

Impacto del Modelo de Clasificación Oncológica en la Lista de Espera no GES Quirúrgica Nacional

Análisis multi-dimensional con evidencia Kaplan-Meier

AUTOR/A

Paulo Villarroel Tapia, Departamento de Análisis e Información para la Gestión — DIGERA MINSAL

FECHA DE PUBLICACIÓN

22 de marzo de 2026

1 Resumen Ejecutivo

Este reporte analiza el impacto de un modelo de clasificación oncológica sobre la Lista de Espera (LE) quirúrgica nacional No GES, utilizando un universo de **732.837 casos únicos** provenientes de 9 cortes mensuales (junio 2025 a febrero 2026).

La evidencia muestra que el modelo no solo clasifica texto clínico, sino que identifica riesgo clínico real y se asocia a cambios medibles en la trayectoria de los pacientes dentro del sistema de salud.

En términos de discriminación, el modelo identifica un subgrupo pequeño — 3.8% de la lista de espera — que concentra un riesgo significativamente mayor, con una mortalidad **7.5 veces superior** a la del resto de los pacientes (33.1 vs 4.4 por 1.000; $p < 0.001$). Esta diferencia se observa desde etapas tempranas, lo que confirma que la clasificación captura señal clínica relevante.

Esa identificación se traduce en cambios operativos: los pacientes sospechosos se resuelven **4.6 veces más rápido** que los no sospechosos (157 vs 720 días; $p < 0.001$), y aquellos correctamente detectados acceden a cirugía hasta **4.8 veces antes** que los no detectados (206 vs 986 días). Es decir, el modelo no solo predice riesgo, sino que la evidencia sugiere que reordena la respuesta del sistema.

Sin embargo, los pacientes sospechosos no validados por los Servicios de Salud presentan peor supervivencia que los confirmados (mortalidad a 6 meses **3.20% vs 2.60%**; log-rank $p = 0.0027$; HR estratificado por SS = 1.24, $p = 0.0123$). Esta asociación admite dos lecturas complementarias: la validación podría ser protectora al facilitar el acceso al circuito oncológico, pero también podría reflejar que los pacientes más graves progresan antes de que el ciclo de validación los alcance. Actualmente, 6.379 casos (23.2%) permanecen sin revisión.

El análisis de resultados clínicos revela que la velocidad de acceso no necesariamente refleja eficiencia. Los pacientes no validados se operan más rápido (97 vs 206 días), pero presentan mayor mortalidad precoz post-cirugía en todas las ventanas temporales (≤ 30 , ≤ 60 , ≤ 90 días). En ambos grupos, los pacientes que fallecen son aquellos que acceden a cirugía en menor tiempo. La rapidez opera como proxy de gravedad clínica, no como indicador de desempeño del sistema.

Finalmente, el modelo muestra una mejora sostenida en su desempeño, con un aumento del valor predictivo positivo desde 71.5% a 81.1% en seis meses, reflejando un círculo virtuoso entre validación clínica, calidad de datos y aprendizaje del sistema.

En conjunto, los resultados muestran que el modelo sí genera valor, pero no como herramienta aislada. Su impacto radica en su capacidad de reordenar la lista de espera en función de riesgo clínico, permitiendo que los pacientes más graves accedan antes — y en mejores condiciones — al circuito oncológico.

La principal conclusión es estructural: el cuello de botella del sistema no está en la detección, sino en la capacidad de procesar oportunamente los casos detectados. Sea porque la validación es protectora o porque los casos no alcanzados son los más graves, la implicación es la misma: el sistema debe ser más rápido en procesar los casos detectados.

Fecha de referencia: 2026-03-21 | **Cortes:** junio 2025 — febrero 2026

2 ¿Sirve el modelo?

El modelo sí sirve, no solo porque clasifica bien, sino porque identifica un subgrupo pequeño de pacientes con riesgo clínico marcadamente mayor. En una LE de 732.837 casos únicos, clasifica como sospechosos a solo 27.515 pacientes (3.8%), pero ese grupo concentra una mortalidad **7.5 veces superior** a la del resto (33.1 vs 4.4 por 1.000). Esto sugiere que el modelo no solo reconoce texto clínico, sino que captura señal relevante de gravedad.

Ese desempeño además es consistente con un uso de priorización. En la cohorte validada, el modelo alcanza 95% de sensibilidad, 95.4% de especificidad, 72.9% de VPP y 99.3% de VPN. En términos operativos, esto significa que logra detectar casi todos los casos oncológicos reales, concentrando la revisión clínica en una fracción acotada de la lista de espera.

La utilidad del modelo también se observa en la trayectoria asistencial. Los pacientes sospechosos se resuelven en una mediana de 157 días, frente a 720 días en los no sospechosos. Entre los oncológicos confirmados, los casos detectados por el modelo acceden a cirugía en 206 días, comparado con 986 días en los no detectados; además, la proporción que efectivamente llega a cirugía es mayor (53% vs 36.3%). En conjunto, estos resultados son consistentes con que la detección temprana se asocia a ingreso más rápido al circuito clínico relevante.

El principal límite no parece estar en la detección, sino en la respuesta del sistema posterior a ella. Actualmente, 6.379 sospechosos (23.2%) permanecen sin validar por los Servicios de Salud, y ese grupo presenta peor supervivencia que los VP confirmados (3.20% vs 2.60% de mortalidad a 6 meses; HR 1.24 ajustado por SS). Esta asociación podría reflejar que la validación facilita el acceso al circuito oncológico, o que los casos más graves progresan antes de ser validados (ver [SD4b](#)). En ambos casos, el valor del modelo depende de su articulación con

validación clínica oportuna.

3 Contexto y Metodología

3.1 El modelo de clasificación oncológica

El modelo de clasificación oncológica del sistema SIGTE utiliza procesamiento de lenguaje natural (NLP), modelamiento matemático y un clasificador XGBoost para analizar el texto clínico de cada caso en la Lista de Espera Quirúrgica (LE IQ) nacional y asignar una clasificación:

- **Sospechoso:** El texto clínico sugiere patología oncológica
- **No sospechoso:** El texto no sugiere patología oncológica
- **Indeterminado:** No se pudo clasificar con confianza suficiente

El sistema opera actualmente en **Fase 1 (screening centralizado)**, donde el modelo procesa retrospectivamente el texto clínico acumulado en la LE IQ nacional. El flujo operativo es: texto clínico → modelo algorítmico → alerta al Servicio de Salud (SS) → validación clínica por el SS → intervención/tratamiento.

Se proyecta una **Fase 2 (clasificación en origen)**, en la que la identificación oncológica se realice directamente desde el registro clínico en los establecimientos de salud. En esta fase, el modelo pasará a cumplir un rol de aseguramiento de calidad, verificando que casos marcados como no oncológicos sean efectivamente no oncológicos.

3.2 Fuentes de datos

Se utilizaron **9 cortes mensuales** de la LE IQ nacional, abarcando el período de **junio 2025 a febrero 2026**. Tras deduplicar por identificador único (primer registro por caso), se obtuvo un universo de **732.837 casos únicos**. El estado actual de cada caso (egreso, defunción, resultado de validación SS) se actualizó desde la base de datos institucional SIGTE.

3.3 Metodología multi-corte

Se utilizan 9 cortes mensuales en vez de uno solo porque la LE IQ es dinámica: casos que egresan entre cortes desaparecen de snapshots posteriores. La estrategia multi-corte permite capturar el universo completo de casos que han transitado por la LE en el período de estudio.

3.4 Definición de caso oncológico

Para este análisis de impacto, se define como **oncológico (positivo)** a los casos cuya validación por SS corresponde a:

- Diagnóstico
- Tratamiento
- Rehabilitación

Se **excluyen los Antecedentes** oncológicos de la definición de positivo, ya que representan pacientes con historia previa de cáncer pero sin proceso diagnóstico o terapéutico activo. Esta decisión difiere de la definición del documento de metodología de métricas (que incluye Antecedentes como oncológico) y se justifica porque el objetivo es medir el impacto sobre pacientes en proceso clínico activo.

4 Métricas del Modelo

4.1 Matriz de confusión

La siguiente tabla muestra la matriz de confusión del modelo sobre los 78.122 casos validados por SS con ingreso desde junio 2025 (sin arrastre), utilizando la definición sin Antecedentes.

