60 a 70% do total de internações hospitalares no país⁴, o SIH/SUS passou a ser considerado de interesse também para a vigilância em saúde, ao apresentar registros de diagnósticos codificados de acordo com a Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID)⁵.

Em estudo conduzido como parte do projeto de Exposição Ocupacional ao Asbesto e seus Efeitos sobre a Saúde no Brasil, Santana et al.⁶, ao buscarem casos de óbitos por mesotelioma e/ou câncer de pleura no SIH/SUS, encontraram dados que permitiram estimar acordos diagnósticos em 70% dos pares. A vinculação entre o SIH/SUS e o Sistema de Informação sobre Mortalidade (SIM) aumentou em 32,2% os registros de mesotelioma e/ou câncer de pleura entre 2002-2012 no estado de São Paulo, os registros recuperados a partir do SIH/SUS que não foram identificados no SIM⁶.

O Registro Hospitalar de Câncer (RHC) e o Registro de Câncer de Base Populacional (RCBP) são as principais bases que compõem o Sistema de Informação do Câncer (SISCAN)

do Instituto Nacional do Câncer (INCA/SUS), destinados à promoção da prevenção, assistência e pesquisa em câncer no país^{3,7}. A fundamentação do registro diagnóstico de câncer é confirmada por meio de exames anatomopatológicos em mais de 90% dos casos⁷⁻⁹.

Estima-se que exposições a cancerígenos no ambiente de trabalho são responsáveis por 5 a 8% da carga de cânceres em homens, notadamente o câncer de pulmão. Entre os cancerígenos ocupacionais, o asbesto responde por mais de 50% dos cânceres ocupacionais de pulmão¹⁰. O asbesto, em todas as suas variedades, é classificado pela Agência Internacional de Pesquisa em Câncer (IARC) no Grupo 1, agente causal e/ou concausal no mesotelioma, câncer de pulmão, de laringe e de ovário¹¹, e possivelmente associado a outros cânceres de cabeça e pescoço, faringe, esôfago, estômago e colorretal¹². Em relação ao mesotelioma, a transição da CID-9 para a CID-10 aumentou expressivamente os registros de causa básica de morte por mesotelioma, pela existência de uma codificação específica para a doença¹³. A utilização de registros de câncer mostrou-se uma fonte válida de diagnósticos de mesotelioma, quando comparados com registros clínicos, de imagem e anatomopatológicos¹⁴.

Pela sua abrangência, o SIH/SUS é uma fonte de busca de registros de doenças, portanto de possíveis casos subnotificados de doenças relacionadas ao asbesto. Seus registros de diagnósticos e procedimentos clínicos, entretanto, podem apresentar problemas, o que requer avaliações de sua acurácia, especialmente nas análises epidemiológicas, seja para a vigilância ou pesquisas específicas^{1-2,4,15-18}. Com vistas ao uso do SIH/SUS como fonte de dados de neoplasias malignas associadas ao asbesto, este estudo tem como objetivo estimar a sua concordância e validade dos diagnósticos de cânceres relacionados à exposição ao asbesto adotando-se como padrão ouro os registros do RHC no estado de São Paulo.

MÉTODOS

Neste estudo, compararam-se os óbitos com diagnósticos de câncer relacionados à exposição ao asbesto no estado de São Paulo, provenientes do SIH/SUS, e todos os casos de câncer com informação de óbito na base do RHC/SP. O período de análise considerou os anos de 2007 a 2014, em indivíduos de ambos os sexos, com 30+ anos.

FONTE DOS DADOS

No SIH/SUS, um único paciente pode apresentar várias internações e diagnósticos. Neste estudo, apenas as internações com alta por óbito e com menção de um dos cânceres de interesse em um dos campos de diagnósticos (CID-10¹⁹) foram consideradas.

Na base de dados do RHC/SP (consolidados pela Fundação Oncocentro de São Paulo – FOSP)⁹, foram selecionados os registros de todas as neoplasias malignas codificados pela Classificação Internacional de Doenças para Oncologia, terceira edição (CID-O3)²⁰, com informação de óbito, independentemente da causa, excluindo-se tumores de pele não melanoma.

Ambas as bases eram identificadas.

DEFINIÇÃO DOS DIAGNÓSTICOS DE INTERESSE

Consideraram-se como neoplasias malignas relacionadas ao asbesto o mesotelioma (C45), o câncer de pulmão (C34), de ovário (C56), de laringe (C32). As neoplasias de esôfago (C15), estômago (C16), cólon e reto (C18-C20) foram consideradas suspeitas de relação. O câncer de pleura (C38.4) foi também selecionado por ser um possível diagnóstico de encobrimento do mesotelioma. Da mesma forma, selecionaram-se casos de câncer de cabeça e pescoço (C00-C14, C30-C31), por serem diagnósticos diferenciais com câncer de laringe, assim como cânceres de localização primária desconhecida (C80)^{10,11}.

