



Revista de Calidad Asistencial

www.elsevier.es/calasis



ORIGINAL

Desarrollo y validación de un algoritmo para identificar recidivas de cáncer a partir de bases de datos hospitalarias



S. Manzaneres-Laya^{a,b}, A. Burón^{a,c,d,*}, C. Murta-Nascimento^{a,c,d}, S. Servitja^{c,e},
X. Castells^{a,c,d} y F. Macià^{a,b,c,d}

^a Servei d'Epidemiologia i Avaluació, Hospital del Mar-Parc de Salut Mar, Barcelona, España

^b Unitat Docent de Medicina Preventiva i Salut Pública, H. MAR-UPF-ASPB, Barcelona, España

^c IMIM Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, España

^d Red de Investigación en Servicios de Salud en Enfermedades Crónicas, REDISSEC, Barcelona, España

^e Servei d'Oncologia, Hospital del Mar-Parc de Salut Mar, Barcelona, España

Recibido el 10 de febrero de 2014; aceptado el 15 de mayo de 2014

Disponible en Internet el 28 de junio de 2014

PALABRAS CLAVE

Registros de
hospitales;
Cáncer;
Sensibilidad;
Especificidad;
Algoritmos

Resumen

Objetivos: Los registros de tumores hospitalarios y las bases de datos hospitalarias son una fuente de información valiosa y eficiente para la investigación de recidivas de cáncer. El objetivo de este estudio fue desarrollar y validar algoritmos para identificar recidivas de cáncer de mama.

Métodos: Estudio observacional retrospectivo de casos de cáncer de mama del registro de tumores de un centro hospitalario universitario de tercer nivel diagnosticados entre 2003 y 2009. A partir del cruce de bases de datos hospitalarias y la construcción de definiciones operativas se obtuvieron diferentes algoritmos de probable recidiva de cáncer con su correspondiente sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo.

Resultados: Se identificaron 1.523 pacientes diagnosticados de cáncer entre 2003 y 2009. La solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico obtuvo la mayor sensibilidad (53,8%) y valor predictivo negativo (93,8%), y la realización de una prueba de anatomía patológica tras 6 meses desde el diagnóstico obtuvo la mayor especificidad (93,8%) y valor predictivo negativo (92,6%). La combinación de definiciones aumentó la especificidad y el valor predictivo positivo pero disminuyó la sensibilidad.

Conclusiones: Se elaboraron diferentes algoritmos diagnósticos cuyas definiciones pueden ser útiles según los intereses y recursos del investigador. Un mayor valor predictivo positivo podría interesar para una estimación rápida del número de casos, y un mayor valor predictivo negativo para dar una estimación más exacta si se dispone de mayores recursos. Estos algoritmos se

* Autor para correspondencia.

Correos electrónicos: aburon@parcdesalutmar.cat, buron.andrea@gmail.com (A. Burón).

KEYWORDS

Hospital registries;
Cancer;
Sensitivity;
Specificity;
Algorithms

configuran como una herramienta versátil y adaptable a otros tumores y a las necesidades del investigador.

© 2014 SECA. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Development and validation of an algorithm to identify cancer recurrences from hospital data bases**Abstract**

Objectives: Hospital cancer registries and hospital databases are valuable and efficient sources of information for research into cancer recurrences. The aim of this study was to develop and validate algorithms for the detection of breast cancer recurrence.

Methods: A retrospective observational study was conducted on breast cancer cases from the cancer registry of a third level university hospital diagnosed between 2003 and 2009. Different probable cancer recurrence algorithms were obtained by linking the hospital databases and the construction of several operational definitions, with their corresponding sensitivity, specificity, positive predictive value and negative predictive value.

Results: A total of 1,523 patients were diagnosed of breast cancer between 2003 and 2009. A request for bone gammagraphy after 6 months from the first oncological treatment showed the highest sensitivity (53.8%) and negative predictive value (93.8%), and a pathology test after 6 months after the diagnosis showed the highest specificity (93.8%) and negative predictive value (92.6%). The combination of different definitions increased the specificity and the positive predictive value, but decreased the sensitivity.