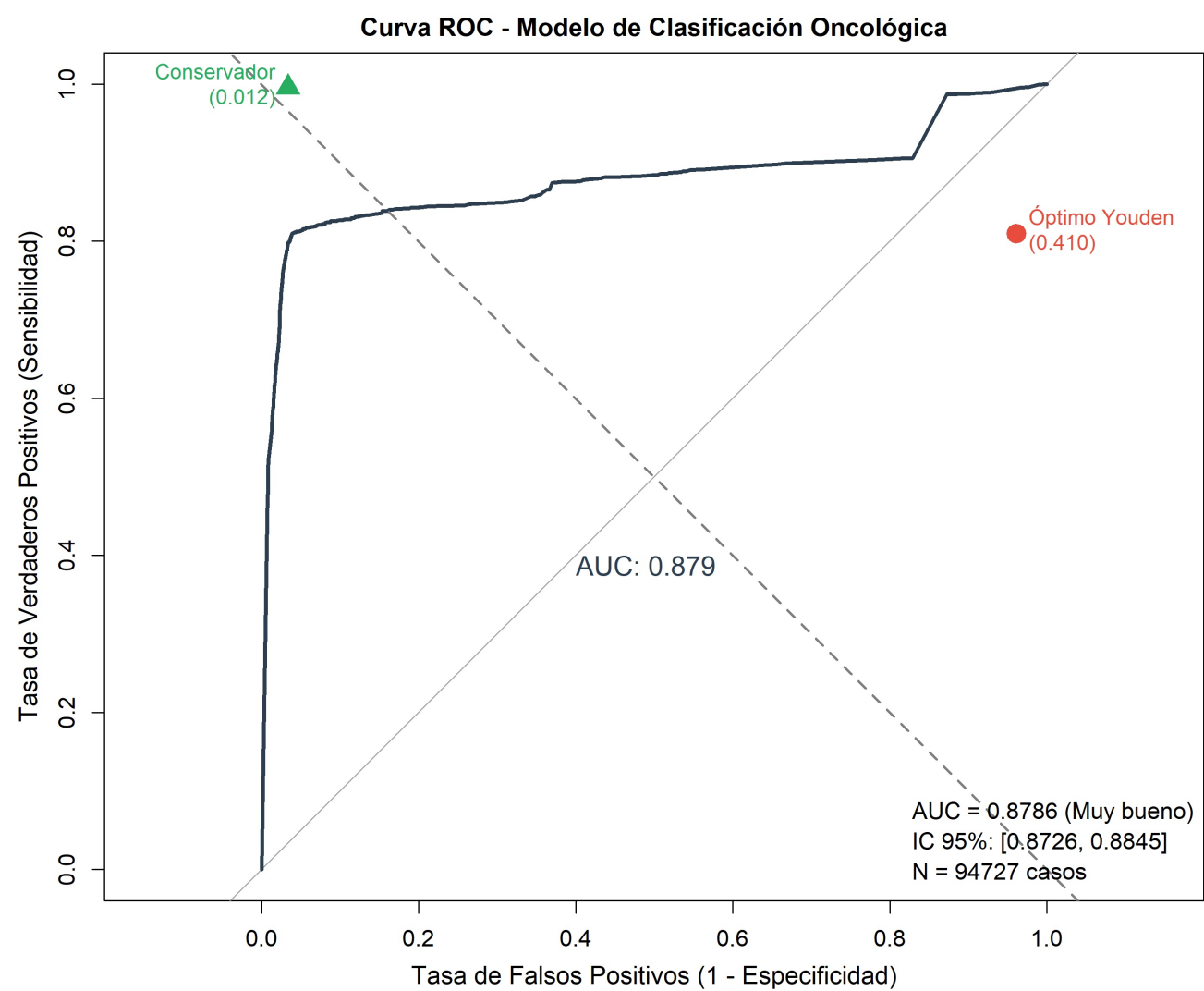
Matriz de confusión		
Clasificación del modelo vs validación del Servicio de Salud (ingresos desde junio 2025)		
Modelo	SS: Oncológico	SS: No oncológico
Sospechoso	8.478	3.158
No sospechoso	448	66.038
VP = 8.478 FP = 3.158 FN = 448 VN = 66.038 Total = 78.122 casos validados		

4.2 Métricas de desempeño

Métricas de desempeño del modelo			
Calculadas sobre casos con ingreso desde junio 2025			
Métrica	Valor (%)	Categoría	Interpretación clínica
Sensibilidad	95.0	Excelente ($\geq 90\%$)	95 de cada 100 oncológicos detectados
Especificidad	95.4	Excelente ($\geq 95\%$)	95 de cada 100 no oncológicos descartados
VPP	72.9	Excelente ($\geq 70\%$)	73 de cada 100 alertas son correctas
VPN	99.3	Excelente ($\geq 99\%$)	99 de cada 100 descartes son correctos
F2-Score	89.6	Bueno (80-89)	Pondera sensibilidad 2x más que VPP (prioriza detección)
MCC	80.8	Excelente (≥ 70)	Correlación predicción vs realidad (robusto ante desbalance)

🔔 **Nota de diseño:** El modelo está calibrado para ser deliberadamente conservador (“alarmista”): ante la duda, clasifica un caso como sospechoso. Esto es intencional — en oncología, es preferible generar una alerta que resulte ser un falso positivo a dejar pasar un caso real de cáncer sin detectar. En términos de decisión, el modelo funciona como un **sistema de triage de alto recall con costo de revisión controlado**: por cada 1.4 casos revisados por los SS, se confirma 1 caso oncológico real (NNR = 1.4).

4.3 Curva ROC



Curva ROC del modelo de clasificación oncológica. El área bajo la curva (AUC) mide la capacidad discriminativa global del modelo.

La curva ROC (Receiver Operating Characteristic) representa el rendimiento del modelo en todos sus posibles umbrales de decisión. El eje vertical muestra la sensibilidad (proporción de oncológicos correctamente detectados) y el eje horizontal muestra la tasa de falsos positivos (1 – especificidad). Cuanto más se acerca la curva a la esquina superior izquierda, mejor es la capacidad del modelo para distinguir entre

casos oncológicos y no oncológicos. La diagonal gris representa un clasificador aleatorio (sin capacidad discriminativa). El área bajo la curva (AUC) resume este rendimiento en un solo número: un valor de 1.0 indica discriminación perfecta, mientras que 0.5 equivale al azar.

4.4 Análisis de métricas

La sensibilidad de 95% indica que el modelo detecta correctamente la gran mayoría de los casos oncológicos, dejando solo 448 falsos negativos. En contexto clínico oncológico, estos FN representan el riesgo más relevante: pacientes con cáncer no priorizados por el modelo.

La especificidad de 95.4% genera falsos positivos, pero en screening oncológico esto es preferible: es mejor generar una alerta innecesaria que perder un caso de cáncer. El trade-off entre sensibilidad y especificidad es deliberado.

Estas métricas se calculan **exclusivamente sobre casos ingresados desde junio 2025** (sin arrastre de casos históricos). Esta decisión responde a que los casos anteriores a esa fecha presentan problemas de calidad de datos: registros con texto clínico incompleto, codificación inconsistente y validaciones SS realizadas bajo criterios distintos a los actuales. Incluirlos introduciría ruido que no refleja el desempeño real del modelo sobre datos contemporáneos. Adicionalmente, se usa la definición **sin Antecedentes** como caso oncológico positivo, ya que el objetivo es medir el impacto sobre pacientes en proceso clínico activo (diagnóstico, tratamiento, rehabilitación), no sobre aquellos con historia oncológica previa sin proceso vigente.

4.5 Umbrales de referencia

Umbrales de referencia para métricas del modelo				
Basado en metodología documentada				
Métrica	Excelente	Bueno	Aceptable	Crítico
Sensibilidad	≥90%	80-89%	70-79%	<70%
Especificidad	≥95%	90-94%	80-89%	<80%
VPP	≥70%	60-69%	50-59%	<50%
VPN	≥99%	95-99%	90-95%	<90%
F2-Score	≥90	80-89	70-79	<70
MCC	≥70	50-69	30-49	<30