Entre os grupos oncológicos estudados, apenas o mesotelioma apresenta diferença de identificação entre os códigos da CID-10, classificado como C45, e a CID-O3, identificado pelos códigos morfológicos 9050/3, 9051/3, 9052/3 e 9053/3. Os demais grupos de interesse apresentam equivalência entre o código de três dígitos da CID-10 e CID-O3. Portanto os códigos da CID-O3 e CID-10 de três dígitos foram utilizados para as neoplasias malignas relacionadas ao asbesto citadas anteriormente, exceto para câncer de pleura (C38.4), para este utilizou-se a classificação de quatro dígitos, pois, nesse caso, o último dígito modifica a localização do tumor²⁰.

VINCULAÇÃO DAS BASES DE DADOS

Após a padronização das variáveis na base de dados do SIH/SUS e do RHC/SP, procedeu-se à realização do relacionamento probabilístico com o programa *Link Plus*, assentando-se em padronização, blocagem e pareamento das variáveis. O *Link Plus* é um software desenvolvido pelo Centro de Controle e Prevenção de Doenças dos Estados Unidos (CDC/USA) destinado ao uso nos registros de câncer e utilizado para vinculações entre a base de dados do registro de câncer e a base de dados externas utilizando identificadores comuns²¹⁻²³.

Na etapa de blocagem, selecionou-se a variável sexo, o que permitiu que as bases fossem logicamente divididas em blocos, segundo a variável de ligação. Dessa forma, as comparações e o cálculo de escores são limitados aos registros que pertencem ao mesmo bloco.

Para a etapa de pareamento, selecionaram-se as variáveis: nome completo do indivíduo e nome da mãe, como variáveis genéricas, e a data de nascimento. O *Link Plus* leva em consideração as diferenças fonéticas, bem como a ausência de partes ou a totalidade do sobrenome no momento de vincular e computar a classificação dos relacionamentos. Assim, o programa gera uma pontuação de *linkage* para cada par estabelecido. Pares de registros em que as variáveis de pareamento coincidem recebem a pontuação máxima, no caso, nome completo do indivíduo, nome da mãe e data nascimento. Quanto menor a combinação dos registros entre os três identificadores, menor será a pontuação atribuída. Selecionou-se a opção “*best match*” para que o software apresentasse a melhor opção de match em casos de múltiplos pareamentos. O *Link Plus* recomenda o valor de ponto de corte entre 7 e 10. Após ensaios realizados com a base de dados utilizando o valor 7, observou-se um alto número de pares falso-positivos e optou-se, portanto, pelo valor de corte em 10, selecionando para a revisão manual pares com escore igual ou maior que 10²¹.

Por fim, foi realizada a revisão manual para classificação como par verdadeiro, na qual foram comparadas, além das variáveis de pareamento, as variáveis código postal de residência, ou seja, o CEP, e data de óbito.

ANÁLISE DA CONCORDÂNCIA

Para análise de concordância, considerou-se o diagnóstico do RHC/SP *versus* o diagnóstico principal SIH/SUS dos pares verdadeiros obtidos do relacionamento probabilístico na etapa anterior. Quando o diagnóstico principal no SIH/SUS não era oncológico, utilizaram-se diagnósticos secundários e outros. Quando havia mais de dois diagnósticos oncológicos no mesmo registro, considerou-se como determinante o diagnóstico principal.

A concordância foi avaliada segundo o código CID-10 e a localização anatômica das neoplasias malignas entre aquelas relacionadas à exposição ao asbesto, de acordo com a mensuração da estatística Kappa de Cohen²⁴, um indicador de confiabilidade que quantifica a concordância entre os observadores, ajustado para aquelas ocorridas ao acaso, sendo, portanto, adequado para a análise deste estudo. Além do Kappa bruto, foi calculado o Kappa ajustado por viés de prevalência e viés de discordância marginal (*prevalence-adjusted and bias-adjusted kappa* — PABAK), o índice de prevalência (*prevalence index* — PI) e o índice de viés (*bias index* — BI). Valores de PI e BI baixos sugerem que os valores de Kappa são menos sujeitos a viés e efeito da prevalência, enquanto valores elevados de PI e BI tendem a diminuir o valor de Kappa^{24,25}. As análises foram realizadas no software R, versão 4.0 (R Development Core Team), utilizando-se o pacote epiR.

ANÁLISE DE VALIDADE

Para a análise de validade, calcularam-se sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN), com seus respectivos intervalos de confiança de 95% (IC95%), do SIH/SUS em relação ao RHC/SP. Consideraram-se os resultados obtidos na vinculação dos registros de óbitos no RHC/SP (padrão ouro) com o SIH/SUS dos tumores relacionados ao asbesto (câncer de pulmão, cabeça e pescoço, laringe, mesotelioma, pleura, esôfago, estômago, cólon e reto, ovário e localização primária desconhecida). As análises foram realizadas no software R, versão 4.0 (R Development Core Team), utilizando-se o pacote epiR.

