PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

MAGISTER EPIDEMIOLOGÍA

FACULTAD DE MEDICINA

VALIDEZ DEL AUTO-REPORTE DE CÁNCER EN UNA POBLACIÓN AGRÍCOLA DE AMÉRICA LATINA

Tesis presentada a la Facultad de Medicina de la Pontificia Universidad Católica de Chile para optar al grado académico de Magíster en Epidemiología.

POR DRA. DANIELA ALFARO SEPÚLVEDA

PROFESORES GUÍA: DRA. CATTERINA FERRECCIO READI

COLABORADORA: DRA. LORENA ROSA

Diciembre 2021

Santiago, Chile

RESUMEN

Introducción: La validez del auto-reporte de cáncer se ha estudiado en países desarrollados mostrando amplia dispersión. Reportamos un estudio de la validez y variabilidad del auto-reporte de cáncer en población adulta de una comuna agrícola de Chile, participantes de la cohorte MAUCO (2014-2020).

Materiales y métodos: De 9.454 participantes de 37 a 75 años, 493 (5,2%) reportaron tener cáncer. Se ingresaron al estudio todos los casos que reportaron cáncer y una muestra pareada por edad, sexo y año de ingreso a la cohorte, de participantes que no reportaron cáncer. Se buscó el diagnóstico y órgano afectado en registros clínicos de tres hospitales. Se calcularon indicadores de calidad para el auto-reporte de cáncer en general y por órgano. Para identificar variables asociadas un verdadero positivo, se realizó una regresión logística múltiple. Se contrastó la precisión para cánceres prevalentes versus incidentes.

Resultados: El valor predictivo positivo (VPP), negativo, sensibilidad y especificidad del auto-reporte de cáncer en general fueron respectivamente: 36%, 93%, 84%, 67%. La sensibilidad del auto-reporte de cáncer de órgano específico (61%) fue menor que la de general (84%). El VPP fue menor para los cánceres prevalentes (28%) que para los incidentes (51%). Se asociaron a un verdadero positivo el sexo masculino, cáncer incidente y los fallecidos.

Conclusión: El auto-reporte de cáncer fue capaz de identificar a la gran mayoría (84%) de los casos de cáncer en esta población de baja escolaridad. Sin embargo, no es confiable para identificar a sujetos sanos: 67% de los que no tenían cáncer en los registros, dijeron no tenerlo.

INTRODUCCIÓN

Los estudios epidemiológicos necesitan datos correctos y precisos. Ejemplos de fuentes de datos son el auto-reporte de los participantes, registros médicos y certificados de defunción. La selección de la fuente depende de la factibilidad técnica y económica de llevar a cabo el estudio, además del tiempo asociado.

La propuesta de este estudio es analizar las limitaciones de la epidemiología observacional del cáncer, como una estrategia para mejorar la validez de los resultados de las investigaciones. En particular, explorar el sesgo del auto-reporte de cáncer.

Auto-reporte de cáncer en general

Para evaluar la validez del auto-reporte de cáncer se compara lo reportado por los participantes con los registros médicos. Las investigaciones que han abarcado este tema presentaron resultados divergentes. La **Tabla 1** muestra la sensibilidad, el valor predictivo positivo (VPP) y el estadístico kappa del auto-reporte de cáncer en general (sin distinguir por tipo de cáncer) reportados en la literatura. La sensibilidad fluctúa entre 21% y 93%,[1–15] el VPP entre 60% y 80%[1–3,5–7,9,12,14] y el kappa entre 0,47 y 0,65. [2,3,6,7,11,13] En cambio, la especificidad y el valor predictivo negativo (VPN) son homogéneos y >97%.[1–7,9–15] Las variaciones en los resultados se relacionan con el diseño del estudio, tipo de población y la temporalidad entre el diagnóstico y el auto-reporte del cáncer.

Los estudios se han realizado en países desarrollados y hasta la fecha, no hay investigaciones en Sudamérica, África, ni en poblaciones agrícolas.[1–16]

Auto-reporte de cáncer según el órgano afectado

Dos estudios[3,5] evaluaron la validez del auto-reporte de cáncer de órgano específico, es decir, cuando se define como verdadero positivo la concordancia entre el órgano afectado por el cáncer según el auto-reporte y el órgano encontrado en los registros. Los resultados fueron: VPP 58% y 79%, y sensibilidad 49% y 74%.[3,5]

A la vez, la sensibilidad del auto-reporte del cáncer varía ampliamente según el órgano involucrado. A modo de ejemplo, la sensibilidad del cáncer de mama es en general >80%,[1–3,5,7–9,11–14,16] mientras que la de cuello uterino es <52%.[2,3,7,11,13,14,16]

Variables asociadas

En globo, las variables asociadas a una mayor validez del auto-reporte de cáncer son una menor edad[2,3,6,7,9,11,13,15] y un menor tiempo entre el diagnóstico y la encuesta.[11,15] El sexo femenino se asocia a una mayor sensibilidad,[6,8,11,12] pero menor VPP.[7,9,11,12] Un nivel educacional alto se relaciona según algunos estudios un mayor *Odds ratio* (OR) de reportar lo correcto,[3,11,14] pero en otros con una menor sensibilidad y VPP.[6,13]

Situación actual en Chile

En Chile no existe un registro nacional de cáncer, y los regionales (Antofagasta, Los Ríos, Bío-Bío) no están disponibles para los investigadores, existiendo solo un reporte que abarcó los años 2003 a 2007.[17] Además, las fichas médicas de servicios de salud públicos son difíciles de conseguir. Tradicionalmente el registro es manual en fichas de papel, y solo recientemente en fichas electrónicas. Dado esto, muchas veces el auto-reporte es la única opción factible para los estudios, siendo también utilizado por la Encuesta Nacional de Salud 2016 - 2017.

El objetivo de este estudio es determinar la validez del auto-reporte de cáncer de participantes de la cohorte MAUCO,[18] entre los años 2014-2020. Además, analizar su variabilidad según factores sociodemográficos de los participantes (sexo, edad, nivel educacional y estado vivo/fallecido) y las características del cáncer (incidente o prevalente, y órgano afectado).