5 Evolución Temporal de Métricas

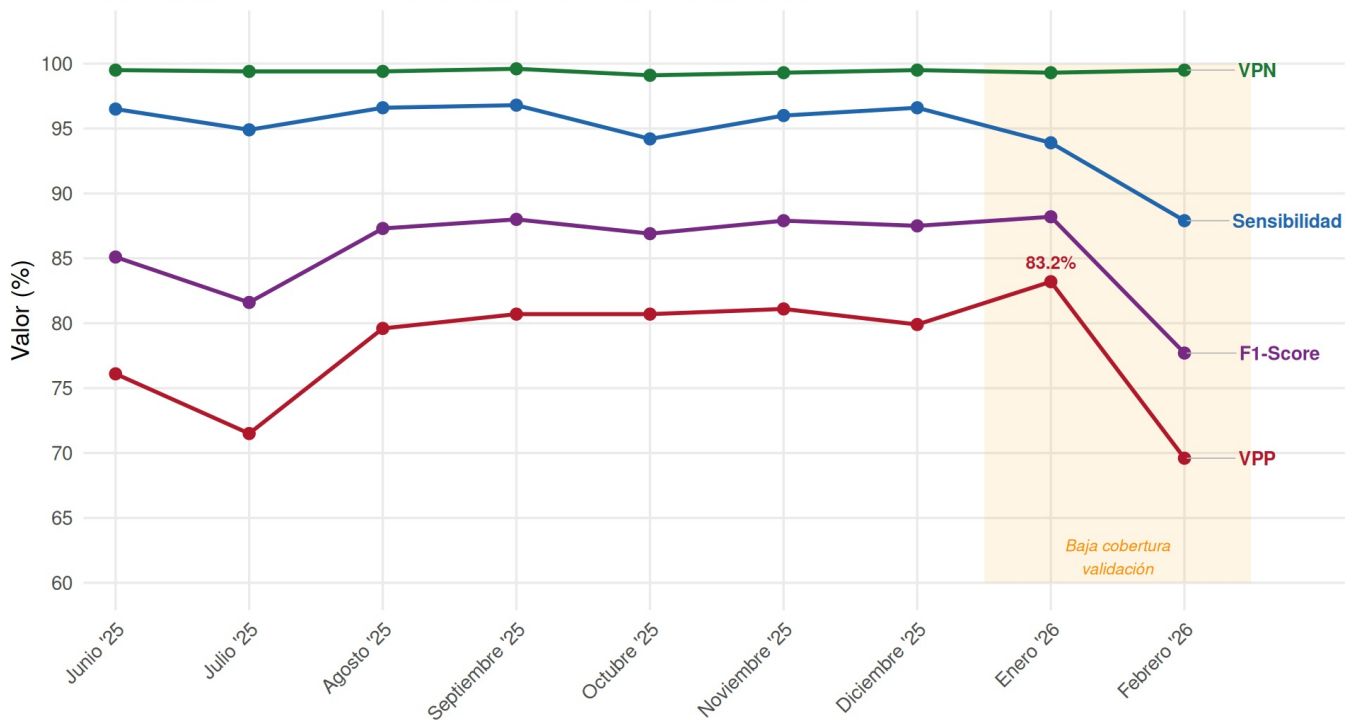
 Advertencia

Nota sobre cobertura de validación: Los meses de enero 2026 (51.8%) y febrero 2026 (17.1%) presentan baja cobertura de validación por parte de los SS. Esto significa que el ground truth es incompleto y las métricas de esos meses están sesgadas a la baja. Se muestran con línea punteada en los gráficos.

5.1 Evolución de métricas principales

Evolución temporal de métricas del modelo

Período junio 2025 — febrero 2026 (definición Diag + Tto + Rehab)



Evolución mensual de las principales métricas del modelo (junio 2025 — febrero 2026). La zona sombreada indica meses con baja cobertura de validación (<60%).

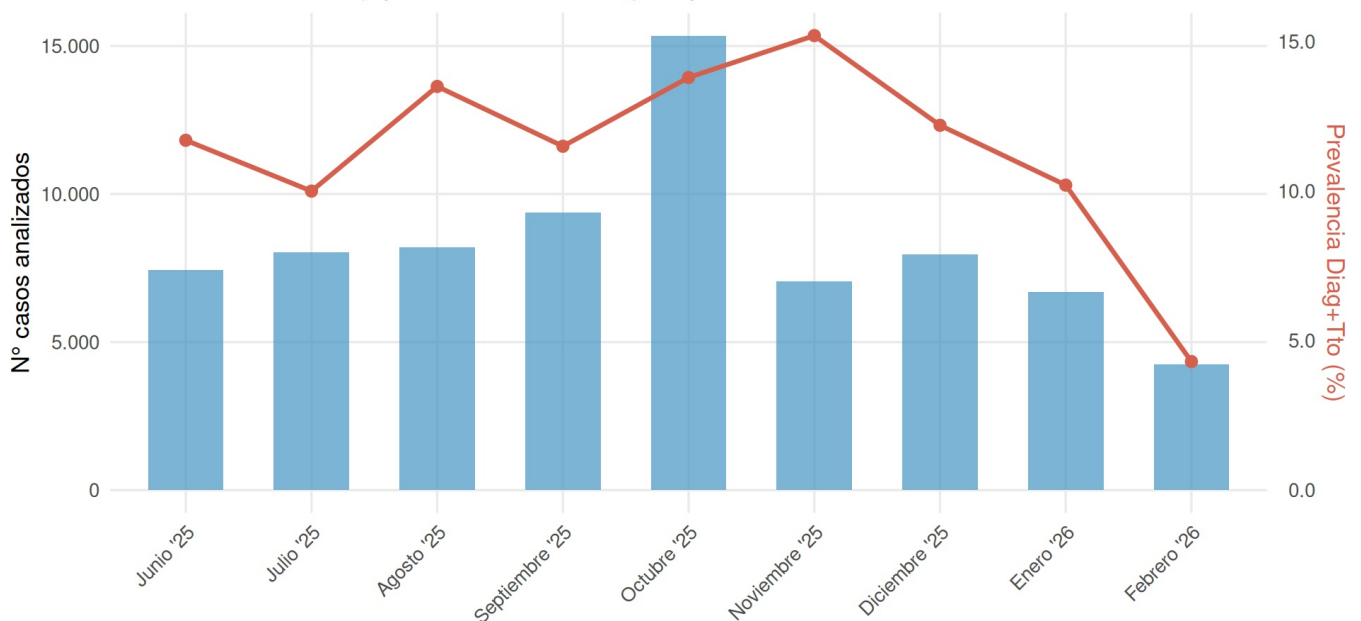
5.2 Análisis de tendencias

La sensibilidad se mantiene estable entre ~94-97% durante junio a diciembre 2025, con descenso en enero 2026 (93.9%) y febrero 2026 (87.9%). Este descenso es explicable por la baja cobertura de validación (17.1% en febrero 2026): con ground truth incompleto, las métricas se sesgan a la baja y no reflejan el desempeño real del modelo.

El hallazgo más relevante es la **tendencia creciente del VPP**: de 71.5% a 81.1% en los meses con cobertura de validación suficiente. Esto sugiere un “círculo virtuoso”: a medida que los SS validan más casos, la calidad del ground truth mejora y las métricas reflejan mejor el desempeño real.

5.3 Evolución de volumen y prevalencia

Volumen de casos y prevalencia oncológica por mes

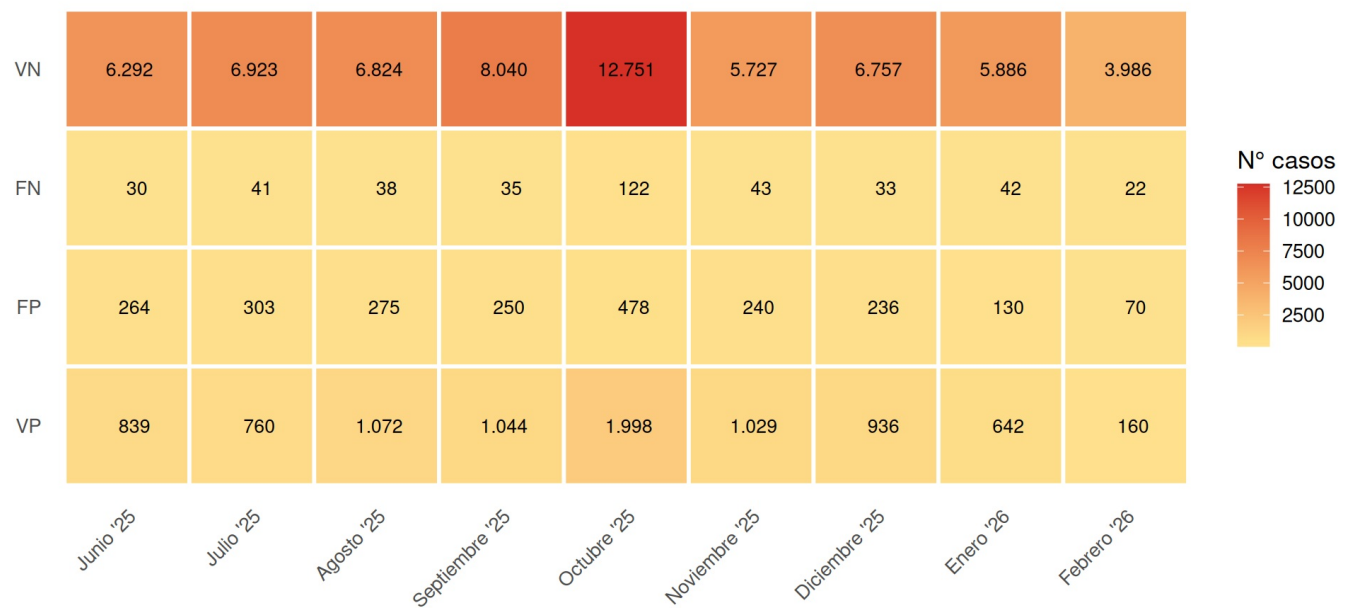


Número total de casos analizados por mes (barras) y prevalencia de casos oncológicos Diag+Tto (línea roja, eje derecho).

Octubre 2025 presenta un pico de 15.349 casos (vs ~4.200-9.400 en otros meses), posiblemente por carga acumulada o ajuste operativo. La prevalencia de casos oncológicos varía entre meses, reflejando tanto la composición real como la cobertura de validación del SS.

5.4 Evolución de la matriz de confusión

Heatmap: componentes de la matriz de confusión por mes



Evolución mensual de los componentes de la matriz de confusión (VP, VN, FP, FN). El color indica la magnitud relativa dentro de cada componente.