ASPECTOS ÉTICOS

O estudo foi registrado no Conselho Nacional de Ética em Pesquisa (CONEP) e aprovado sob o Certificado de Apresentação para Apreciação Ética (CAAE) 36547514 9 0000 5030, Addendum no. 962 145 e 1 761 856, Instituto de Saúde Coletiva da Universidade Federal da Bahia.

RESULTADOS

A base de dados do SIH/SUS para doenças e neoplasias malignas associadas à exposição ao asbesto continha 360.443 registros múltiplos de internações no período de 2007 a 2014, no estado de São Paulo. Após a identificação dos registros únicos com desfecho óbito, selecionaram-se os diagnósticos oncológicos, portanto obteve-se o total de 56.623 registros no SIH/SUS. No mesmo período, no RHC/SP identificou-se o total de 151.766 óbitos, excluindo os casos de câncer de pele não melanoma.

CONCORDÂNCIA

Tanto no SIH/SUS quanto no RHC/SP houve predomínio do sexo masculino, com 61,3 e 56,8%, respectivamente. A população acima de 50 anos correspondeu a 87,0% no SIH/SUS e a 84,8% no RHC/SP (Tabela 1). Após o relacionamento probabilístico, 64,4% dos casos pareados resultaram no sexo masculino e 56,2% acima de 60 anos (Tabela Suplementar 1).

Tabela 1. Características dos óbitos de acordo com as fontes dos dados, Registros Hospitalares de Câncer do Estado de São Paulo (RHC/SP) e Sistema de Informação Hospitalar (SIH/SUS), estado de São Paulo, Brasil, 2007–2014.

Variáveis	RHC/SP		SIH/SUS	
	n	%	n	%
Sexo				
Masculino	86.137	56,8	34.682	61,3
Feminino	65.629	43,2	21.941	38,7
Faixa etária (anos)				
30–49	23.128	15,2	7.420	13,1
50–69	78.290	51,6	30.329	53,6
70+	50.348	33,2	18.874	33,4
Ano				
2007	16.874	11,1	6.824	12,1
2008	18.989	12,5	5.653	10,0
2009	19.883	13,1	6.410	11,3
2010	19.983	13,2	7.025	12,4
2011	20.202	13,3	7.125	12,6
2012	19.173	12,6	7.560	13,4
2013	18.553	12,2	7.786	13,8
2014	18.109	11,9	8.240	14,6
UF				
São Paulo	143.708	94,7	55.254	97,6
Outros estados	8.058	5,3	1.369	2,4
Total	151.766	100,0	56.623	100,0

UF: Estado de origem do indivíduo.

O relacionamento probabilístico das bases obteve 26.196 pares com escore entre 10.0 e 28.3. Após a revisão manual dos pares com variáveis divergentes e/ou escores abaixo de 20.0, 19.458 foram considerados pares verdadeiros.

Na análise de concordância, os resultados foram apresentados para grupos de CID-10 de cânceres relacionados ou possivelmente relacionados à exposição ao asbesto. Destacou-se a proporção de acordos para o câncer de pleura (99,7%), de mesotelioma (99,6%) e de ovário (98,7%), enquanto as menores proporções foram estimadas para a localização primária desconhecida (92,4%) e o de pulmão (94,7%). Entretanto, os valores do Kappa bruto, distintamente, variaram de 0,05, para o câncer de pleura, a 0,85, para o câncer de pulmão, sugerindo desempenhos inversos em relação aos achados de proporções de acordo. Esses contrastes aparecem com a variação dos números de casos de diagnósticos específicos: 5.179 para o câncer de pulmão e de apenas 99 de mesoteliomas e 71 de câncer de pleura. A Tabela Suplementar 2 relaciona as observações de cada câncer de interesse entre as duas bases. Verifica-se que, entre os 70 casos pareados de mesotelioma registrados no SIH/SUS, apenas 20 (28,6%) apresentaram o mesmo diagnóstico no RHC, e, entre os 73 registrados como câncer de pleura, 6 (8,2%) apresentaram o diagnóstico de mesotelioma ou câncer de pleura no RHC (Tabela Suplementar 3). O Kappa bruto da tabela geral foi de 0,74.

Para o câncer de pulmão, o Kappa bruto foi de 0,85 (IC95% 0,84 – 0,87), e de 0,84 para o de estômago (IC95% 0,83 – 0,86). Para tumores raros, a concordância observada foi de 0,33 (IC95% 0,32 – 0,35) para o mesotelioma, e de 0,05 (IC95% 0,04 – 0,07) para câncer de pleura, enquanto para a localização primária desconhecida foi de 0,33 (IC95% 0,32 – 0,35). Para os índices de prevalência e viés, obteve-se 0,52 a 0,99 e < 0,01 a 0,02, respectivamente (Tabela 2).

Tabela 2. Concordância de diagnósticos de câncer relacionados ao asbesto provenientes da vinculação entre óbitos registrados no Registros Hospitalares de Câncer do Estado de São Paulo (RHC/SP) e Sistema de Informação Hospitalar (SIH/SUS), São Paulo, Brasil, 2007–2014.