Conclusions: Several diagnostic algorithms were obtained, and the different definitions could be useful depending on the interest and resources of the researcher. A higher positive predictive value could be interesting for a quick estimation of the number of cases, and a higher negative predictive value for a more exact estimation if more resources are available. It is a versatile and adaptable tool for other types of tumors, as well as for the needs of the researcher.

© 2014 SECA. Published by Elsevier España, S.L. All rights reserved.

Introducción

El cáncer sigue siendo una de las principales causas de muerte en todo el mundo. Fue responsable de 7,6 millones de muertes en 2008, lo cual representó el 13% de todas ellas¹. Aunque globalmente la supervivencia relativa del cáncer ha aumentado en los últimos años²⁻⁴, la aparición de recidivas tumorales empeora considerablemente el pronóstico y sigue siendo frecuente en ciertos tipos de cáncer⁵⁻⁷.

Se considera recidiva tumoral la aparición de cualquier signo o síntoma de la enfermedad tras un periodo de remisión. Un estudio francés siguió durante un mínimo de 12 años a pacientes diagnosticadas de cáncer de mama invasivo y observó un índice de primera recidiva locorregional del 14,6%, con una media de seguimiento de 5,8 años⁷. El conocimiento del comportamiento de estas es determinante en el pronóstico de la enfermedad, ya que impacta en la calidad de vida y se relaciona con la aparición de nuevas recidivas y la muerte debida al tumor⁶. Sin embargo, el estudio epidemiológico de las recidivas tumorales se ve frecuentemente limitado por las dificultades que existen para identificarlas.

La fuente de información habitual para estudiar las características de los tumores y su evolución son los registros de tumores hospitalarios (RTH) que, a diferencia de los registros poblacionales, generalmente disponen de información detallada sobre la histología y otras características del

tumor y el tratamiento del paciente⁸. En cambio, los registros poblacionales son esenciales en el control del cáncer a nivel poblacional, ya que al recoger exhaustivamente todos los tumores de un área geográfica constituyen el principal sistema de vigilancia oncológico y permiten el cálculo de incidencias y prevalencias⁹⁻¹². La exhaustividad de los datos incluidos en los RTH es variable según el centro, y la detección de recidivas tumorales a partir de ellos es complicada, debido a que es difícil la distinción, en la práctica clínica, entre progresión de enfermedad y recidiva, así como entre tumores múltiples, bilaterales, primarios o metastásicos.

En este contexto, el cruce de datos de RTH con otras bases de datos hospitalarias, cuyo papel también es reconocido en la investigación en salud¹³, podría configurarse como fuente de identificación de posibles recidivas. La elaboración de algoritmos definitorios de caso mediante el cruce de registros informatizados en el ámbito sanitario está siendo un método cada vez más utilizado para la investigación clínica y epidemiológica de diferentes enfermedades, principalmente crónicas¹⁴⁻²³. En nuestro medio existen diversos estudios que lo han aplicado en el ámbito de la oncología²⁴⁻²⁶, en particular en el estudio del cáncer de mama²⁷⁻³⁰. La calidad de los algoritmos diagnósticos es clave para determinar su utilidad, por lo que deben ser validados de manera conveniente³¹.

El objetivo de este estudio fue desarrollar y evaluar un algoritmo que identificase de forma semiautomática y eficiente posibles recidivas entre aquellos pacientes diagnosticados de cáncer. Por motivos prácticos el primer algoritmo se diseñó para detectar recidivas en el cáncer de mama (CM), dado que las recidivas de este tumor son relativamente frecuentes (pudiéndose dar a nivel locorregional, contralateral o a distancia) y con la intención de posteriormente adaptarlo a otro tipo de tumores.