Los FP disminuyen en proporción en los meses con buena cobertura, indicando que el modelo genera menos falsas alarmas a medida que la calidad de datos y la cobertura de validación mejoran. Los FN se mantienen bajos y estables, reflejando una sensibilidad sostenida.

5.5 Tabla resumen mensual

Métricas mensuales del modelo						
Definición: positivo = Diagnóstico + Tratamiento + Rehabilitación						
Mes	N casos	Sens. (%)	VPP (%)	VPN (%)	F1	Cobertura (%)
Junio 2025	7.425	96.5	76.1	99.5	85.1	77.2
Julio 2025	8.027	94.9	71.5	99.4	81.6	71.5
Agosto 2025	8.209	96.6	79.6	99.4	87.3	80.2
Septiembre 2025	9.369	96.8	80.7	99.6	88.0	87.8
Octubre 2025	15.349	94.2	80.7	99.1	86.9	85.8
Noviembre 2025	7.039	96.0	81.1	99.3	87.9	85.5
Diciembre 2025	7.962	96.6	79.9	99.5	87.5	84.5
Enero 2026	6.700	93.9	83.2	99.3	88.2	51.8
Febrero 2026	4.238	87.9	69.6	99.5	77.7	17.1

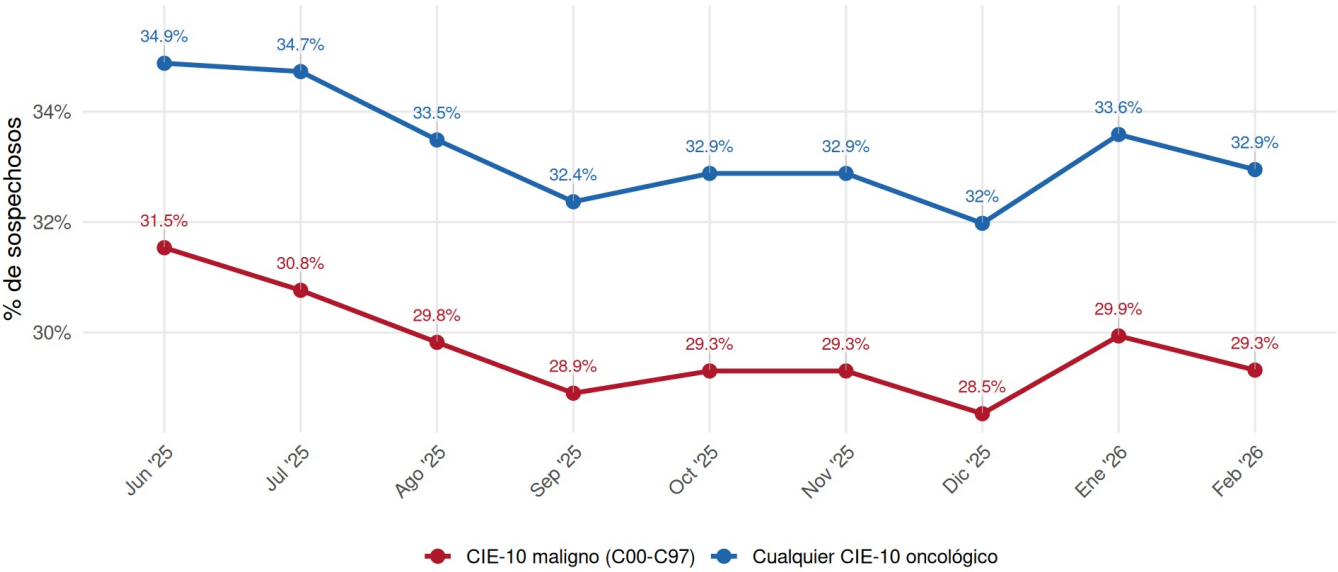
Filas destacadas: cobertura de validación < 60% (métricas menos confiables)

6 Calidad de Codificación CIE-10 en Casos Sospechosos

El Ministerio de Salud impulsa activamente el uso de codificación CIE-10 estructurada en las redes asistenciales. Esta sección evalúa la presencia de códigos CIE-10 en el texto clínico de los casos clasificados como **sospechosos** por el modelo, considerando **solo ingresos desde junio 2025** (sin arrastre histórico) para evitar que la masa de casos antiguos diluya los cambios recientes en codificación.

6.1 Proporción de sospechosos con códigos CIE-10

Codificación CIE-10 en casos sospechosos (sin arrastre)
Ingresos desde junio 2025 — proporción cuyo texto clínico contiene códigos CIE-10



Proporción de casos sospechosos (ingresados desde junio 2025) con al menos un código CIE-10 detectado en el texto clínico, por corte mensual.

6.2 Desglose por tipo de código CIE-10

Codificación CIE-10 en sospechosos por corte (sin arrastre)						
Ingresos desde junio 2025 — proporción con cada tipo de código CIE-10 en su texto clínico						
Corte	N sosp.	% Maligno	% Benigno	% No neoplásico	% Sospechoso	% Cualquier CIE-10
Jun '25	1.557	31.5	0.1	2.1	1.6	34.9
Jul '25	2.828	30.8	0.4	2.6	1.8	34.7
Ago '25	3.930	29.8	0.3	2.2	1.7	33.5
Sep '25	4.761	28.9	0.3	2.2	1.6	32.4
Oct '25	6.402	29.3	0.4	2.3	1.8	32.9
Nov '25	6.402	29.3	0.4	2.3	1.8	32.9
Dic '25	6.905	28.5	0.4	1.9	1.8	32.0
Ene '26	7.018	29.9	0.5	2.0	1.9	33.6
Feb '26	7.299	29.3	0.6	2.1	1.8	32.9

6.3 Análisis de codificación CIE-10

La codificación CIE-10 en los casos sospechosos ingresados desde junio 2025 muestra qué proporción tiene códigos oncológicos explícitos en su texto clínico. Al excluir el arrastre histórico, se aísla mejor la señal de adopción reciente de la codificación estructurada. Los códigos malignos (C00-C97) son los más relevantes porque representan diagnósticos de neoplasias confirmadas.

Los casos ingresados desde junio 2025 presentan tasas de codificación CIE-10 maligno entre 28.5% y 31.5%, notablemente superiores a las observadas cuando se incluye el arrastre histórico (~23-25%). Esto indica que los casos más recientes tienden a tener mejor codificación estructurada. Sin embargo, la tendencia dentro del período analizado es estable: no se observa una adopción creciente mes a mes, lo que sugiere que la mejora es atribuible a la composición de los casos nuevos más que a un cambio sistemático en las prácticas de codificación.

La presencia de códigos benignos o no neoplásicos en sospechosos no es necesariamente un error: un paciente puede tener un código benigno previo y una sospecha de malignidad posterior registrada en el mismo texto clínico.

7 Modelo Conceptual de Impacto

El impacto observado en este reporte no es atribuible exclusivamente al modelo algorítmico, sino a su integración en un **sistema socio-técnico** donde la detección automatizada activa procesos clínicos humanos: validación por especialistas, entrada al circuito oncológico, estadificación, comité y tratamiento.

El modelo no mejora el diagnóstico ni el tratamiento en sí mismo. Lo que optimiza es el **orden en que el sistema de salud responde**: identifica qué pacientes necesitan atención prioritaria entre cientos de miles de casos. Y esa reordenación — combinada con la acción clínica de los Servicios de Salud — se asocia a diferencias medibles en tiempos, acceso y resultados.

La evidencia presentada en las dimensiones de este reporte permite trazar el mecanismo completo:

Cadena de valor: detección → validación → desenlace		
Cada eslabón del sistema se asocia a diferencias clínicas observables		
Eslabón	Cuando funciona	Cuando falla
1. Detección algorítmica	El paciente entra a cola prioritaria. Accede a cirugía 4.8x más rápido que si no es detectado (§D6)	Queda como 'no sospechoso' (FN), invisible entre 732.837+ casos. Mediana a cirugía: 986 días vs 206 días
2. Validación por SS	El paciente ingresa al circuito oncológico (estadificación, comité, tratamiento). Mejor supervivencia ajustada por tiempo (§D4)	Queda como sospechoso sin acción clínica. 6.379 pacientes (23.2%) en esta situación. Peor supervivencia (p = 0.0027)
3. Sistema completo	Detección + validación → tratamiento oportuno e integral. Mortalidad 7.5x mayor en sospechosos confirma que el modelo identifica riesgo clínico real (§D3)	Sin detección ni validación, el paciente oncológico espera como caso general sin priorización

El modelo actúa como un **marcador de riesgo clínico temprano**: no solo clasifica texto, sino que identifica una población con desenlaces significativamente peores (mortalidad, acceso a cirugía, supervivencia). La validación por SS no es un trámite administrativo: es la puerta de entrada al circuito oncológico y se asocia a diferencias significativas en la supervivencia del paciente.