CID-10	n	Concordância (%)	Kappa (IC95%)	PI	BI	PABAK
Sistema Respiratório						
Pulmão (C34)	5.179	94,7	0,85 (0,84 – 0,87)	0,52	< 0,01	0,89
Laringe (C32)	1.150	96,7	0,61 (0,60 – 0,63)	0,91	< 0,01	0,93
Mesotelioma (C45)	99	99,5	0,33 (0,32 – 0,35)	0,99	< 0,01	0,99
Pleura (C38.4)	71	99,6	0,05 (0,04 – 0,07)	0,99	< 0,01	0,99
Sistema Digestivo						
Esôfago (C15)	2.407	96,6	0,82 (0,81 – 0,83)	0,79	< 0,01	0,93
Estômago (C16)	3.694	95,7	0,84 (0,83 – 0,86)	0,66	< 0,01	0,91
Cólon e Reto (C18-C20)	2.711	96,6	0,84 (0,83 – 0,86)	0,76	< 0,01	0,93
Outros						
Cabeça e Pescoço (C00-C14)	2.956	95,0	0,77 (0,76 – 0,79)	0,75	< 0,01	0,90
Ovário (C56)	777	98,7	0,79 (0,78 – 0,81)	0,93	< 0,01	0,97
Localização primária desconhecida (C80)	1.930	92,4	0,33 (0,32 – 0,35)	0,88	0,02	0,85

CID-10: Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde; n: pares vinculados; IC95%: intervalo de confiança de 95%; PI: prevalence index; BI: bias index; PABAK: prevalence-adjusted and bias-adjusted kappa.

VALIDADE

Na análise de validade identificou-se maior sensibilidade entre os óbitos por câncer de pulmão (0,90, IC95% 0,90 – 0,91) e estômago (0,86, IC95% 0,85 – 0,87), enquanto a sensibilidade foi menor para aqueles tumores raros, como mesotelioma (0,41, IC95% 0,27 – 0,56) e câncer de pleura (0,08, IC95% 0,01 – 0,25), e classificados como localização primária desconhecida (0,45, IC95% 0,42 – 0,49). O valor preditivo positivo também foi menor para aqueles registrados como localização primária desconhecida (0,32, IC95% 0,29 – 0,34), mesotelioma (0,29, IC95% 0,18 – 0,41) e câncer de pleura (0,04; IC95% 0,01 – 0,16). No entanto, a especificidade foi alta para todos os tumores analisados, assim como o valor preditivo negativo (Tabela 3). Os valores absolutos utilizados para análise de validade estão apresentados na Tabela Suplementar 2.

Tabela 3. Análise de sensibilidade e especificidade entre os registros de óbitos por cânceres associados ao asbesto, vinculados entre Registros Hospitalares de Câncer do Estado de São Paulo (RHC/SP) e Sistema de Informação Hospitalar (SIH/SUS), estado de São Paulo, Brasil, 2007–2014.

CID-10	n	Sensibilidade (IC95%)	Especificidade (IC95%)	VPP (IC95%)	VPN (IC95%)
Sistema Respiratório					
Pulmão (C34)	5.179	0,90 (0,90 – 0,91)	0,96 (0,96 – 0,96)	0,88 (0,87 – 0,89)	0,97 (0,97 – 0,97)
Laringe (C32)	1.150	0,68 (0,64 – 0,71)	0,98 (0,98 – 0,98)	0,58 (0,55 – 0,62)	0,99 (0,98 – 0,99)
Mesotelioma (C45)	99	0,41 (0,27 – 0,56)	1,00 (1,00 – 1,00)	0,29 (0,18 – 0,41)	1,00 (1,00 – 1,00)
Pleura (C38.4)	71	0,08 (0,01 – 0,25)	1,00 (1,00 – 1,00)	0,04 (0,01 – 0,16)	1,00 (1,00 – 1,00)
Sistema Digestivo					
Esôfago (C15)	2.407	0,87 (0,85 – 0,88)	0,98 (0,97 – 0,98)	0,81 (0,80 – 0,83)	0,98 (0,98 – 0,99)
Estômago (C16)	3.694	0,86 (0,85 – 0,87)	0,98 (0,97 – 0,98)	0,88 (0,87 – 0,89)	0,97 (0,97 – 0,97)
Cólon e Reto (C18-C20)	2.711	0,86 (0,84 – 0,87)	0,98 (0,98 – 0,98)	0,86 (0,85 – 0,88)	0,98 (0,98 – 0,98)
Outros					
Cabeça e Pescoço (C00-C14)	2.956	0,75 (0,74 – 0,77)	0,98 (0,98 – 0,98)	0,86 (0,84 – 0,87)	0,96 (0,96 – 0,97)
Ovário (C56)	777	0,86 (0,83 – 0,89)	0,99 (0,99 – 0,99)	0,74 (0,71 – 0,77)	1,00 (0,99 – 1,00)
Localização primária desconhecida (C80)	1.930	0,45 (0,42 – 0,49)	0,95 (0,95 – 0,95)	0,32 (0,29 – 0,34)	0,97 (0,97 – 0,97)

CID-10: Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde; n: pares vinculados; IC95%: intervalo de confiança de 95%; VPP: valor preditivo positivo; VPN: valor preditivo negativo.