MATERIALES Y MÉTODOS

El estudio fue aprobado por el Comité de Ética de la Pontificia Universidad Católica de Chile.

Es un estudio de información secundaria, de validación del auto-reporte de cáncer en comparación con los registros médicos (*gold-standard*).

Se utilizaron dos estrategias descritas por Fox et al. (2020),[19] para calcular correctamente, desde un punto de vista estadístico, distintos indicadores.

Brevemente, la primera estrategia (Estrategia 1) consiste en configurar la muestra a partir del auto-reporte de cáncer. Se seleccionan sujetos con auto-reporte positivo para cáncer, y se realiza un pareo con sujetos que no auto-reportaron cáncer. Luego, se determina la frecuencia de cáncer según los registros médicos, obteniendo los verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Dada las características del modelo y considerando que no se fija la real prevalencia de la enfermedad, se pueden obtener los valores predictivos.

De manera inversa, la segunda estrategia (Estrategia 2) conforma la muestra a partir de los registros médicos (*gold-standard*). Se seleccionan sujetos con cáncer según los registros médicos, y se realiza un pareo con sujetos sin cáncer en los registros médicos. Luego, se compara con la frecuencia de cáncer según lo que habían auto-reportado los participantes. Dada las características del modelo se puede obtener la sensibilidad y especificidad.

Cohorte MAUCO

El 2014 se inicia la cohorte MAUCO, ubicada en la comunidad agrícola de Molina, Región del Maule. Se reclutaron 9.454 participantes de 37 a 75 años. La mitad de ellos habían sido agricultores y tenían una educación <8 años. Además, 87% eran usuarios del sistema de salud público. Más detalles de la cohorte se han descrito previamente en otras publicaciones.[18]

Muestra

Los criterios de inclusión fueron: participantes de MAUCO, que respondieron la pregunta sobre el antecedente de cáncer y con información sobre ficha clínica revisada o en proceso de revisión. Por diseño no hubo criterios de exclusión.

Siguiendo la metodología de la Estrategia 1, se seleccionaron todos los participantes que auto-reportaron cáncer (n = 493). Se realizó un pareo 1:1, según sexo, edad (± 2 años) y año de ingreso a la cohorte (± 3 años), con participantes que no reportaron cáncer (n = 493).

El tamaño muestral mínimo se calculó con el programa EPIDAT 3.1. Se consideró un VPP = 75%, VPN = 90% y un nivel de significancia = 5%, obteniéndose un n total = 776. Por las posibles pérdidas, se incluyeron a todos los participantes con auto-reporte positivo y su respectivo par (n = 986).

En la Estrategia 2, se seleccionaron todos los participantes con cáncer según los registros médicos (n = 209). Se realizó un pareo 1:1, utilizando los criterios ya mencionados, con participantes sin registros médicos de cáncer (n = 209).

Dado que el tamaño muestral total por disponibilidad fue 418, se calculó la precisión absoluta, considerando una sensibilidad = 70% y una especificidad = 90%, utilizando el programa EPIDAT 3.1. La precisión absoluta de la sensibilidad fue 6% y la de la especificidad = 4%.

Auto-reporte de cáncer

La fuente de información sobre el auto-reporte de los participantes fue la base de datos de MAUCO. En el reclutamiento y en los seguimientos se les preguntó a los participantes: "¿Usted padece o ha padecido de alguna enfermedad que haya sido diagnosticada por un médico o personal de salud? ¿Algún cáncer? Como cáncer de estómago, vesícula u otro cáncer". Si la respuesta era otro cáncer, se les solicitaba especificar.

Para los análisis, se definió "auto-reporte de cáncer en general", cuando un participante reportaba el antecedente de cáncer (si/no). Mientras, "auto-reporte de cáncer de órgano específico" consideraba también el órgano afectado, estableciendo como verdadero

positivo la concordancia entre el órgano afectado por el cáncer, según el auto-reporte y el encontrado en los registros.

Además, se creó la variable "oportunidad diagnóstica", para diferenciar los cánceres auto-reportados en el enrolamiento (cánceres prevalentes) de los cánceres que ocurrieron posterior al inicio de la cohorte (cánceres incidentes).

Registros médicos

En tres hospitales públicos de la zona se buscaron registros de cáncer y el órgano afectado. Se revisaron fichas de hospitalizaciones e informes de Anatomía Patológica. En los Servicios de Anatomía Patológica se almacenan los informes en dos bases de datos electrónicas, creadas recientemente, o como informes impresos. También se buscaron certificados de defunción. En uno de los Servicios de Anatomía Patológica la búsqueda de informes impresos continúa, pero se revisaron las dos bases de datos electrónicas.

La búsqueda de fichas de hospitalizaciones la realizó personal capacitado de MAUCO, distinto a los investigadores. La búsqueda en Anatomía Patológica la realizó personal entrenado del propio servicio. La búsqueda de certificados de defunción la realizó la investigadora DA.

Dado que en Chile el cáncer de cuello uterino cuenta con un programa de tamizaje, se analizó como una entidad diferente al cáncer de útero, considerando también la diferente evolución y prevalencia de ambas neoplasias. No se consideraron como cáncer lesiones pre-neoplásicas ni tumores de carácter benigno. En el caso de las metástasis, se consideró como órgano afectado el del cáncer primario. Se encontraron tres registros sugerentes, pero no explícitos de metástasis. Dado esto, se consideraron como órganos afectados primariamente.

Datos demográficos y socioeconómicos

Utilizando la base de MAUCO, se obtuvo la información respecto a las variables sexo; edad (promedio y rangos 37-52, 53-60 y 61-75 años); nivel educacional (≤ 8, 9-12, ≥ 13 años de estudio); previsión (FONASA, Isapre, Fuerzas Armadas, sin previsión, no

sabe/no contesta); y año de ingreso a la cohorte (desde 2014 hasta 2019). El estado del participante (vivo o fallecido) se obtuvo mediante los certificados de defunción.