Material y métodos

Se trató de un estudio observacional retrospectivo cuyas fuentes de información fueron diversas bases de datos hospitalarias. La fuente de datos principal y denominador fue el RTH de un centro hospitalario universitario de tercer nivel. Operativo desde 1978, el RTH incluye toda neoplasia diagnosticada tanto por ingreso hospitalario como por atención en el servicio de urgencias o en consultas externas, así como los casos diagnosticados por autopsia. Este registro identifica los casos de tumores primarios diagnosticados, no pacientes, sin hacer seguimiento de la enfermedad (excepto del estado vital de la persona), por lo que no se incluyen las posibles recidivas que pueden existir. Sus principales fuentes de información fueron el registro de anatomía patológica (AP) y del conjunto mínimo básico de datos (CMBD). Cuando un mismo paciente tiene más de un registro se debe a que tiene otra neoplasia diferente. En el caso de los pacientes con CM un segundo registro de CM puede significar la aparición de una neoplasia en la mama contralateral. Así pues, se identificaron los casos de CM diagnosticados por primera vez según la Clasificación Internacional de Enfermedades en Oncología (CIE-O 2.^a edición)³² entre 2003 y 2009, por ser este el periodo en el cual, en el momento de llevar a cabo el estudio, se disponían de todos los datos y variables necesarias. Se incluyeron todos los casos de CM tanto en mujeres como en hombres, sin límite de edad.

Para la identificación de recidivas se cruzó la base de datos de los casos de CM del RTH con el registro de AP, CMBD, el registro de gammagrafías óseas, el de tratamientos citostáticos y el del comité de tumores, que conformaron los numeradores. El cruce de bases de datos se consiguió mediante el número de historia clínica, un código de identificación unívoco.

Se establecieron diferentes definiciones operativas de probable recidiva de CM (pRCM) mediante el consenso de expertos para las diferentes bases de datos, resumidas en la [tabla 1](#).

Posteriormente, se construyeron diferentes combinaciones de fuentes de información y definiciones operativas para elaborar el algoritmo definitorio de pRCM.

Para cada definición operativa simple, así como para las definiciones combinadas, se calculó el número de pRCM detectadas, y para el estudio de validez de estas definiciones se calcularon la sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN) y sus correspondientes intervalos de confianza del 95%^{33,34}. Para ello se utilizaron 2 fuentes de datos con recidivas reales. Para la sensibilidad se obtuvo un registro manual de 104 casos documentados de recidiva de CM diagnosticados entre 2003 y 2009 que fueron atendidos en la consulta de oncología de

dicho hospital, identificando el porcentaje de los casos de recidiva reportados por el servicio de oncología que habían sido clasificados como recidiva por el algoritmo (verdaderos positivos), así como los que no habían sido clasificados como recidivas por este (falsos negativos). Para el cálculo de la especificidad, y dado que los casos que no desarrollan recidiva no llegan a dicha consulta, el personal médico revisó manualmente las historias clínicas de los casos de CM diagnosticados en 2003, con el fin de disponer del máximo tiempo de seguimiento y de posibilidad de aparición de recidiva. Cada uno de estos pacientes fue clasificado en pacientes con recidiva o sin recidiva hasta 2009, obteniendo el porcentaje de verdaderos negativos y de falsos positivos identificados por el algoritmo.

El cruce sucesivo de las diferentes bases de datos y los análisis estadísticos se realizaron utilizando el paquete estadístico SPSS versión 17.0³⁵.

Resultados

Se identificaron 1.523 pacientes diagnosticados de CM a partir del RTH entre 2003 y 2009. La [tabla 2](#) muestra el número de pRCM identificadas para cada una de las definiciones operativas simples. Los registros y definiciones operativas que mayor número de probables recidivas identificaron fueron el de gammagrafías óseas que identificó 208 recidivas (13,7% del total de casos inicialmente identificados por el RTH) según la definición operativa de «solicitud de realización de gammagrafía ósea posterior a los 6 meses desde la fecha del primer tratamiento oncológico» y el registro de AP que identificó 184 (12,1%) según «prueba de AP a los 6 meses o más del diagnóstico o a los 3 meses o más del tratamiento quirúrgico». Sin embargo, solamente se hallaron 55 recidivas (3,6%) según la definición operativa de «diagnóstico al alta de CM tras 6 meses desde el diagnóstico inicial», utilizando el registro del CMBD, y 43 recidivas (2,8%) según «nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado» cuando se usó el registro de tratamientos citostáticos.