DISCUSSÃO

O relacionamento probabilístico entre óbitos de cânceres associados ao asbesto encontrados no SIH/SUS e no RHC/SP no período de 2007 a 2014 no estado de São Paulo revelou boa concordância para cânceres mais frequentes, como pulmão [Kappa = 0,85 (IC95% 0,84 – 0,87)], laringe, ovário e cânceres digestivos. Entretanto, a concordância mostrou-se baixa para cânceres raros como o mesotelioma [Kappa = 0,33 (IC95% 0,32 – 0,35)] e o câncer de pleura [Kappa = 0,05 (IC95% 0,04 – 0,07)], ambos com valor preditivo positivo (VPP) muito baixo, 0,29 e 0,04, respectivamente. O Kappa geral foi de 0,74 (Tabela Suplementar 3), classificado como substancial, conforme os critérios de Landis e Koch²⁶, ou considerado bom, segundo Fleiss²⁷ e Altman²⁸. Entretanto, nesse indicativo observamos o efeito da baixa prevalência dos tumores como o mesotelioma, o câncer de pleura e de localização primária desconhecida. Esse efeito ocasionou menores valores de concordância, se comparados com a concordância específica para tumores mais prevalentes, um dos paradoxos do coeficiente Kappa de Cohen descritos por Feinstein e Cicchetti²⁹.

Na tentativa de melhor analisar o VPP para o mesotelioma e o câncer de pleura em razão dos pequenos números, restringiu-se o número de observações total nas tabelas de contingência para alguns diagnósticos específicos, passíveis de encobrimento desses tumores: câncer de pulmão (C34), coração, timo e pleura (C38), tecido conjuntivo e outros tecidos moles do tórax (C49.3), tecidos moles do abdômen (C49.4), tecidos moles da pelve (C49.5), outras localizações e localizações mal definidas (C76.1, C76.2 e C76.3), localização primária desconhecida (C80). Esse exercício mostrou não haver mudanças em relação aos cálculos de sensibilidade, especificidade e VPP, pois os denominadores mantiveram-se altos. Especificamente, o registro do câncer de pleura (C38.4) no SIH/SUS apresentou sensibilidade muito baixa (Tabela 3): dos 47 casos, apenas 2 foram confirmados pelo RHC (Tabela Suplementar 4). Possivelmente, o registro de internação não corroborado pelo registro de câncer se deve à pleura ser local comum de metástases por outros tumores. Um estudo sobre acurácia das declarações de óbito por câncer mostrou que, quando há um percentual maior de registros de óbito por determinado câncer, comparativamente aos registros específicos de câncer, as declarações de óbito possivelmente expressam uma superestimativa da doença³⁰.

Há um potencial de emprego de dados do SIH/SUS para propósitos epidemiológicos. Veras e Martins¹⁶ analisaram 1.934 registros de Autorização de Internação Hospitalar (AIH), comparando diagnósticos constantes com prontuários médicos em hospitais privados conveniados à rede pública. As autoras encontraram boa concordância para os dados sociodemográficos e administrativos, porém, para os diagnósticos selecionados, houve variação de acordo com a agregação: Kappa = 0,81 (IC95% 0,77 – 0,85) para três dígitos e Kappa = 0,72 (IC95% 0,68 – 0,76) para quatro dígitos. Mathias e Soboll¹⁷, em outro estudo de concordância entre as CIDs constantes nas AIHs de 1.595 internações, tomando como base o diagnóstico nos prontuários médicos, encontraram uma baixa concordância com três dígitos de agregação entre diagnósticos de neoplasias (Kappa = 0,46 IC95% 0,34 – 0,57), contrariamente às doenças do aparelho circulatório (Kappa = 0,92 IC95% 0,89 – 0,95) e do aparelho respiratório (Kappa = 0,91

IC95% 0,88 – 0,94). As autoras concluíram que a concordância foi maior para diagnósticos mais comuns. Um estudo sobre a qualidade dos dados em casos de óbito por infarto agudo do miocárdio no município do Rio de Janeiro, comparando-se registros do SIH/SUS com do SIM, identificou a ausência de registro de óbitos ocorridos em serviços de urgência, o emprego de outros diagnósticos no diagnóstico principal e ausência de preenchimento de diagnósticos secundários e o erro de codificação no tipo de saída na AIH³¹. Os autores comentam que, ao contrário do que acontece no SIM, não há regras claras para a inclusão e revisão de códigos da CID nos formulários de internação hospitalar e o preenchimento dos dados clínicos é realizado por pessoal administrativo sem treinamento específico^{16,17,31}.

Os sistemas de classificação de doenças empregados nas duas bases unidas são diferentes. O SIH/SUS emprega a CID-10, que representa doenças determinadas, enquanto o RHC utiliza a CID-O3, uma classificação biaxial que leva em conta a topografia e a morfologia dos cânceres, sendo, portanto, aberta a diferentes combinações. Embora distintas, em um estudo de integração, as classificações CID-10 e CID-O3 concordaram conceitualmente em 88%³². Na presente análise, os códigos de interesse apresentavam correspondência topográfica direta entre as classificações, exceção feita ao mesotelioma, identificado no RHC pela sua morfologia.