Análisis de los datos

Se realizó un análisis descriptivo de los grupos. Para detectar diferencias se utilizó el test t de student (variables continuas) y chi-cuadrado (variables categóricas).

Utilizando la Estrategia 1, se calculó el VPP y VPN del cáncer en general. Luego, se calculó el VPP del cáncer de órgano específico y se comparó con el del cáncer en general. No se calculó el VPN del cáncer de órgano específico, dado que alguien que no auto-reportó cáncer, no tiene un órgano que especificar. Por lo tanto, los verdaderos negativos se mantuvieron constantes. Si el órgano afectado fue más de uno, se consideraba concordante con los registros cuando todos coincidían. Si no se indicó el órgano (n=10), no se consideró concordante.

Para el auto-reporte de cáncer en general, los valores predictivos se calcularon además según los estratos de las siguientes covariables: sexo, rango etario, nivel educacional, y oportunidad diagnóstica.

Para detectar variables asociadas a un verdadero positivo, se construyó una regresión logística múltiple.

Utilizando la Estrategia 2, se calculó la sensibilidad del auto-reporte de cáncer en general y específico. La especificidad se calculó solo para cáncer en general, y no para cáncer de órgano específico, dado que los verdaderos negativos eran constantes. Además, se analizó el estadístico kappa e índice de Youden del auto-reporte de cáncer en general.

También, se calculó la sensibilidad del auto-reporte de cáncer en general según los estratos de distintas covariables, y la sensibilidad de dos cánceres con programas de tamizaje en Chile: mama y cuello uterino.

Los intervalos de confianza 95% de todos los indicadores se calcularon según lo descrito por Collett et al. (1999).[20]

Todos los cálculos se realizaron con el lenguaje de programación R Studio, versión 1.2.5033© (2009-2019 RStudio, Inc) y se utilizó un nivel de significancia = 5%.

RESULTADOS

Estrategia 1 (muestra a partir del auto-reporte)

Las características de la muestra se describen en la **Tabla 2**. 76% eran mujeres, la mediana de la edad fue 57 años (rango intercuartil: 50 – 64 años) y 87% fueron reclutados entre los años 2014 y 2016. Más de la mitad de los participantes (56%) tenía ≤ 8 años de educación. La mayoría estaba afiliado a FONASA y solo 1,72% a una Isapre.

Al comparar las características del grupo de personas que auto-reportó cáncer, con el grupo que no auto-reportó, el sexo, edad y año de reclutamiento fueron iguales, concordantemente con el pareo. No hubo diferencia en las previsiones. Si diferían en el nivel educacional, teniendo el grupo con auto-reporte negativo menos años de educación.

Se encontró que 209 participantes contaban con algún registro clínico de cáncer. En total encontramos 284 registros de cáncer. Algunos participantes tenían más de un órgano afectado o un mismo participante podía tener más de un registro (**Figura 1**). La mayoría de los registros de cáncer correspondió a informes de Anatomía Patológica (n = 159). Los registros de participantes que auto-reportaron cáncer se encontraron más frecuentemente en Anatomía Patológica. En tanto, la mayoría de los registros de participantes que no auto-reportaron cáncer se encontraron en fichas clínicas y certificados de defunción. Hubo 8 casos en que los órganos encontrados eran discordantes según los diferentes registros médicos (**Tabla suplementaria 1**). En 39 participantes hospitalizados, cuyo diagnóstico de cáncer se encontró en Anatomía Patológica o certificados de defunción, no aparecía el registro de cáncer en sus fichas de hospitalización.

Como se muestra en la **Figura 2**, El órgano más auto-reportado fue útero (n = 118), el segundo mama (n = 80), que además obtuvo la mayor cantidad de registros (n = 55), tercero próstata con más auto-reportes (n = 40), seguido por tiroides (n = 35) y piel (n = 32).

En el gráfico se incluyeron las neoplasias intraepiteliales del cuello uterino (NIE), porque pese a ser una condición pre-neoplásica, algunas pacientes lo reportaron como cáncer (n = 21). El auto-reporte de cáncer de útero fue mayor que el de cuello uterino, y este a su

vez mayor que el de NIE. Contrariamente, la frecuencia de los registros, de mayor a menor, fueron NIE, cáncer de cuello uterino y cáncer de útero.

Entre las 493 personas que auto-reportaron cáncer, en 176 se confirmó este diagnóstico (verdaderos positivos) y en 317 (64,3%) no se pudo confirmar el cáncer (falsos positivos). En tanto, entre las 493 personas que no auto-reportaron cáncer, en 460 efectivamente no había registros de cáncer (verdaderos negativos) pero en 33 (6,7%) se encontró este diagnóstico en al menos un registro (falsos negativos). **Tabla suplementaria 2**.

El VPP del auto-reporte de cáncer en general fue 36% [IC 95%: 31% - 40%] y el VPN 93% [IC 95%: 91% - 95%]. El VPP para cáncer de órgano específico fue menor: 27% [IC 95%: 23% - 32%], pero la diferencia no fue estadísticamente significativa (**Tabla suplementaria 3**).

Al no considerar los auto-reportes de NIE, cáncer de cuello uterino y de útero, el VPP del auto-reporte de cáncer en general aumentó a 47% [IC 95%: 41% - 52%], alcanzando significación estadística.

Los valores predictivos del auto-reporte de cáncer en general no variaron de modo importante según sexo, rango etario, nivel educacional y oportunidad de estudio (**Tabla 3**). En todos los análisis el VPN fue alto (>90%) y el VPP fue bajo (<52%). Aunque no alcanzó significación estadística, el VPP fue menor en mujeres (33%) que en hombres (45%). Por edad, el grupo de 37 a 52 años tuvo el menor VPP (21%) alcanzando significación estadística, comparado con los grupos de 53-60 (42%) y 61-75 años (45%). Las personas con ≤ 8 años de estudio tuvieron mayor VPP (42%), en comparación a 9-12 años (31%) y ≥13 años de escolaridad (24%), sin alcanzar significancia estadística. El VPP de los cánceres incidentes (51%) fue mayor que el de los prevalentes (28%), siendo estadísticamente significativos.