7.1 Tres efectos distintos

Este reporte mide tres efectos que operan en cascada pero son analíticamente separables:

1. **Asociación con detección** — ¿Ser clasificado como sospechoso se asocia a una trayectoria diferente? Evidencia: VP acceden a cirugía **4.8x más rápido** que FN (§D6).
2. **Asociación con validación** — ¿Ser validado por el SS se asocia a diferencias en pronóstico? Evidencia: sospechosos sin validar tienen **peor supervivencia** que VP confirmados, p = 0.0027 (§D4).
3. **Identificación de riesgo** — ¿El modelo identifica riesgo clínico real? Evidencia: mortalidad **7.5x mayor** en sospechosos vs no sospechosos (§D3).

Los datos muestran que cada eslabón se asocia a diferencias en desenlaces. Cuando alguno falta — cuando el modelo no detecta, o cuando detecta pero el SS no valida — se observan peores resultados.

ⓘ Importante

Sobre la evidencia disponible: Este es un estudio observacional sin línea base histórica (no existen datos del período previo a la implementación del modelo). La evidencia se basa en comparaciones transversales entre grupos (detectados vs no detectados, validados vs no validados). Sin embargo, la **consistencia, magnitud y plausibilidad clínica** de los efectos observados — gradiente de riesgo, asociación fuerte, temporalidad correcta y efecto dosis-respuesta entre validación y desenlaces — configuran un patrón consistente con un efecto operativo de la detección y validación sobre los desenlaces clínicos, aunque la demostración causal formal requeriría un diseño prospectivo o cuasi-experimental.

Las siguientes dimensiones presentan la evidencia que sustenta cada eslabón de esta cadena.

8 Dimensión 1: Reducción Potencial de la LE

8.1 Pregunta

¿Cuánto puede el modelo reducir la carga de revisión oncológica en la LE IQ?

8.2 Distribución de la LE por clasificación

Distribución de la LE IQ por clasificación del modelo

Universo: 9 cortes mensuales agregados (jun 2025 — feb 2026), deduplicados por caso único

Clasificación	N° casos	% del total
No sospechoso	703.982	96.1%
Sospechoso	27.515	3.8%
Indeterminado	1.340	0.2%
Total	732.837	100.0%

8.3 Análisis

La tabla muestra cómo el modelo distribuye los 732.837 casos de la LE IQ. Solo un 3.8% (27.515 casos) son clasificados como sospechosos de patología oncológica. Esto significa que el modelo permite a los Servicios de Salud focalizar su revisión clínica en un grupo pequeño y manejable, en vez de tener que revisar la totalidad de la lista sin ningún criterio de priorización.

9 Dimensión 2: Valor de Detección (Contrafactual)

9.1 Pregunta

¿Cuál es el valor operativo del modelo comparado con no tener modelo?

9.2 Escenarios

Comparación de escenarios: sin modelo vs con modelo		
Casos con ingreso desde junio 2025		
Escenario	Casos a revisar	Factor de reducción
Sin modelo	328.066 (100%)	—
Con modelo	~16.420 (5%)	~20x reducción

9.3 Falsos negativos: el costo del modelo

Los 448 FN representan oncológicos clasificados como no sospechoso: el costo inevitable de la priorización. Estos pacientes no reciben alerta prioritaria del modelo, aunque pueden ser detectados por los SS durante su revisión habitual.

El modelo permite una reducción de ~20x en la carga de revisión prioritaria. Los SS deben seguir revisando activamente los no sospechosos con tiempos de espera prolongados, ya que estos podrían incluir FN con patología avanzada.

10 Dimensión 3: Mortalidad (Kaplan-Meier)

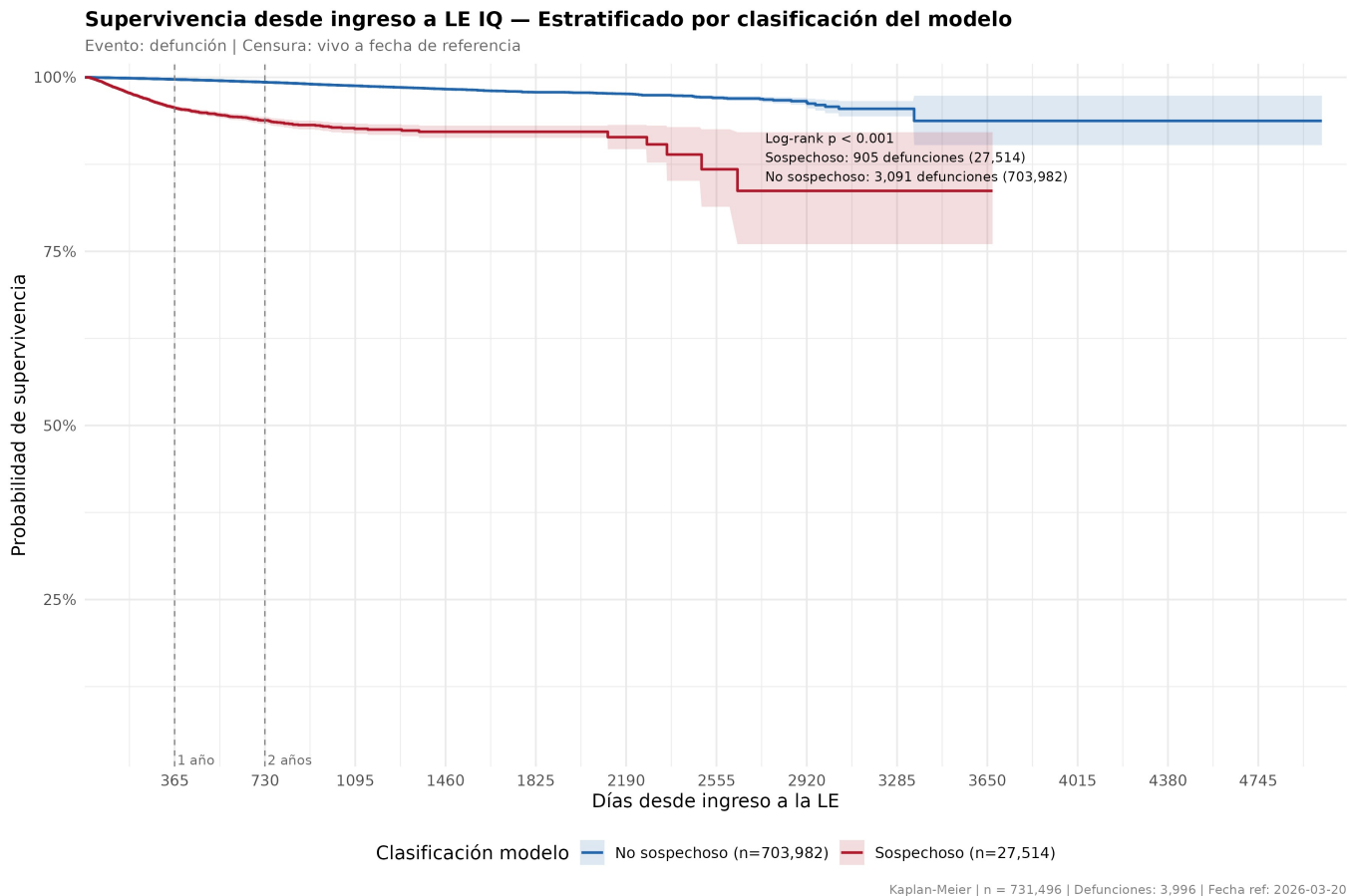
10.1 Pregunta

¿Los casos clasificados como sospechosos por el modelo tienen mayor mortalidad?

10.2 Mortalidad por grupo

Mortalidad por clasificación del modelo			
Tasa bruta de mortalidad por 1.000 casos			
Grupo	N casos	Defunciones	Tasa (×1.000)
Sospechoso	27.515	910	33.1
No sospechoso	703.982	3.101	4.4
Log-rank test: p = < 0.001. Razón de tasas: 7.5x			

10.3 Curvas de supervivencia



Curvas Kaplan-Meier de supervivencia por clasificación del modelo. La curva de sospechosos (rojo) desciende más rápido que la de no sospechosos (azul), con separación significativa desde los primeros meses.

10.4 Análisis

Los sospechosos tienen una tasa de mortalidad **7.5 veces mayor** (33.1 vs 4.4 por 1.000), con separación estadísticamente significativa (log-rank $p = < 0.001$). Las curvas se separan desde los primeros meses, indicando que la clasificación captura diferencias reales en riesgo desde el inicio.

Este resultado es la evidencia más contundente de que el modelo no solo clasifica texto: **identifica pacientes con desenlaces clínicos significativamente peores**. La mortalidad actúa como proxy de gravedad de la patología subyacente, validando que la etiqueta “sospechoso” tiene significado clínico real.