La [tabla 2](#) presenta también la sensibilidad, especificidad, VPP y VPN de cada registro y definición operativa. Destaca un VPN superior al 90% en la detección de recidivas por CM por parte de los registros de AP y gammagrafías óseas, lo que significa que los casos identificados como probable no recidiva tienen una probabilidad superior al 90% de realmente no ser recidivas. Asimismo, la solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico tuvo una mayor sensibilidad (53,8%; IC 95%: 43,8-63,6%) y un mayor VPN (93,8%; IC 95%: 88,1-96,9%) que si esta se solicitaba tras el segundo tratamiento oncológico. Entre las diferentes definiciones operativas relativas a la prueba de AP, la correspondiente a su realización tras 6 meses desde el diagnóstico resultó tener mayor especificidad (93,8%; IC 95%: 88,5-96,8%) y VPN (92,6%; IC 95%: 87,1-95,9%).

La [tabla 3](#) muestra distintas combinaciones de las definiciones operativas estudiadas, conformando así nuevas definiciones de probable recidiva de CM. De forma global, mediante la combinación de registros se consiguió una especificidad cercana al 100% en la mayoría de las definiciones

Tabla 1 Definiciones operativas de probable recidiva de CM (pRCM) consensuadas según base de datos y criterios de inclusión

Base de datos	Criterios de inclusión	Definiciones operativas de pRCM
AP	Piezas anatómicas de CM ^a Metástasis de origen no especificado ^a Tumores malignos ^a	«Prueba positiva de AP tras 6 meses desde el diagnóstico» «Prueba positiva de AP tras 3 meses desde el tratamiento quirúrgico» «Prueba positiva de AP tras 6 meses desde diagnóstico o tras 3 meses desde el tratamiento quirúrgico»
CMBD	Casos con diagnóstico al alta (primario o secundario) de CM ^b Casos con diagnóstico de CM <i>in situ</i> , de neoplasia de mama de evolución incierta y neoplasia de mama de naturaleza no especificada ^b	«Diagnóstico al alta de CM tras 6 meses desde diagnóstico inicial»
Registro de gammagrafías	Casos a los cuales se les había solicitado una gammagrafía ósea por parte de los servicios de oncología, cirugía general o ginecología, independientemente del resultado	«Solicitud de gammagrafía ósea posterior a los 6 meses desde la fecha del primer tratamiento oncológico» «Solicitud de gammagrafía ósea posterior a los 6 meses desde la fecha del segundo tratamiento oncológico», entendiendo que en oncología puede administrarse el tratamiento en 2 períodos diferentes (neoadyuvancia y adyuvancia)
Registro de tratamiento con citostáticos	Tratamientos dirigidos al CM	«Nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado»
Comité de tumores	Casos de CM presentados en el comité de tumores	«Paciente presentado más de una vez en el comité de tumores tras 5 meses desde la última vez que fue presentado»

^a Según la topografía y morfología de la Clasificación Internacional de Enfermedades en Oncología (CIE-O 2.^a edición)

^b Según los códigos de la Clasificación Internacional de Enfermedades 9.^a revisión Modificación Clínica (CIE 9-MC)

de caso, con intervalos de confianza estrechos. Cabe destacar una disminución global en la sensibilidad respecto al uso de un solo registro, así como un aumento notable del VPP, ambos parámetros con intervalos de confianza amplios. Se observó una ligera disminución del VPN, no obstante, este fue cercano al 90% con intervalos de confianza estrechos, especialmente en aquellas definiciones de caso que combinan los registros de AP y gammagrafías óseas.

Discusión

Los resultados de este estudio indican que es posible desarrollar sistemas factibles de detección de recidivas a partir del cruce semiautomático de bases de datos hospitalarias. El índice de recidivas hallado en las definiciones operativas simples de los registros de AP y de gammagrafías óseas es similar al hallado en otros estudios⁶. Los altos valores predictivos y especificidad obtenidos reducen de manera importante el número de casos a revisar de manera manual con márgenes de error razonables, especialmente porque, hasta la fecha, la búsqueda de recidivas se hace de manera manual. La combinación de diferentes bases de datos y definiciones operativas aporta una especificidad cercana al 100% y un aumento del VPP, disminuyendo ligeramente el VPN.