Se realizó una regresión logística múltiple, para identificar predictores de un verdadero positivo (encontrar un registro de cáncer, dado que se auto-reporto cáncer). En el modelo final (*Area Under the Curve* = 0,7), las variables significativas fueron sexo, cáncer incidente/prevalente, y estado del participante vivo o muerto al momento de este estudio (**Tabla suplementaria 4**). El OR ajustado de tener un verdadero positivo en las mujeres versus los hombres fue 0,63 [IC 95%: 0,40 - 0,98]. El OR ajustado de los cánceres incidentes en comparación a los prevalentes fue 2,31 [IC 95%: 1,55 - 3,44]. Finalmente,

el OR ajustado de tener un verdadero positivo si el participante falleció durante el seguimiento, en comparación con participantes actualmente vivos, fue 46,12 [IC 95%: 9,81 – 823,89]. El gran intervalo de confianza se explica por el bajo n de participantes que auto-reportaron cáncer y fallecieron (n = 26).

Estrategia 2 (muestra a partir del registro médico)

Las características de la muestra de la Estrategia 2 se describen en la **Tabla 4.** Al igual que en la muestra de la Estrategia 1, el sexo femenino fue el más prevalente (70%). La mayoría (90%) fueron reclutados entre los años 2014 y 2016. Solo 1,44% estaban afiliados a una Isapre y más de la mitad de los participantes (61%) tenían ≤ 8 años de educación. La mediana de la edad fue levemente mayor, 60 años (rango intercuartil 53 – 67 años).

Al comparar las características del grupo de personas con y sin registro de cáncer, no hubo diferencias según sexo, edad, año de reclutamiento, previsión ni nivel educacional.

De las 209 personas que tenían registros de cáncer, 33 no lo auto-reportaron. De las 209 que no tenían registros de cáncer, 68 lo auto-reportaron. La sensibilidad del auto-reporte de cáncer en general fue 84% [IC 95%: 79% - 89%], la especificidad 67% [IC 95%: 61% - 74%], el estadístico kappa 0,51 [IC 95%: 0,42 - 0,61] y el índice de Youden 51 (**Tabla suplementaria 5**).

La sensibilidad del auto-reporte de cáncer de órgano específico fue 61% [IC 95%: 54% - 67%], es decir, menor que la de cáncer en general, siendo la diferencia estadísticamente significativa (**Tabla suplementaria 6**).

Al no considerar los registros de cáncer de cuello uterino y útero, la sensibilidad para el auto-reporte de cáncer en general no varió: 83% [IC 95%: 77% - 89%]. La especificidad aumento a 76% [IC 95% 70% – 81%], pero de forma no estadísticamente significativa.

La **Tabla 5** muestra la sensibilidad y especificidad del auto-reporte de cáncer en general, estratificado por sexo, rango etario, nivel educacional y oportunidad de estudio. Sin importar la variable por la cual se estratificó, la sensibilidad fue alta (≥ 70%) y no hubo diferencias entre los distintos estratos.

La especificidad, al estratificar por rango etario, fue menor en el grupo de 37 a 52 años (45%) y mayor en los grupos de 53-60 (77%) y 61-75 años (74%). La diferencia alcanzó significancia estadística. Además, los cánceres incidentes tuvieron mayor especificidad que los prevalentes (90% vs 74% respectivamente), también de forma estadísticamente significativa.

Se calculó la sensibilidad del auto-reporte de cánceres con programa de tamizaje en Chile: cáncer de mama y de cuello uterino (**Tabla suplementaria 7**). La sensibilidad del auto-reporte de cáncer de mama fue alta, 89% [IC 95%: 77% - 96%], mientras que la de cuello uterino fue baja: 40% [IC 95%: 19% - 64%]. Luego, se consideró el auto-reporte de cáncer de útero y NIE como sinónimos del auto-reporte del cáncer de cuello uterino, aumentando la sensibilidad a 80% [IC 95%: 56% - 94%]. Cabe destacar que el número de registros de cáncer de cuello uterino fue bajo (n = 20).

DISCUSIÓN

Este estudio exploró la validez del auto-reporte de cáncer en una población agrícola, con bajo nivel educacional y usuarios del sistema de salud público. Este es el primer estudio realizado en Chile y Latinoamérica, y también el primero a nivel internacional que utilizó una población con estas características.

El VPP obtenido en este estudio fue menor que el de estudios internacionales (36% versus > 60%).[1,2,5,7–9,11,12,14] El VPN fue alto, pero menor a lo reportado por series internacionales (93% versus >97%).[2,7–9,11,12,14] La sensibilidad fue alta (84%) y similar a la de otros estudios.[9,12] La especificidad fue notablemente menor, 67% en comparación a >97% (**Tabla suplementaria 8**).[1–9,11,12,14,15]

En esta población agrícola y de baja escolaridad, hubo un sobre-reporte del antecedente de cáncer. Consecuentemente, la probabilidad de que un paciente con cáncer sea identificado como tal utilizando el auto-reporte, es alta (sensibilidad = 84%), pero la especificidad fue baja 67%. Esto probablemente se debe a que las personas consideran como cáncer entidades benignas. El fenómeno se ilustró cuando se restó de los análisis los registros de cáncer de útero y cuello uterino, ya que algunas mujeres que tenían NIE las reportaron como cáncer. Al hacer esto, la especificidad aumentó a 76%.

La probabilidad de que un sujeto que auto-reporta cáncer, efectivamente tenga cáncer es baja (VPP = 36%). Esto concuerda con la baja especificidad encontrada, ya que una proporción considerable de los participantes que dijeron tener cáncer, probablemente tenían entidades pre-neoplásicas, benignas o no tenían cáncer. Nuevamente, esto se comprueba al excluir el auto-reporte de cáncer de cuello uterino, útero y NIE, aumentando el VPP a 47%. La implicancia de esto es que cuando un participante reporta el antecedente de cáncer, se deben buscar estrategias de *triage* para determinar si efectivamente es o fue un cáncer. Por ejemplo, indagar si efectivamente es una entidad de carácter maligno, una pre-neoplásica o una de carácter benigno. También preguntar por el lugar de tratamiento, tratamiento recibido, tipos de exámenes de confirmación, y eventualmente confirmar la información con algún registro clínico.