11 Dimensión 4: Relevancia de la Validación SS

11.1 Pregunta

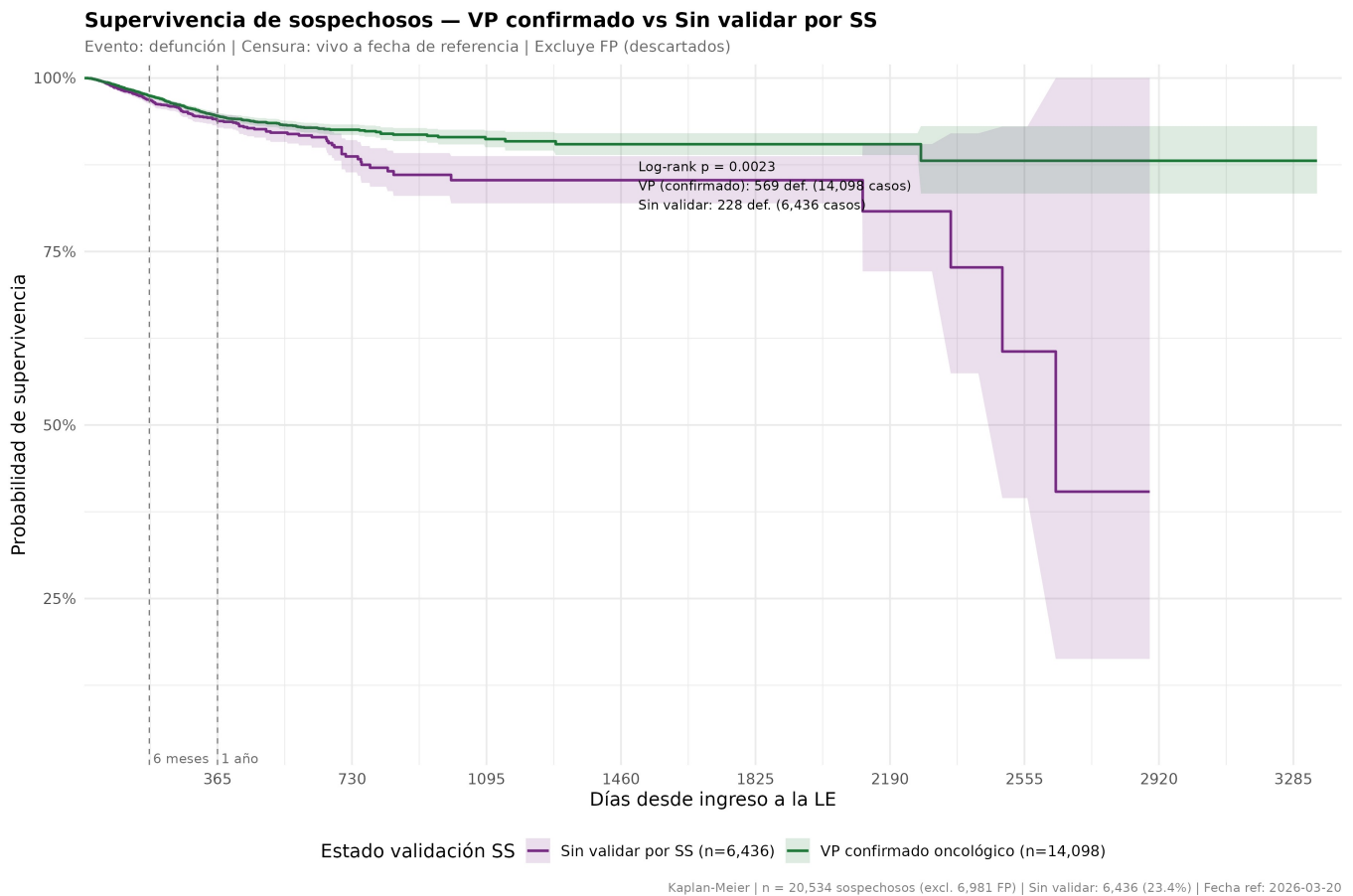
¿Importa que el Servicio de Salud valide los casos sospechosos? ¿Los sospechosos sin validar tienen peor pronóstico?

11.2 Estado de validación de sospechosos

Mortalidad de sospechosos por estado de validación			
Tasa bruta por 1.000 casos			
Estado	N casos	Fallecidos	Tasa (×1.000)
Sin validar por SS	6.379	228	35.7
VP (confirmado oncológico)	14.180	574	40.5
FP (descartado por SS)	6.956	108	15.5

11.3 Supervivencia KM: VP confirmado vs Sin validar

Supervivencia Kaplan-Meier por estado de validación		
Excluyendo FP (descartados por SS)		
Grupo	Mortalidad a 6 meses	Mortalidad a 1 año
VP (confirmado)	2.60%	5.46%
Sin validar	3.20%	6.19%
Log-rank test: p = 0.0027 n = 20.559 sospechosos		



Curvas KM de supervivencia: VP confirmado (verde) vs Sin validar (púrpura). Los sospechosos sin validar tienen peor supervivencia ajustada por tiempo.

11.4 Análisis

Hallazgo contraintuitivo: Los VP confirmados (oncológicos bajo tratamiento) tienen tasa bruta más alta (40.5 por 1.000) que sin validar (35.7 por 1.000), pero la curva KM muestra que los sin validar tienen **peor supervivencia ajustada por tiempo** (p = 0.0027). Los VP llevan más tiempo en la LE (mayor exposición al riesgo), pero al ajustar por tiempo de seguimiento, los sin validar fallecen más rápido.

Los 6.379 sospechosos sin validar (23.2% de todos los sospechosos) constituyen una cohorte de alto riesgo. Su perfil de mortalidad es similar al de los VP confirmados, lo que sugiere que una proporción importante podría ser oncológica. Sin embargo, la asociación entre falta de validación y peor pronóstico admite dos interpretaciones: la validación podría ser protectora (acceso al circuito oncológico), o la falta de validación podría ser consecuencia de una trayectoria clínica tan rápida que el ciclo de validación no alcanza a completarse. La sección siguiente explora esta pregunta con indicadores proxy.

Se excluyeron los FP de la comparación KM porque su baja mortalidad (15.5 por 1.000) diluye la señal al mezclarlos con VP, dado que no son casos oncológicos.

ⓘ Importante

Los casos clasificados por el SS como “Antecedentes” **no están incluidos** en el grupo VP. Solo Diagnóstico, Tratamiento y Rehabilitación se consideran oncológicos positivos. Los de Antecedentes caen en el grupo FP junto con “No oncológico”, y están excluidos de la comparación KM principal.

11.5 Robustez geográfica: ¿El efecto se sostiene entre Servicios de Salud?

Una pregunta legítima es si el efecto observado a nivel nacional está concentrado en pocos Servicios de Salud o es un patrón distribuido. Para evaluarlo, se ajustaron modelos de regresión de Cox (riesgos proporcionales) que controlan por la variabilidad geográfica.

Riesgo de mortalidad: Sin validar vs VP confirmado			
Hazard Ratio (HR > 1 = sin validar muere más rápido)			
Modelo	HR	IC 95%	p-valor
Cox simple (sin ajustar por SS)	1.26	1.08 — 1.47	0.0028
Cox estratificado por SS	1.24	1.05 — 1.48	0.0123
Cox con SS como covariable	1.27	1.07 — 1.51	0.0053
n = 20.558 sospechosos Eventos (defunciones): 801 29 Servicios de Salud			

El modelo **Cox estratificado por SS** es la prueba más rigurosa: permite que cada Servicio de Salud tenga su propia tasa basal de mortalidad, y evalúa si dentro de cada SS los sin validar mueren más rápido. El efecto persiste con significancia estadística.