En nuestro estudio se proponen diferentes escenarios en los que el investigador puede preferir tener un alto VPN o

VPP, dependiendo de sus intereses y necesidades. Un alto VPP puede ser útil cuando el investigador quiere tener una estimación del número de recidivas ocurridas y dispone de pocos o inexistentes recursos para revisar los casos manualmente, ya que el elevado VPP le asegura que aquellos casos clasificados como pRCM son realmente recidivas con una alta probabilidad, pudiendo asumir que son recidivas sin realizar revisión. Bajo este supuesto, la definición operativa compuesta «prueba de AP tras 6 meses desde el diagnóstico, solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado» presenta un VPP del 100% (IC 95%: 97,1-99,9%), teniendo en cuenta que tendría un gran número de falsos negativos (sensibilidad del 12,5%; IC 95%: 7,1-20,8%). Actualmente los investigadores han de realizar una búsqueda manual de los casos de recidivas para cuantificarlas, por lo que este método de detectarlas supone una gran ayuda, ya que detecta de forma semiautomática a un subconjunto de estas, disminuyendo el volumen a revisar.

Por otro lado, al investigador le podría interesar tener un alto VPN y una alta especificidad si dispone de más recursos para la revisión manual de las pRCM, ya que el elevado VPN le aseguraría que con una gran probabilidad los casos clasificados como *probables no recidivas* por el algoritmo realmente son *no recidivas*, que serían los casos más numerosos, ya que la mayor parte de pacientes diagnosticados de

Tabla 2 Número de probables recidivas de cáncer de mama (CM), identificadas como tales según diferentes definiciones operativas simples (n = 1.523)

Definición operativa simple	N (% sobre total de CM del RTH)	Sensibilidad, % (IC 95%) ^a	Especificidad, % (IC 95%) ^b	Valor predictivo positivo, % (IC 95%) ^b	Valor predictivo negativo, % (IC 95%) ^b
Prueba positiva de AP tras 6 meses desde el diagnóstico	169 (11,1)	44,2 (34,6-54,3)	93,8 (88,5-96,8)	65,5 (45,7-81,4)	92,6 (87,1-95,9)
Prueba positiva de AP tras 3 meses desde el tratamiento quirúrgico	95 (6,2)	36,5 (27,5-46,6)	93,1 (87,7-96,4)	59,3 (39,0-77,0)	90,9 (85,1-94,6)
Prueba positiva de AP tras 6 meses desde diagnóstico o tras 3 meses desde el tratamiento quirúrgico	184 (12,1)	45,2 (35,5-55,2)	91,3 (85,5-94,6)	57,6 (39,4-74,1)	92,4 (86,8-95,8)
Diagnóstico al alta de CM tras 6 meses desde diagnóstico inicial	55 (3,6)	31,7 (23,1-41,7)	99,4 (96,0-100,0)	91,7 (59,8-99,6)	88,8 (83,1-92,9)
Solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico	208 (13,7)	53,8 (43,8-63,6)	84,4 (77,6-89,5)	46,8 (32,4-61,8)	93,8 (88,1-96,9)
Solicitud de gammagrafía ósea tras 6 meses desde el segundo tratamiento oncológico	175 (11,5)	47,1 (37,3-57,1)	86,3 (79,7-91,0)	46,3 (31,0-62,4)	92,0 (86,1-95,6)
Nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado	43 (2,8)	18,3 (11,6-27,3)	100,0 (97,1-99,9)	100,0 (67,9-99,2)	88,9 (83,1-92,9)
Paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado	116 (7,6)	24,0 (16,4-33,6)	96,9 (92,5-98,8)	66,7 (38,7-87,0)	88,1 (82,1-92,3)

^a Calculada a partir de los casos de recidiva diagnosticados en el servicio de oncología como prueba de referencia.

^b Calculada a partir de los casos de recidiva identificados a partir de la revisión manual de historias clínicas de mujeres con cáncer de mama como prueba de referencia.

CM no desarrollará recidivas. Para este supuesto el investigador puede optar por la solicitud de gammagrafías óseas a los 6 meses del primer tratamiento oncológico, que se muestra como la definición operativa con un mayor VPN (93,8%; IC 95%: 88,1-96,9%), con un número asumible de pRCM a revisar de 208, lo que supone el 13,7% de los casos de CM en el RTH. Esto implica que el 86,3% de los casos clasificados como «no recidivas» según esta definición tienen una probabilidad del 93,8% de ser «no recidivas» y no necesitarían ser revisadas. Este algoritmo cuenta con una alta especificidad y VPN, por lo que podría ser muy útil utilizarlo en este sentido.