Este estudo demonstra a melhor concordância entre as informações constantes no SIH/SUS e no RHC/SP para algumas neoplasias de maior incidência e uma baixa concordância para neoplasias raras ou mal definidas. Entretanto, esses resultados foram obtidos comparando-se as duas bases nas quais as informações estão consolidadas, não sendo viável o acesso e a verificação das informações clínicas em todas as unidades hospitalares de ocorrência. Em contrapartida, o elevado número de pares analisados e a alta proporção de diagnósticos com confirmação histológica no RHC/SP sugerem que os achados são confiáveis. Destaca-se que, em contraste com as AIHs, o registro dos casos de câncer na base de dados do RHC segue protocolos nacionais e internacionais, sendo realizado por registradores de câncer treinados e as informações coletadas em prontuários hospitalares⁷⁻⁹.

Avaliando a confiabilidade das causas básicas de óbitos por câncer entre o RCBP e o SIM em Goiânia, Oliveira et al.³³ observaram que tumores raros (com baixa incidência) apresentaram concordância bruta baixa, no entanto, quando ajustados os valores, tanto pela prevalência quanto pelo viés de discordância marginal (PABAK), obteve-se o aumento dos valores de concordância ajustada. A mesma tendência foi aqui observada.

Entre as limitações deste estudo, cita-se a restrição das análises para o período de 2007 a 2014, apesar de o SIH/SUS possuir dados informatizados desde 1984 e o RHC/SP desde 2000. Porém, em virtude da ausência de um identificador único comum entre as bases, limitou-se a seleção a partir do ano de 2007, quando a base do SIH/SUS passou a apresentar melhor preenchimento para as variáveis de nome da mãe do indivíduo, importante para a distinção entre homônimos³⁴. Outra limitação é o pequeno número de registros para tumores raros, evidenciando as limitações da estatística Kappa de Cohen, uma vez que esse teste é extremamente sensível às distribuições dos totais marginais e pode produzir resultados espúrios. No entanto, os valores de prevalência e Kappa ajustado de viés (PABAK) podem ser utilizados na compreensão de quão sensível é a estatística à distribuição dos totais marginais³⁵. Lima et al. verificaram que a estatística Kappa tem se mostrado útil em estudos de

confiabilidade para dados do SIH/SUS, do SIM, do Sistema de Informação de Agravos de Notificação (SINAN) e do Sistema de Informações de Nascidos Vivos (SINASC), e a análise de validade também tem sido aplicada a essas bases de dados. Nessa perspectiva, este estudo considerou, portanto, que as duas análises são complementares de interpretação³⁶. Anteriormente às análises aqui apresentadas, testou-se a concordância entre as mesmas bases em um ensaio com 226 casos em que havia o diagnóstico de câncer de pleura e mesotelioma em óbitos (SIH/SUS) no estado de São Paulo e não possuíam correspondência no SIM⁶. Esses casos foram vinculados com o banco do RHC/SP e encontraram-se 104 pares verdadeiros. Da mesma forma, a análise do Kappa, com as caselas mais bem balanceadas, evidenciou baixa concordância, sendo que boa parte dos casos era de câncer de pulmão e/ou metástases pleurais de outras neoplasias. Por fim, existe a limitação vinculada à temporalidade dos registros, pois o RHC recebe as informações dos hospitais da rede participante após o diagnóstico estabelecido, sendo que o evento óbito pode ocorrer tempos depois, sendo possível, em casos de cânceres de baixa letalidade, ter ocorrido por causas não relacionadas à neoplasia e, portanto, não capturado pelo registro de internação.

Apesar de os SIS do SUS possuírem, em sua maioria, finalidade administrativa, a utilização de seus dados em estudos epidemiológicos deve ser considerada. Quanto ao SIH/SUS, as informações devem ser utilizadas com cautela, levando-se em consideração diagnósticos específicos ou o grupo de diagnósticos a serem analisados. O SIH/SUS, por sua abrangência nacional e boa cobertura dos eventos hospitalares, com ajustes nas regras e, principalmente, no treinamento dos agentes administrativos responsáveis pelo seu preenchimento^{15,37}, pode constituir-se em uma rica fonte de informação em saúde.

Reforça-se, além de um cuidado maior na inserção de códigos corretos dos diagnósticos de internações, a importância do preenchimento de uma chave única nos SIS para melhorar a precisão e segurança na identificação dos mesmos indivíduos entre as bases, contribuindo assim na utilização de séries históricas dos sistemas de informações em saúde^{1,38}.

AGRADECIMENTOS

As bases de dados não anônimas de internações no sistema público foram gentilmente disponibilizadas por Simone Santos e Rosemeire Norye Inamine, Secretaria de Saúde do Estado de São Paulo.