La probabilidad de que el auto-reporte detecte si el participante tiene o ha tenido algún tipo de cáncer (84%), es mayor que la probabilidad de identificar el órgano específico afectado por el cáncer (61%). Posiblemente esto se deba a que reportar el órgano es más exigente y requiere un mayor conocimiento de la enfermedad.

Cuando el antecedente de cáncer fue reportado por los más jóvenes (37 a 52 años), la capacidad del auto-reporte para identificar a sujetos sanos (45%) fue menor en comparación a los grupos etarios mayores (77% en el grupo de 53-60 años y 74% en el de 53-60 años). La tendencia del grupo más joven a reportar un cáncer cuando eran sanos, puede deberse a la mayor presencia de lesiones pre-neoplásicas, dado que ciertos tipos de cáncer requieren más años para poder desarrollarse.

Si el participante reportaba tener un cáncer durante el seguimiento, la probabilidad de que efectivamente fuese enfermo era mayor que si lo reportaba en el reclutamiento: 51% versus 28%. Concordantemente, el auto-reporte fue más eficaz para detectar a personas sanas si el cáncer era incidente (90%), que si era prevalente (74%). Creemos que los cánceres incidentes están sujetos a un menor sesgo de memoria, en comparación a los prevalentes, cuyo diagnóstico puede haber sido muchos años atrás.

La chance de reportar cáncer y efectivamente presentar el diagnóstico era menor si el sexo era femenino, en compasión al masculino (OR ajustado 0,63). Esto se puede explicar por la confusión entre cáncer de útero, cáncer de cuello uterino y NIE, entidades que correspondían al 40% de los cánceres auto-reportados por las mujeres. Además, la concordancia entre reportar el antecedente de cáncer y presentarlo en los registros era mayor si el cáncer era incidente, en comparación a los prevalentes (OR ajustado 2,31), y mayor si el participante falleció durante el seguimiento (OR ajustado 46,12). Nuevamente, creemos que la mayor concordancia si el cáncer era incidente se debe a un menor sesgo de memoria. Respecto al estado del participante, si este falleció, es posible que el cáncer estuviese en una etapa más avanzada. Por lo tanto, el participante podría haber estado más consciente de su diagnóstico.

En este estudio, la capacidad del auto-reporte para identificar casos de cáncer de mama fue alta (89%) y concuerda con la literatura.[1,2,16,3,6–9,12–14] Mientras, el auto-reporte no fue confiable para identificar casos de cáncer de cuello uterino, ya que solo 40% de las que tenían el diagnóstico, reportaron tenerlo. Esto fue similar a lo obtenido en estudios internacionales (**Tabla suplementaria 9**).[1,3,6,7,12,16]

Al analizar la frecuencia de los auto-reportes de cáncer de útero, cuello uterino y NIE, la distribución es contraria a la prevalencia esperada[21,22] y a la encontrada en los registros médicos. Esto corrobora la ya mencionada confusión respecto a la diferencia entre útero y cuello uterino, al igual que entre una entidad benigna (NIE) y una maligna.

En esta misma línea, al considerar el auto-reporte de NIE y de cáncer de útero como sinónimos de cáncer de cuello uterino, la probabilidad de que el diagnóstico de cáncer de cuello fuera identificado a través del auto-reporte, aumento de 40% a 80%. Sin embargo, no alcanzó significancia estadística. Esto último se puede deber al bajo número de registros de cáncer de cuello uterino (n = 20).

Respecto a la calidad de la información, las fichas de hospitalizaciones estaban escritas a mano y eran guardadas en una bodega, lo que da pie a extravíos, pérdidas y deterioro de las hojas, sin tener respaldo de la información. Similarmente, la mayoría de los informes de Anatomía Patológica estaban impresos, sin respaldo electrónico, y el resto en dos bases electrónicas independientes, por lo que la información estaba dispersa. Por ley, los informes con más de diez años de antigüedad pueden ser eliminados. En este estudio, el informe más antiguo de Anatomía Patológica fue escrito en 1997. Todo lo mencionado anteriormente puede ser perjudicial para el paciente, dada la posibilidad de perder antecedentes tan relevantes como un cáncer.

Como se mencionó, 40 pacientes que tenían cáncer fueron hospitalizados, pero el antecedente no se encontró en las fichas de hospitalización. Esto refleja una calidad sub-óptima de la información registrada.

Además, puede que actualmente el rol educativo del personal de la salud sea deficiente, dada la confusión de los participantes respecto al sitio afectado por el cáncer y la diferencia entre una entidad benigna y una maligna.

Una de las limitaciones de este estudio es que la búsqueda de informes impresos de un Servicio de Anatomía Patológica continúa en proceso. Sin embargo, ya sea realizó la búsqueda electrónica y todos los pacientes se buscaron en fichas de hospitalizaciones, certificados de defunción y en bases electrónicas e informes impresos de otro Servicio de Anatomía Patológica. Otra limitación es que al utilizar fichas de hospitalizaciones se pueden encontrar casos de cánceres más graves, y por lo tanto, participantes con más conciencia de su diagnóstico. También, ciertos tipos de cánceres no se biopsian cuando

están en estados avanzados, como el de pulmón, y por lo tanto, no se encontrará un informe en Anatomía Patológica. Solo se revisaron registros del sistema de salud público. No obstante, menos de 2% de los participantes estaban afiliados a una Isapre y 1,72% pertenecían a las Fuerzas Armadas y del Orden. Finalmente, dado el contexto de la pandemia por COVID-19, el acceso a fuentes de registro se vio limitado.