Este estudio no permite ni está diseñado para el cálculo de incidencia de recidivas, al no disponer de tiempo de seguimiento suficiente ni haberse realizado una revisión exhaustiva de cada caso de cáncer. Además, el algoritmo

está solamente diseñado para la identificación de primeras recidivas, no de recidivas sucesivas.

Entre las limitaciones se encuentran la inclusión de un periodo de estudio relativamente estrecho en el que no se da opción a la aparición de todas las recidivas, sino solo a las más tempranas, siendo necesaria la replicación del estudio utilizando un periodo mayor. Otra de las limitaciones que puede afectar tanto la sensibilidad como a la especificidad obtenidas es que algunas bases de datos empleadas pueden no ser suficientemente exhaustivas y, por tanto, tener pérdidas de seguimiento, entre otros porque la paciente haya acudido a otro centro para diagnosticarse o para tratarse. Sin embargo, al cruzar las bases de manera sucesiva estas posibles pérdidas se minimizan. Otras limitaciones son los posibles errores del RTH, que en los últimos años se han ido

Tabla 3 Número de probables recidivas de cáncer de mama (CM), identificadas como tales según diferentes definiciones operativas compuestas (n = 1.523)

Definición operativa compuesta	N (% sobre total de CM del RTH)	Sensibilidad, % (IC 95%) ^a	Especificidad, % (IC 95%) ^b	Valor predictivo positivo (%) ^b	Valor predictivo negativo (%) ^b
Prueba positiva de AP tras 6 meses desde el diagnóstico y solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico	61 (4,0)	29,8 (21,4-39,7)	98,1 (94,2-99,5)	81,3 (53,7-95,0)	89,7 (84,0-93,6)
Prueba positiva de AP tras 6 meses desde el diagnóstico, solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y diagnóstico al alta de CM tras 6 meses desde diagnóstico inicial	27 (1,8)	16,4 (10,1-25,2)	99,4 (96,0-100,0)	87,5 (46,7-99,3)	86,9 (80,9-91,3)
Prueba positiva de AP tras 6 meses desde el diagnóstico, solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado	19 (1,2)	11,5 (6,4-19,7)	100,0 (97,1-99,9)	100,0 (51,7-98,5)	86,5 (80,5-90,9)
Prueba positiva de AP tras 6 meses desde el diagnóstico, solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado	24 (1,6)	12,5 (7,1-20,8)	100,0 (97,1-99,9)	100,0 (51,7-98,5)	86,5 (80,5-90,9)
Prueba positiva de AP tras 6 meses desde el diagnóstico y solicitud de gammagrafía ósea tras 6 meses desde el segundo tratamiento oncológico	54 (3,5)	27,9 (19,8-37,7)	98,1 (94,2-99,5)	80,0 (51,4-94,7)	89,2 (83,4-93,2)
Diagnóstico al alta de CM tras 6 meses desde diagnóstico inicial y solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico	40 (2,6)	25,0 (17,3-34,6)	99,4 (96,0-99,9)	90,0 (54,1-99,5)	87,9 (82,0-92,1)
Solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado	32 (2,1)	15,4 (9,3-24,1)	100,0 (97,1-99,9)	100,0 (59,8-98,8)	87,4 (81,5-91,7)
Solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico y paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado	34 (2,2)	13,5 (7,8-21,9)	98,1 (94,2-99,5)	70,0 (35,4-91,9)	86,7 (80,7-91,2)
Solicitud de gammagrafía ósea tras 6 meses desde el primer tratamiento oncológico, nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado y paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado	10 (0,7)	6,7 (3,0-13,9)	100,0 (97,1-99,9)	100,0 (19,8-95,1)	84,7 (78,5-89,3)
Diagnóstico al alta de CM tras 6 meses desde diagnóstico inicial y nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado	14 (0,9)	11,5 (6,4-19,7)	100,0 (97,1-99,9)	100,0 (39,6-97,7)	85,6 (79,5-90,1)
Nuevo tratamiento citostático tras 3 meses desde la finalización del último ciclo administrado y paciente nuevamente presentado en el comité de tumores tras 5 meses desde la última vez que fue presentado	13 (0,9)	8,7 (4,3-16,2)	100,0 (97,1-99,9)	100,0 (31,0-96,8)	85,1 (79,0-89,7)

^a Calculada a partir de los casos de recidiva diagnosticados en el servicio de oncología como prueba de referencia.