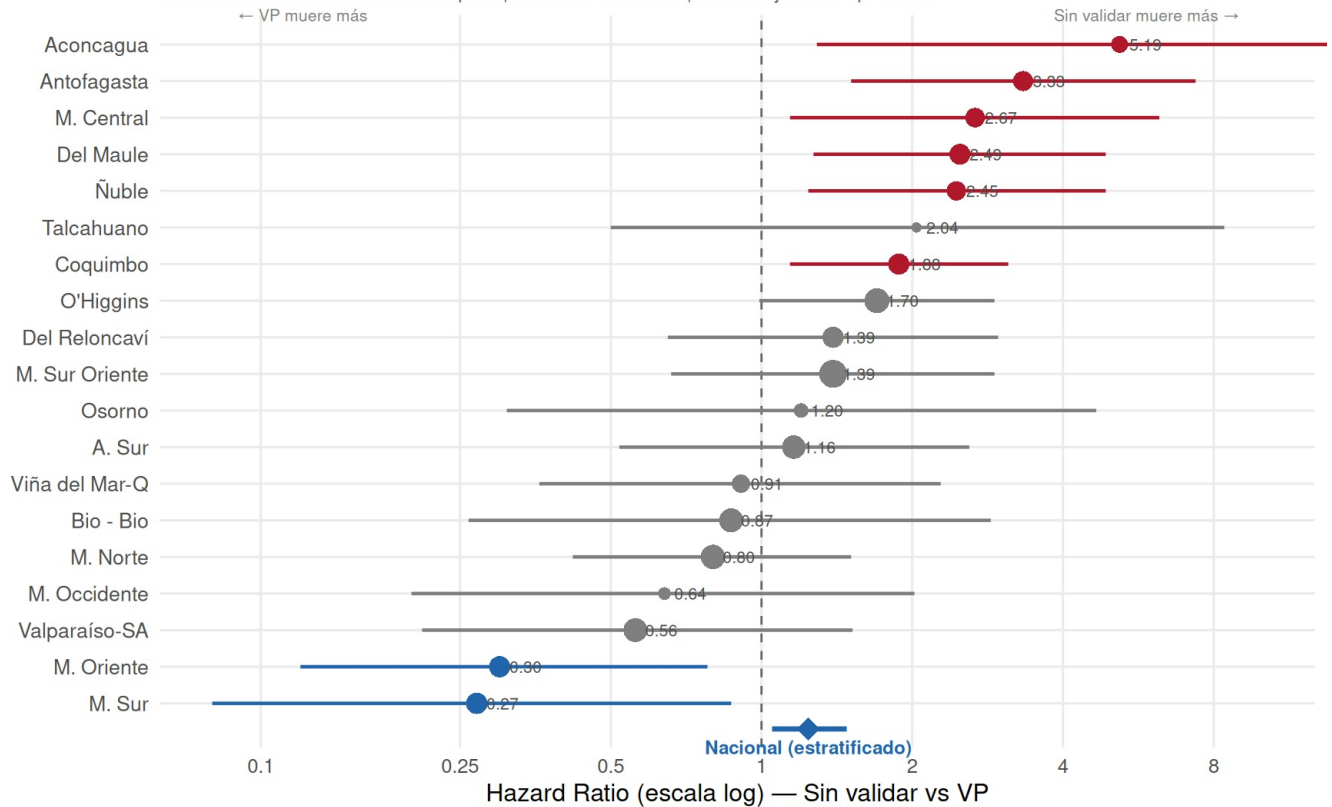
⚠ Advertencia

Nota metodológica: El Modelo 3 (Cox con SS como covariable de efecto fijo) estima 28 parámetros adicionales (uno por cada Servicio de Salud). Se excluyó el SS Arauco (61 registros) por presentar cero eventos (fallecimientos) en ambos grupos (VP y Sin validar), lo que produce separación completa y hace imposible estimar un coeficiente finito para ese estrato. Su exclusión no afecta las estimaciones del efecto principal ya que, al no tener eventos, Arauco no aporta información al modelo. El Modelo 2 (estratificado) es preferible porque controla por SS sin estimar coeficientes individuales. Los tres modelos convergen en la misma dirección y magnitud del efecto, lo que refuerza la robustez del hallazgo.

Servicio de Salud	n	Def. VP	Def. SV	HR	IC 95%	p-valor	Dirección
M. Sur Oriente	1.961	20	11	1.39	0.66 — 2.92	0.3850	SV muere más
O'Higgins	1.439	40	20	1.70	0.99 — 2.92	0.0532	SV muere más
M. Norte	1.359	36	13	0.80	0.42 — 1.51	0.4850	VP muere más
Bio - Bio	1.300	39	3	0.87	0.26 — 2.87	0.8200	VP muere más
Valparaíso-SA	1.274	91	4	0.56	0.21 — 1.52	0.2563	VP muere más
A. Sur	1.221	27	8	1.16	0.52 — 2.60	0.7233	SV muere más
M. Sur	941	41	3	0.27	0.08 — 0.87	0.0284	VP muere más
Del Reloncaví	926	9	30	1.39	0.65 — 2.97	0.3964	SV muere más
M. Oriente	913	32	5	0.30	0.12 — 0.78	0.0133	VP muere más
Coquimbo	900	25	44	1.88	1.14 — 3.11	0.0142	SV muere más
Del Maule	871	14	23	2.49	1.27 — 4.87	0.0078	SV muere más
Antofagasta	784	20	11	3.33	1.51 — 7.36	0.0030	SV muere más
M. Central	722	29	7	2.67	1.14 — 6.23	0.0230	SV muere más
Ñuble	705	19	15	2.45	1.24 — 4.87	0.0101	SV muere más
Viña del Mar-Q	612	19	6	0.91	0.36 — 2.28	0.8388	VP muere más
Aconcagua	489	7	4	5.19	1.29 — 20.96	0.0208	SV muere más
Osorno	350	7	3	1.20	0.31 — 4.66	0.7950	SV muere más
M. Occidente	281	11	4	0.64	0.20 — 2.02	0.4507	VP muere más
Talcahuano	246	7	3	2.04	0.50 — 8.40	0.3215	SV muere más

Riesgo de mortalidad por Servicio de Salud

HR > 1: sin validar muere más rápido | Tamaño = n casos | Color: rojo/azul = $p < 0.05$



Forest plot: Hazard Ratio (Sin validar vs VP) por Servicio de Salud. Línea punteada = HR 1 (sin diferencia). Diamante azul = estimación nacional estratificada.

De los 19 SS con suficientes eventos para análisis individual, **12** muestran la misma dirección que el hallazgo nacional (sin validar muere más rápido, HR > 1), y **8** alcanzan significancia estadística. Solo 7 SS muestran dirección contraria.

Nota

Interpretación: El efecto de la validación sobre la supervivencia **no es un artefacto de unos pocos Servicios de Salud**. El Cox estratificado — que controla por la variabilidad geográfica en mortalidad basal — confirma que los sospechosos sin validar tienen un riesgo de mortalidad ~24% mayor que los VP confirmados (HR = 1.24, IC 95%: 1.05 — 1.48, $p = 0.0123$), controlando por diferencias en riesgo basal entre Servicios de Salud. Si bien existe heterogeneidad en la magnitud del efecto entre Servicios de Salud, la consistencia en la dirección sugiere un fenómeno estructural más que local. Algunos Servicios de Salud (ej. M. Oriente) muestran resultados en dirección opuesta, lo que podría reflejar diferencias en procesos de validación, tiempos de registro o características clínicas. La persistencia del efecto en modelos estratificados sugiere que no se explica únicamente por confusión geográfica.

11.6 ¿La falta de validación es causa o consecuencia del peor pronóstico?

Los análisis anteriores muestran que los sospechosos sin validar tienen peor supervivencia que los VP confirmados. Existen dos interpretaciones posibles:

- **Hipótesis A (validación protectora):** La validación es la puerta de entrada al circuito oncológico; no validar priva al paciente de estadificación, comité y tratamiento oncológico estructurado.
- **Hipótesis B (causalidad reversa):** Los pacientes más graves progresan tan rápido que el ciclo de validación SS — que toma semanas o meses — no alcanza a completarse. La falta de validación no es causa del peor pronóstico sino marcador de una trayectoria clínica acelerada.

Para evaluar estas hipótesis, se analizan cinco indicadores proxy. No se dispone de un campo `fecha_validacion` en la BD, por lo que los análisis son indirectos.

11.6.1 Perfil de desenlace de los sospechosos sin validar

Desenlace de los sospechosos sin validar

n = 6.379 casos

Perfil de desenlace	N casos	% del total
No operado + Fallecido	84	1.3%

No operado + Vivo (pendiente)	3.000	47.0%
Operado + Fallecido	144	2.3%
Operado + Vivo	3.151	49.4%

11.6.2 Velocidad de resolución: VP vs Sin validar

Entre los casos que ya egresaron, la mediana de días entre ingreso y egreso indica qué tan rápido transitan por el sistema.

Días entre ingreso y egreso (casos egresados)				
VP confirmado vs Sin validar				
Grupo	N egresados	Mediana (días)	P25	P75
Sin validar	4.024	36	17	79
VP	9.402	50	17	116
Wilcoxon p = < 0.001				

11.6.3 Tiempo a muerte: ¿Los sin validar fallecen antes?

Entre los fallecidos de cada grupo, se compara la mediana de días entre ingreso a LE y defunción. Una mediana menor en sin validar sugiere una enfermedad de progresión más rápida.

Días entre ingreso a LE y defunción		
Solo fallecidos: VP vs Sin validar		
Grupo	N fallecidos	Mediana (días)
VP (confirmado)	574	162
Sin validar	228	160
Wilcoxon p = 0.6606		

11.6.4 Egreso precoz sin validación

¿Qué proporción de sin validar egresaron en menos de 90 días? Un porcentaje alto indicaría que el sistema no tuvo tiempo de validarlos.

De los 4.024 sin validar que egresaron, **3.116 (77.4%)** lo hicieron en menos de 90 días desde su ingreso a LE.

11.6.5 Landmark analysis: ¿Persiste la diferencia después de 6 meses?

Si la diferencia de supervivencia entre VP y Sin validar se debe a muertes tempranas (trayectoria acelerada), al excluir los primeros 180 días la diferencia debería atenuarse o desaparecer.