Una fortaleza es lo innovador del estudio, tanto por el tema como por la doble metodología, que permitió calcular múltiples parámetros. Además, el tamaño muestral fue grande. Un aporte del estudio es el conocimiento generado, dado que es el primero que evaluó la validez del auto-reporte de cáncer en Chile. Además, se creó una nueva base de datos para MAUCO, con cánceres confirmados por registros. Esto es importante para investigaciones a futuro, dada las limitaciones expuestas del auto-reporte para obtener el antecedente de cáncer. Además, se evidenció la necesidad de un registro nacional de cáncer y de un registro médico de calidad. Finalmente, se evidenció una posible carencia en el rol educativo actual del personal de la salud.

REFERENCIAS

1. Stavrou E, Vajdic CM, Loxton D, Pearson SA. The validity of self-reported cancer diagnoses and factors associated with accurate reporting in a cohort of older Australian women. *Cancer Epidemiol*. 2011;35(6):e75-80.

2. Loh V, Harding J, Koshkina V, Barr E, Shaw J, Magliano D. The validity of self-reported cancer in an Australian population study. *Aust N Z J Public Health*. 2014;38(1):35-8.

3. Cowdery SP, Stuart AL, Pasco JA, Berk M, Campbell D, Williams LJ. Validity of self-reported cancer: Comparison between self-report versus cancer registry records in the Geelong Osteoporosis Study. *Cancer Epidemiol*. 2020;68:101790.

4. Schrijvers CT, Stronks K, van de Mheen DH, Coebergh JW, Mackenbach JP. Validation of cancer prevalence data from a postal survey by comparison with cancer registry records. *Am J Epidemiol.* 1994;139(4):408-14.

5. Bergmann MM, Calle EE, Mervis CA, Miracle-McMahill HL, Thun MJ, Heath CW. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol*. 1998;147(6):556-62.

6. Koller KR, Wilson AS, Asay ED, Metzger JS, Neal DE. Agreement between self-report and medical record prevalence of 16 chronic conditions in the Alaska EARTH Study. *J Prim Care Community Health.* 2014;5(3):160-5.

7. Nash SH, Day G, Hiratsuka VY, Zimpelman GL, Koller KR. Agreement between self-reported and central cancer registry-recorded prevalence of cancer in the Alaska EARTH study. *Int J Circumpolar Health.* 2019;78(1):1571383.

8. Yoshinaga A, Sasaki S, Tsugane S. JPHC Study Group (Japan Public Health Center). Sensitivity of self-reports of cancer in a population-based prospective study: JPHC Study Cohort I. *J Clin Epidemiol*. 2001;54(7):741-6.

9. Inoue M, Sawada N, Shimazu T, et al. Japan Public Health Center-based Prospective Study Group. Validity of self-reported cancer among a Japanese population: recent results

from a population-based prospective study in Japan (JPHC Study). *Cancer Epidemiol.* 2011;35(3):250-3.

10. Cho LY, Kim CS, Li L, et al. Validation of self-reported cancer incidence at follow-up in a prospective cohort study. *Ann Epidemiol*. 2009;19(9):644-6.

11. Cho S, Shin A, Song D, et al. Validity of self-reported cancer history in the health examinees (HEXA) study: A comparison of self-report and cancer registry records. *Cancer Epidemiol*. 2017;50(Pt A):16-21.

12. Berthier F, Grosclaude P, Bocquet H, Faliu B, Cayla F, Machelard-Roumagnac M. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *Br J Cancer.* 1997;75(3):445-7.

13. Manjer J, Merlo J, Berglund G. Validity of self-reported information on cancer: determinants of under- and over-reporting. *Eur J Epidemiol*. 2004;19(3):239-47.

14. Navarro C, Chirlaque MD, Tormo MJ, et al. Validity of self reported diagnoses of cancer in a major Spanish prospective cohort study. *J Epidemiol Community Health*. 2006;60(7):593-9.

15. Desai MM, Bruce ML, Desai RA, Druss BG. Validity of self-reported cancer history: a comparison of health interview data and cancer registry records. *Am J Epidemiol*. 2001;153(3):299-306.

16. Parikh-Patel A, Allen M, Wright WE. California Teachers Study Steering Committee. Validation of self-reported cancers in the California Teachers Study. *Am J Epidemiol.* 2003;157(6):539-45.

17. Ministerio de Salud. Primer informe de registros poblacionales de cáncer de Chile, quinquenio 2003-2007. 2012.

18. Ferreccio C, Huidobro A, Cortés S, et al. Cohort Profile: The Maule Cohort (MAUCO). *Int J Epidemiol*. 2020;49(3):760-1.

19. Fox MP, Lash TL, Bodnar LM. Common misconceptions about validation studies. Int *J Epidemiol*. 2020;49(4):1392-6.

20. Collett D (1999). Modelling Binary Data. Chapman & Hall/CRC, Boca Raton Florida, p. 24.

21. Ministerio de Salud. Plan Nacional de Cáncer 2018-2028. 2019.

22. Ministerio de Salud. Guías Clínicas AUGE Cáncer Cérvico Uterino. 2015.

TABLAS

Tabla 1. Sensibilidad y valores predictivos positivos del auto-reporte de cáncer en general, reportados en estudios internacionales, según el año de publicación, país, población de estudio y oportunidad de estudio del cáncer (incidente o prevalente).