^b Calculada a partir de los casos de recidiva identificados a partir de la revisión manual de historias clínicas de mujeres con cáncer de mama como prueba de referencia.

corrigiendo, y que el registro de casos de recidivas del servicio de oncología solo permitiera calcular la sensibilidad por no ser un registro exhaustivo. Sin embargo, la disponibilidad de la revisión manual de casos con diagnóstico de CM diagnosticados en 2003 permitió el cálculo del resto de parámetros para la validación.

El establecimiento de los cortes de tiempo para la definición operativa en cada registro no están basados en definiciones preestablecidas, sino en la experiencia clínica y el consenso de expertos, así como en las particularidades intrínsecas de los registros y funcionamiento del centro de estudio, pudiendo afectar a la validez externa. Sin embargo, no existen definiciones operativas consensuadas disponibles, y las empleadas en este estudio pueden ser revisadas en futuras versiones de la herramienta. Por otro lado, la metodología empleada sí puede ser extrapolable a otros centros sanitarios mediante la utilización de códigos propios de cada centro.

Una mayor definición de criterios clínicos y una mayor calidad de los registros sería deseable para la mejor aplicación de los algoritmos definitorios de caso^{7,12}. Por otro lado, la inclusión de las recidivas atendidas en otros hospitales de área sería interesante para captar todas aquellas que se puedan haber perdido en el seguimiento. Esta metodología, que se ha mostrado útil y válida en cada vez un mayor número de estudios^{17,18,24,28-30}, aún debe ser depurada para conseguir una mayor validez en este estudio, pero los resultados obtenidos muestran su potencial para conseguir una herramienta de gran utilidad en la investigación sanitaria. En el contexto actual en el que impera la maximización de la utilidad de los recursos existentes, y dado el gran potencial y la frecuente infrautilización de las bases de datos hospitalarias, la versatilidad de esta herramienta es particularmente interesante, permitiendo adecuar y replicar el algoritmo a otras localizaciones tumorales y otras definiciones operativas, y así adaptarse a las diferentes necesidades e intereses en el ámbito de la investigación de recidivas.

Conflicto de intereses

Los autores declaran que no existe ningún conflicto de intereses.

Agradecimientos

A Eugenio Calciati por su trabajo inicial, que ha facilitado realizar este estudio.