REFERÊNCIAS

1. Ali MS, Ichihara MY, Lopes LC, Barbosa GCG, Pita R, Carreiro RP, et al. Administrative Data Linkage in Brazil: Potentials for Health Technology Assessment. *Front Pharmacol* 2019; 10: 984. <http://dx.doi.org/10.3389/fphar.2019.00984>
2. Mendes ACG, Silva Junior JB, Medeiros KR, Lyra TM, Melo Filho DA, Sá DA. Avaliação do sistema de informações hospitalares - SIH/SUS como fonte complementar na vigilância e monitoramento de doenças de notificação compulsória. *Inf Epidemiol*

- Sus 2000; 9(2): 67-86. <http://dx.doi.org/10.5123/S0104-1673200000200002>
3. Brasil. Ministério da Saúde. Departamento de Ciência e Tecnologia. Secretaria de Ciência e Tecnologia e Insumos Estratégicos. Integração de informações dos registros de câncer brasileiros. *Rev Saúde Pública* 2007; 41(5): 865-8.
 4. Brasil. Ministério da Saúde. Pesquisa Nacional de Saúde: 2013 acesso e utilização dos serviços de saúde, acidentes e violências: Brasil, grandes regiões e unidades da federação. Rio de Janeiro: IBGE; 2015.
 5. Campos MR, Martins M, Noronha JC, Travassos C. Proposta de integração de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS) para pesquisa. *Inf Epidemiol Sus* 2000(1); 9: 51-8. <http://dx.doi.org/10.5123/S0104-1673200000100005>
 6. Santana VS, Algranti E, Campos F, Cavalcante F, Salvi L, Santos SA, et al. Recovering missing mesothelioma deaths in death certificates using hospital records. *Am J Ind Med* 2018; 61(7): 547-55. <http://dx.doi.org/10.1002/ajim.22846>
 7. Instituto Nacional de Câncer. Registros hospitalares de câncer: rotinas e procedimentos. Rio de Janeiro: INCA; 2000.
 8. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* 2009; 45(5): 747-55. <http://dx.doi.org/10.1016/j.ejca.2008.11.032>
 9. Fundação Oncocentro de São Paulo. Registro Hospitalar de Câncer: Conceitos, rotinas e instruções de preenchimento [Internet]. 2ª ed. São Paulo: Fundação Oncocentro de São Paulo; 2013 [acessado em 4 out. 2020]. Disponível em: http://www.fosp.saude.sp.gov.br:443/epidemiologia/docs/ManualRHC_2013.pdf
 10. Takala J. Eliminating occupational cancer in Europe and globally [Internet]. [acessado em 4 dez. 2020]. Disponível em: https://oshwiki.eu/wiki/Eliminating_occupational_cancer_in_Europe_and_globally
 11. Straif K, Benbrahim-Tallaa L, Baan R, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens-Part C: metals, arsenic, dusts, and fibres. *Lancet Oncol* 2009; 10(5): 453-4. [http://dx.doi.org/10.1016/S1470-2045\(09\)70134-2](http://dx.doi.org/10.1016/S1470-2045(09)70134-2)
 12. Kim SJ, Williams D, Cheresch P, Kamp DW. Asbestos-Induced Gastrointestinal Cancer: An Update. *J Gastrointest Dig Syst* 2013; 3: 135. <http://dx.doi.org/10.4172/2161-069X.1000135>
 13. Camidge DR, Stockton DL, Bain M. Factors affecting the mesothelioma detection rate within national and international epidemiological studies: insights from Scottish linked cancer registry-mortality data. *Br J Cancer* 2006; 95(5): 649-52. <http://dx.doi.org/10.1038/sj.bjc.6603293>
 14. Labrèche F, Case BW, Ostiguy G, Chalaoui J, Camus M, Siemiatyki J. Pleural meothelioma surveillance: validity of cases from a tumour registry. *Can Respir J* 2012; 19(2): 103-7. <https://doi.org/10.1155/2012/650935>
 15. Bittencourt SA, Camacho LAB, Leal MC. O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva. *Cad Saúde Pública* 2006; 22(1): 19-30. <https://doi.org/10.1590/S0102-311X2006000100003>
 16. Veras CMT, Martins MS. A confiabilidade dos dados nos formulários de autorização de internação hospitalar (AIH), Rio de Janeiro, Brasil. *Cad Saúde Pública* 1994; 10(3): 339-55. <https://doi.org/10.1590/S0102-311X1994000300014>
 17. Mathias TAF, Soboll MLMS. Confiabilidade de diagnósticos nos formulários de autorização de internação hospitalar. *Rev Saúde Pública* 1998; 32(6): 26-32. <https://doi.org/10.1590/S0034-89101998000600005>
 18. Machado JP, Martins M, Leite IC. Quality of hospital databases in Brazil: some elements. *Rev Bras Epidemiol* 2016; 19(3): 567-81. <https://doi.org/10.1590/1980-54972016000300008>
 19. World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems: tenth revision. 2ª ed. Genebra: WHO; 2004.
 20. Percy C, van Holten V, Munir C, editores. CID-O: Classificação Internacional de Doenças para Oncologia. 3ª ed. São Paulo: EDUSP; 2005.
 21. National Program of Cancer Registries. Registry Plus™ Link Plus Features and Future Plans [Internet]. Atlanta: Centers for Disease Control and Prevention; 2018 [acessado em 11 jan. 2019]. Disponível em: https://www.cdc.gov/cancer/npcr/tools/registryplus/lp_features.htm
 22. Campbell KM, Deck D, Krupski A. Record linkage software in the public domain: A comparison of Link Plus, The Link King, and a 'basic' deterministic algorithm. *Health Informat J* 2008; 14(1): 5-15. <https://doi.org/10.1177/1460458208088855>
 23. Garvin JH, Herget KA, Hashibe M, Kirchhoff AC, Hawley CW, Bolton D, et al. Linkage between Utah All Payers Claims Database and Central Cancer Registry. *Health Serv Res* 2019; 54(3): 707-13. <https://doi.org/10.1111/1475-6773.13114>
 24. Cohen JA. Coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20(1): 37-46. <https://doi.org/10.1177%2F001316446002000104>
 25. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46(5): 423-9. [https://doi.org/10.1016/0895-4356\(93\)90018-v](https://doi.org/10.1016/0895-4356(93)90018-v)