Referencia y año	País	Población	Oportunidad de estudio [a] (incidente o prevalente)	Sensibilidad (%)	VPP [b] (%)
[2]2014	Australia	General	Incidente y prevalente	71	66
[3]2020	Australia	General	Incidente y prevalente	64	60
[9]1998	EEUU [c]	General	Incidente	93	75
[8]2001	Japón	General	Incidente	36	
[5]2011	Japón	General	Incidente	53	60
[10]2009	R. de Corea [d]	General	Incidente	40	
[4]1994	Holanda	General	Prevalente	55	
[15]2001	EEUU [c]	General	Prevalente	61	
[13]2004	Suecia	General	Prevalente	53	
[14]2006	España	General	Prevalente	58	70
[1]2011	Australia	Específica [e]	Incidente	88	80
[6]2014	EEUU [c]	Específica [f]	Prevalente	64	62
[7]2019	EEUU [c]	Específica [f]	Prevalente	87	61
[11]2017	R. de Corea [d]	Específica [g]	Prevalente	72	
[12]1997	Francia	Específica [h]	Prevalente	21	

[a] Oportunidad de estudio: Evaluación del auto-reporte de cánceres reportados en el enrolamiento (prevalentes) o cánceres que ocurrieron posterior al inicio de la cohorte (incidentes); [b] VPP = valor predictivo positivo; [c] EEUU = Estados Unidos; [d] R. de Corea = República de Corea; [e] Mujeres adultas mayores; [f] Nativos de Alaska; [g] Asistentes a centros de salud y hospitales formadores; [h] Adultos mayores.

Tabla 2. Caracterización de los participantes de MAUCO (2014 - 2020) que auto-reportaron cáncer y los que no auto-reportaron cáncer, respecto al sexo, edad, año de ingreso a la cohorte, nivel educacional y previsión.

	Total (n=986)	Auto-reporte cáncer (+) (n= 493)	Auto-reporte cáncer (-) (n=493)	Valor-p entre auto-reporte (+) y (-)
	(%)			
Sexo femenino	76,06	76,1	76,1	1
Edad, mediana (rango intercuartil)	57 (50, 64)	57 (50, 64)	57 (49, 65)	1
Año de reclutamiento				
2014	26,97	26,17	27,79	0,08
2015	25,35	25,35	25,35	
2016	34,17	33,26	35,09	
2017	3,95	3,65	4,25	
2018	4,56	4,62	2,66	
2019	4,96	7,09	2,83	
Nivel Educacional (años)				
≤ 8 años	56,19	51,3	61,1	**0,007**
9-12 años	32,76	36,1	29,4	
≥ 13 años	10,95	12,6	9,33	
Previsión				
FONASA	82,76	82,2	83,4	0,80
B	38,95	40,24	37,80	
Isapre	1,72	2,23	1,22	
Otro/no sabe	15,5	15,6	15,4	

Los grupos fueron pareados por sexo, edad y año de ingreso a la cohorte. Los valores son presentados como porcentajes, salvo que se indique lo contrario. Los valores-p fueron obtenidos con los test t de student (variables continuas) y chi-cuadrado (variables categóricas). Los valores-p en negrita son estadísticamente significativos = valor-p < 0,05.

Tabla 3. Valores predictivos positivos y negativos del auto-reporte de cáncer en general de participantes de la cohorte MAUCO (2014 - 2020), estratificado por sexo, rango etario, nivel educacional y oportunidad de estudio.

	Valor predictivo positivo	Valor predictivo negativo
	% [IC 95%]	
Sexo		
Femenino	33 [28 - 38]	94 [91 - 96]
Masculino	45 [36 - 54]	92 [85 - 96]
Rango etario (años) [a]		
37-55	**21 [15 - 28]**	95 [90 - 98]
56-68	**42 [34 - 51]**	95 [90 - 98]
69-75	**45 [38 - 53]**	91 [85 - 94]
Nivel Educacional (años)		
≤ 8	42 [36 - 48]	93 [90 - 96]
9-12	31 [24 - 38]	94 [89 - 98]
≥ 13	24 [14 - 37]	91 [79 - 98]
Oportunidad de estudio		
Incidente	**51 [43 - 59]**	93 [91 - 95]
Prevalente	**28 [23 - 33]**	93 [91 - 95]

El cálculo de los intervalos de confianza se realizó según lo descrito por Collett et al. (1999)[20]; Las variables en negrita tienen valores predictivos positivos con diferencias estadísticamente significativas entre los estratos. [a] Rango etario según edad al reclutamiento.

Tabla 4. Caracterización de los participantes de MAUCO (2014 – 2020) con registro de cáncer y sin registro de cáncer, respecto al sexo, edad, año de ingreso a la cohorte, nivel educacional y previsión.

	Total (n=418)	Registro médico cáncer (+) (n=209)	Registro médico cáncer (-) (n=209)	Valor-p entre Registro médico (+) y (-)
		(%)		
Sexo femenino	69,85	69,86	69,86	1
Edad años, mediana (rango intercuartil)	60 (53-67)	60 (53-66)	60 (52-67)	0,9827
Año de reclutamiento				
2014	28,71	27,75	29,67	0,9789
2015	26,32	27,27	25,36	
2016	34,69	33,97	35,41	
2017	5,02	5,74	4,31	
2018	2,39	2,39	2,39	
2019	2,87	2,87	2,87	
Nivel Educacional (años)				
≤ 8 años	60,48	60,76	60,19	0,7686
9-12 años	29,40	30,14	28,64	
≥ 13 años	10,12	9,09	11,17	
Previsión				
FONASA	84,92	87,08	82,78	0,2185
B	41,35	46,41	36,06	
Isapre	1,44	0,96	1,92	
Otro/no sabe	13,63	11,96	15,3	

Los grupos fueron pareados por sexo, edad y año de ingreso a la cohorte. Los valores son presentados como porcentajes, salvo que se indique lo contrario. Los valores-p fueron obtenidos con los test t de student (variables continuas) y chi cuadrado (variables categóricas).

Tabla 5. Sensibilidad y especificidad del auto-reporte de cáncer en general de participantes de la cohorte MAUCO (2014-2020), estratificado por sexo, rango etario, nivel educacional y oportunidad de estudio.

	Sensibilidad	Especificidad
	% [IC 95%]	
Sexo		
Femenino	84 [77 - 90]	64 [55 - 71]
Masculino	84 [73 - 92]	76 [64 - 86]
Rango etario (años) [a]		
37-52	80 [66 – 91]	**45 [32 - 59]**
53-60	88 [80 - 94]	**77 [66 - 85]**
61-75	81 [70 - 90]	**74 [62 - 84]**
Nivel Educacional (años)		
≤ 8	83 [76 - 89]	74 [66 - 82]
9-12	87 [77 - 94]	61 [47 - 73]
≥ 13	79 [54 - 94]	43 [23 - 66]
Oportunidad de estudio		
Incidente	70 [61 - 78]	**90 [85 - 95]**
Prevalente	71 [62 - 79]	**74 [67 - 80]**

El cálculo de los intervalos de confianza se realizó según lo descrito por Collett et al. (1999)[20]; Las variables en negrita tienen valores predictivos positivos con diferencias estadísticamente significativas entre dos o más estratos; [a]Rango etario según edad al reclutamiento.