Bibliografía

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Globocan 2008 v1.2, Cancer incidence and mortality worldwide: IARC CancerBase No. 10 [sede Web]. Lyon, France: International Agency for Research on Cancer; 2010 [acceso 28 Feb 2011]. Disponible en: <http://globocan.iarc.fr/>
2. Galceran J, Puigdefàbregas A, Ribas G, Izquierdo A, Pareja L, Marcos-Gragera R. Evolución de la supervivencia del cáncer en Cataluña y comparación con Europa. *Med Clin (Barc)*. 2008;131 Supl 1:19-24.
3. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: Results of EUROcare-5-a population-based study. *Lancet Oncol*. 2014;15:23-34.
4. Macià F, Porta M, Murta-Nascimento C, Servitja S, Guxens M, Burón A, et al. Factors affecting 5- and 10-year survival of women with breast cancer: An analysis based on a public general hospital in Barcelona. *Cancer Epidemiol*. 2012;36:554-9.
5. Fisher B, Anderson S, Fisher ER, Redmond C, Wickerham DL, Wolmark N, et al. Significance of ipsilateral breast tumour recurrence after lumpectomy. *Lancet*. 1991;338:327-31.
6. Holmäng S, Hedelin H, Anderström C, Johansson SL. The relationship among multiple recurrences, progression and prognosis of patients with stages Ta and T1 transitional cell cancer of the bladder followed for at least 20 years. *J Urol*. 1995;153:1823-6.
7. Rondeau V, Mathoulin-Pélissier S, Tanneau L, Sasco AJ, Macgrogan G, Debled M. Separate and combined analysis of successive dependent outcomes after breast-conservation surgery: Recurrence, metastases, second cancer and death. *BMC Cancer*. 2010;10:697.
8. Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. The hospital-based cancer registry cancer registration: Principles and methods. Lyon: International Agency for Research on Cancer; 1991. p. 177-84.
9. Muir CS, Démaret E, Boyle P. The cancer registry in cancer control: An overview. *IARC Sci Publ*. 1985;13-26.
10. Armstrong BK. The role of the cancer registry in cancer control. *Cancer Causes Control*. 1992;3:569-79.
11. Parkin DM. The role of cancer registries in cancer control. *Int J Clin Oncol*. 2008;13:102-11.
12. World Health Organization. National cancer control programmes. Policies and managerial guidelines. 2nd ed Geneva: World Health Organization; 2002. p. 180.
13. De Coster C, Quan H, Finlayson A, Gao M, Halfon P, Humphries KH, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: Report from an international consortium. *BMC Health Serv Res*. 2006;6:77.
14. Tu K, Campbell NR, Chen ZL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Med*. 2007;1:e18-26.
15. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying patients with physician-diagnosed asthma in health administrative databases. *Can Respir J*. 2009;16:183-8.
16. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: Determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care*. 2002;25:512-6.
17. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. *PLoS One*. 2011;6:e21748.
18. Guttmann A, Nakhla M, Henderson M, To T, Daneman D, Cauch-Dudek K, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. *Pediatr Diabetes*. 2010;11:122-8.
19. Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. *Pharmacoepidemiol Drug Saf*. 2009;18:1086-93.
20. Prosser RJ, Carleton BC, Smith MA. Identifying persons with treated asthma using administrative data via latent class modelling. *Health Serv Res*. 2008;43:733-54.
21. Bernatsky S, Lix L, Hanly JG, Hudson M, Badley E, Peschken C, et al. Surveillance of systemic autoimmune rheumatic diseases using administrative data. *Rheumatol Int*. 2011;31:549-54.
22. Roberts CL, Bell JC, Ford JB, Hadfield RM, Algert CS, Morris JM. The accuracy of reporting of the hypertensive disorders of pregnancy in population health data. *Hypertens Pregnancy*. 2008;27:285-97.

23. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care*. 1997;35:932–47.
24. Clèries R, Ribes J, Gálvez J, Melià A, Moreno V, Bosch FX. Cálculo automatizado de la supervivencia relativa vía web. El proyecto WAERS del Instituto Catalán de Oncología. *Gac Sanit*. 2005;19:71–5.
25. Ribes J, Gálvez J, Melià A, Clèries R, Messeguer X, Bosch FX. Automatización de un registro hospitalario de tumores. *Gac Sanit*. 2005;19:221–8.
26. Márquez Cid M, Valera Niñirola I, Chirlaque López MD, Tortosa Martínez J, Párraga Sánchez E, Navarro Sánchez C. Validación de los códigos diagnósticos de cáncer de colon y recto del conjunto mínimo básico de datos. *Gac Sanit*. 2006;20:266–72.
27. Nattinger AB, Laud PW, Bajorunaite R, Sparapani RA, Freeman JL. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. *Health Serv Res*. 2004;39:1733–49.
28. Armstrong K, Long JA, Shea JA. Measuring adherence to mammography screening recommendations among low-income women. *Prev Med*. 2004;38:754–60.
29. Leung KM, Hasan AG, Rees KS, Parker RG, Legorreta AP. Patients with newly diagnosed carcinoma of the breast: Validation of a claim-based identification algorithm. *J Clin Epidemiol*. 1999;52:57–64.
30. Warren JL, Feuer E, Potosky AL, Riley GF, Lynch CF. Use of Medicare hospital and physician data to assess breast cancer incidence. *Med Care*. 1999;37:445–56.
31. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011;64:821–9.
32. WHO. International Classification of Diseases of Oncology (ICD-O). 2nd ed Geneva: World Health Organization; 1990.
33. Altman DG, Bland JM. Diagnostic tests. Sensitivity and specificity. *BMJ*. 1994;11:1552.
34. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ*. 1994;309:102.
35. SPSS., Inc. SPSS Statistics for Windows, version 17.0. Chicago, IL, SPSS Inc; 2008. SPSS Inc. Released SPSS Statistics for Windows. Version 17. 0. Chicago: SPSS Inc; 2008.