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159-74. <https://doi.org/10.2307/2529310>
27. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76(5): 378-82.
28. Altman DG. *Practical statistics for medical research*. Londres/Nova York: Chapman and Hall; 1991.
29. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43(6): 543-9. [https://doi.org/10.1016/0895-4356\(90\)90158-1](https://doi.org/10.1016/0895-4356(90)90158-1)
30. German RR, Fink AK, Heron M, Stewart SL, Johnson CJ, Finch JL, et al. The accuracy of cancer mortality statistics based on death certificates in the United States. *Cancer Epidemiol* 2011; 35(2): 126-31. <https://doi.org/10.1016/j.canep.2010.09.005>
31. Melo ECP, Travassos C, Carvalho MS. Qualidade dos dados sobre óbitos por infarto agudo do miocárdio, Rio de Janeiro. *Rev Saúde Publica* 2004; 38(3): 385-91. <https://doi.org/10.1590/S0034-89102004000300008>
32. Nikiema J, Jouhet V, Mougín F. Integrating cancer diagnosis terminology based on logical definitions of SNOMED CT concepts. *J Biomed Inform* 2017; 74: 46-58. <https://doi.org/10.1016/j.jbi.2017.08.013>
33. Oliveira PPV, Silva GA, Curado MP, Malta DC, Moura L. Confiabilidade da causa básica de óbito por câncer entre Sistema de Informações sobre Mortalidade do Brasil e Registro de Câncer de Base Populacional de Goiânia, Goiás, Brasil. *Cad Saúde Pública* 2014; 30(2): 296-304. <https://doi.org/10.1590/0102-311X00024813>
34. Peres SV, Latorre MRDO, Tanaka LF, Michels FAS, Teixeira MP, Coeli CM, et al. Melhora na qualidade e completude da base de dados do Registro de Câncer de Base Populacional do município de São Paulo: uso de técnicas de linkage. *Rev Bras Epidemiol* 2016; 19(4): 753-65. <https://doi.org/10.1590/1980-5497201600040006>
35. Flight L, Julious SA. The disagreeable behaviour of the kappa statistic. *Pharm Stat* 2015; 14(1): 74-8. <https://doi.org/10.1002/pst.1659>
36. Lima CRA, Schramm JMA, Coeli CM, Silva MEM. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cad Saúde Pública* 2009; 25(10): 2095-109. <https://doi.org/10.1590/S0102-311X2009001000002>
37. Machado JP, Martins M, Leite IC. Qualidade das bases de dados hospitalares no Brasil: alguns elementos. *Rev Bras Epidemiol* 2016; 19(3): 567-81. <https://doi.org/10.1590/1980-5497201600030008>
38. Queiroz OV, Guerra Júnior AA, Machado CJ, Andrade EIG, Meira Junior W, Acurcio FA, et al. Relacionamento de registros de grandes bases de dados: estimativa de parâmetros e validação dos resultados, aplicados ao relacionamento dos registros das autorizações de procedimentos ambulatoriais de alta complexidade com os registros de sistema de informações hospitalares. *Cad Saúde Coletiva* 2010; 18(2): 298-308.

Recebido em: 13/01/2021

Revisado em: 07/04/2021

Aprovado em: 23/04/2021

Contribuição dos autores: Silva DRM participou do planejamento, análise, interpretação e discussão dos resultados e redação do artigo. Luizaga CTM participou da análise e interpretação dos resultados e revisão crítica do conteúdo. Toporcov TN participou da discussão dos resultados, redação e revisão crítica do conteúdo. Algranti E participou do planejamento, interpretação e discussão dos resultados e redação do artigo. Todos os autores aprovaram a versão final a ser publicada.