FIGURAS

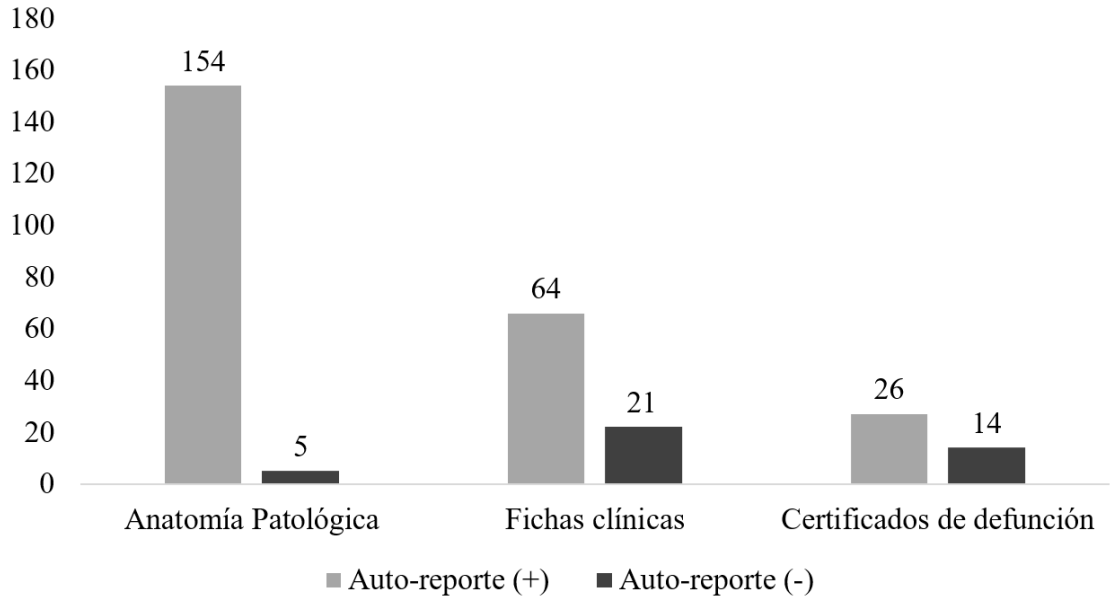

Figura 1. Gráfico de barras. Frecuencia de registros médicos de cáncer de participantes de la cohorte MAUCO (2014 - 2020), según la fuente de detección (Anatomía Patológica, fichas de hospitalización y certificados de defunción). Las barras grises corresponden a registros de participantes que auto-reportaron cáncer y las negras corresponden a participantes que no reportaron cáncer.

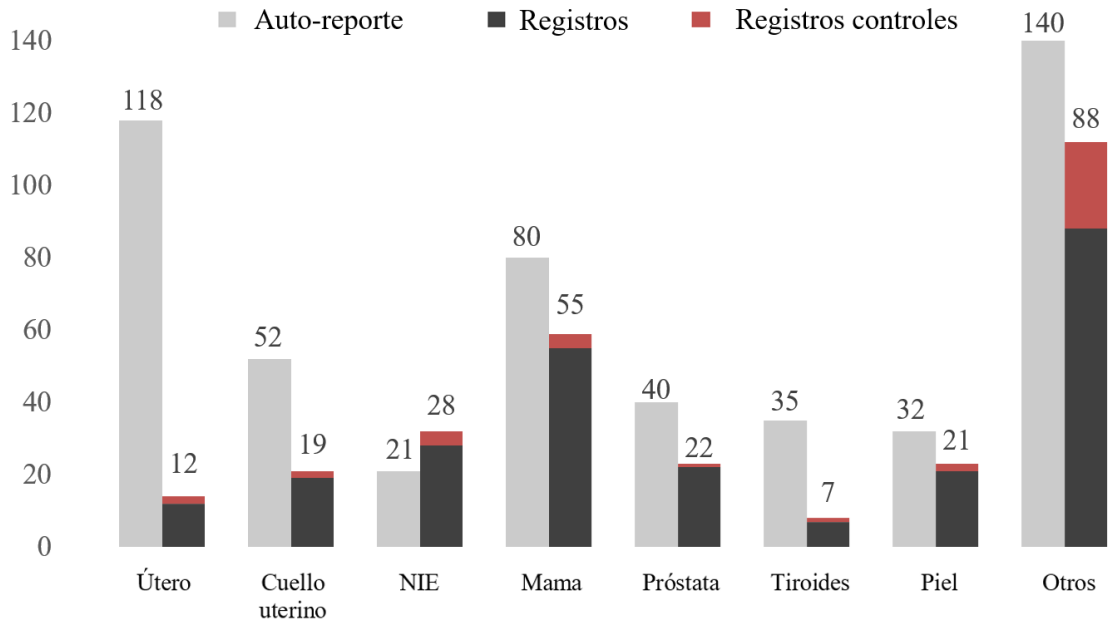

Figura 2. Gráfico de barras. Frecuencia de ciertos órganos afectados por el de cáncer, de participantes de la cohorte MAUCO (2014-2020), según el auto-reporte (barras grises) y según los registros médicos (barras negras-rojas). El segmento rojo representa visualmente la proporción de registros médicos pertenecientes a personas que no reportaron tener cáncer, y el negro a personas que reportaron el antecedente de cáncer. Pese a ser una lesión pre-neoplásica, se incluyeron las neoplasias intraepiteliales (NIE) de cuello uterino, dado que algunas participantes lo reportaron como un tipo de cáncer.