Landmark analysis: HR Sin validar vs VP				
¿La diferencia persiste al excluir muertes tempranas (<180 días)?				
Análisis	n	HR	IC 95%	p-valor
Cohorte completa	20.558	1.26	1.08 — 1.47	0.0028
Landmark >180 días	12.358	1.31	1.04 — 1.66	0.0232
HR > 1: sin validar muere más rápido que VP				

11.6.6 Síntesis de los indicadores proxy

Resumen de indicadores proxy	
A = validación protectora B = causalidad reversa (trayectoria acelerada)	
Resultado cuantitativo	

Indicador	Dirección	
Velocidad de resolución (egresados)	Mediana 36 vs 50 días (Wilcoxon $p = < 0.001$)	Apoya hipótesis B
Tiempo a muerte (fallecidos)	Mediana 160 vs 162 días (Wilcoxon $p = 0.6606$)	No concluyente ($p \geq 0.05$)
Egreso precoz <90 días	77.4% de sin validar egresaron en <90 días	Apoya hipótesis B
Landmark >180 días	HR = 1.31 (IC 1.04 — 1.66, $p = 0.0232$)	Apoya hipótesis A (efecto persiste)

Los datos muestran un cuadro mixto. Por un lado, los sin validar egresan significativamente más rápido que los VP (mediana 36 vs 50 días, $p < 0.001$) y el 77.4% egresó en menos de 90 días, lo que indica que una proporción importante transita por el sistema antes de que el ciclo de validación pueda alcanzarlos. Esto es consistente con la hipótesis de causalidad reversa: la falta de validación sería consecuencia de la velocidad de la enfermedad, no causa del peor pronóstico. Sin embargo, el landmark analysis muestra que la diferencia de supervivencia **persiste** al excluir los primeros 6 meses (HR = 1.31, IC 95%: 1.04 — 1.66, $p = 0.0232$), lo que sugiere que la falta de validación no se explica únicamente por muertes tempranas. El tiempo a muerte entre fallecidos no difiere significativamente entre grupos ($p = 0.6606$), lo que no apoya ni descarta ninguna hipótesis.

En conjunto, la evidencia sugiere que **ambos mecanismos operan simultáneamente**: una fracción de los sin validar corresponde a pacientes cuya trayectoria clínica fue más rápida que el ciclo administrativo (causalidad reversa), pero la persistencia del efecto a largo plazo indica que la validación también se asocia a un beneficio que va más allá de los primeros meses — probablemente el acceso al circuito oncológico estructurado (estadificación, comité, tratamiento planificado). La implicación práctica es la misma bajo ambas lecturas: el sistema debe cerrar más rápido la distancia entre la detección algorítmica y la acción clínica.

12 Dimensión 5: Tiempo a Resolución

12.1 Pregunta

¿Los sospechosos se resuelven (egresan de la LE) más rápido o más lento que los no sospechosos?

Nota metodológica: El análisis Kaplan-Meier incluye **todos los casos** (egresados y pendientes). Los casos aún pendientes se incluyen como observaciones censuradas a la fecha de referencia, evitando el sesgo de selección que ocurriría al analizar solo los casos ya resueltos.

12.2 Análisis

Las medianas KM de resolución muestran una diferencia significativa: los sospechosos tienen una mediana de 157 días vs 720 días para los no sospechosos (log-rank $p < 0.001$).

Los sospechosos se resuelven significativamente más rápido, probablemente porque reciben atención prioritaria y protocolos oncológicos más definidos que aceleran el proceso diagnóstico-terapéutico. Los no sospechosos tienen mediana muy alta (720 días), reflejando que muchos casos no oncológicos permanecen largo tiempo en LE sin resolución urgente.

13 Dimensión 6: Tiempo a Cirugía

13.1 Pregunta

¿El modelo y la validación SS aceleran el acceso a cirugía oncológica?

Nota metodológica: El análisis Kaplan-Meier incluye **todos los casos** (operados, pendientes y egresados por otra causa). Los casos sin cirugía se incluyen como observaciones censuradas, evitando el sesgo de selección que ocurriría al analizar solo los pacientes que ya fueron operados.

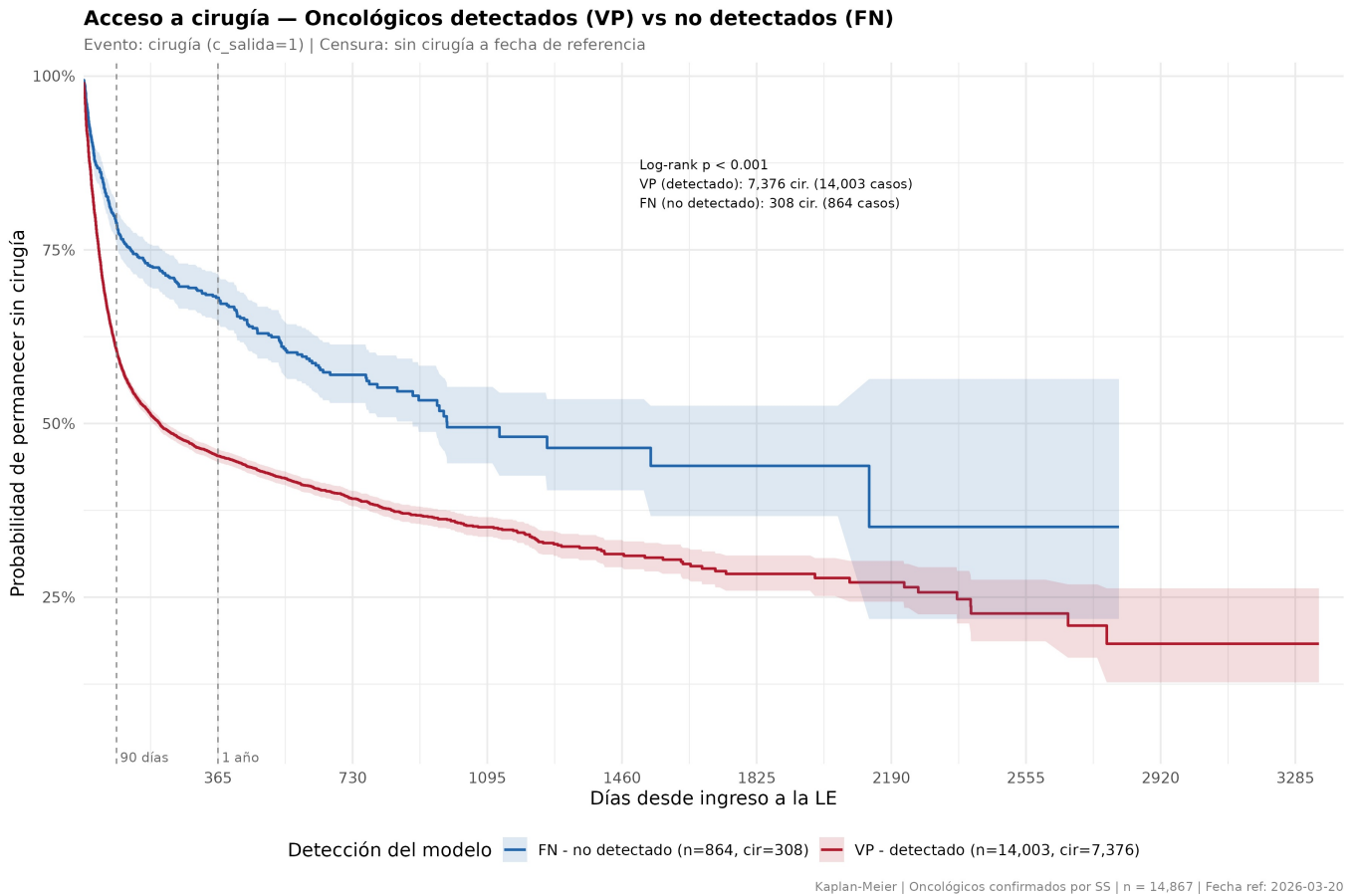
13.2 Tasa de acceso quirúrgico

Acceso a cirugía oncológica por grupo			
Proporción de casos que acceden a intervención quirúrgica			
Grupo	N total	Cirugías realizadas	% acceso quirúrgico
VP (confirmado oncológico)	14.180	7.518	53.0%
Sin validar	6.379	3.295	51.7%
FN (no detectado)	879	319	36.3%

FP (descartado)	6.956	2.423	34.8%
Las medianas KM de la tabla siguiente incluyen todos los casos (operados + censurados)			

13.3 KM: VP detectado vs FN no detectado

Tiempo a cirugía: VP vs FN (oncológicos confirmados)			
Medianas Kaplan-Meier			
Grupo	Mediana KM	IC 95%	N° cirugías
VP (detectado por modelo)	206 días	190 - 224	7.518
FN (no detectado)	986 días	893 - NA	319
Log-rank test: p < 0.001 Razón: VP acceden 4.8x más rápido			



Curvas KM de acceso a cirugía: VP detectados por el modelo (rojo) vs FN no detectados (azul). Los oncológicos detectados acceden a cirugía significativamente más rápido.

13.4 Análisis VP vs FN

Los VP acceden a cirugía con mediana KM de 206 días vs 986 días para los FN: **4.8 veces más rápido** (p < 0.001). Además, la tasa de acceso quirúrgico es 17 puntos porcentuales mayor (53% vs 36.3%).

Esta es **evidencia consistente** de una asociación significativa entre la detección por el modelo y el acceso a cirugía oncológica. Ser clasificado como sospechoso y confirmado por el SS se asocia a un acceso significativamente más rápido al tratamiento quirúrgico y a una mayor probabilidad de recibirlo.

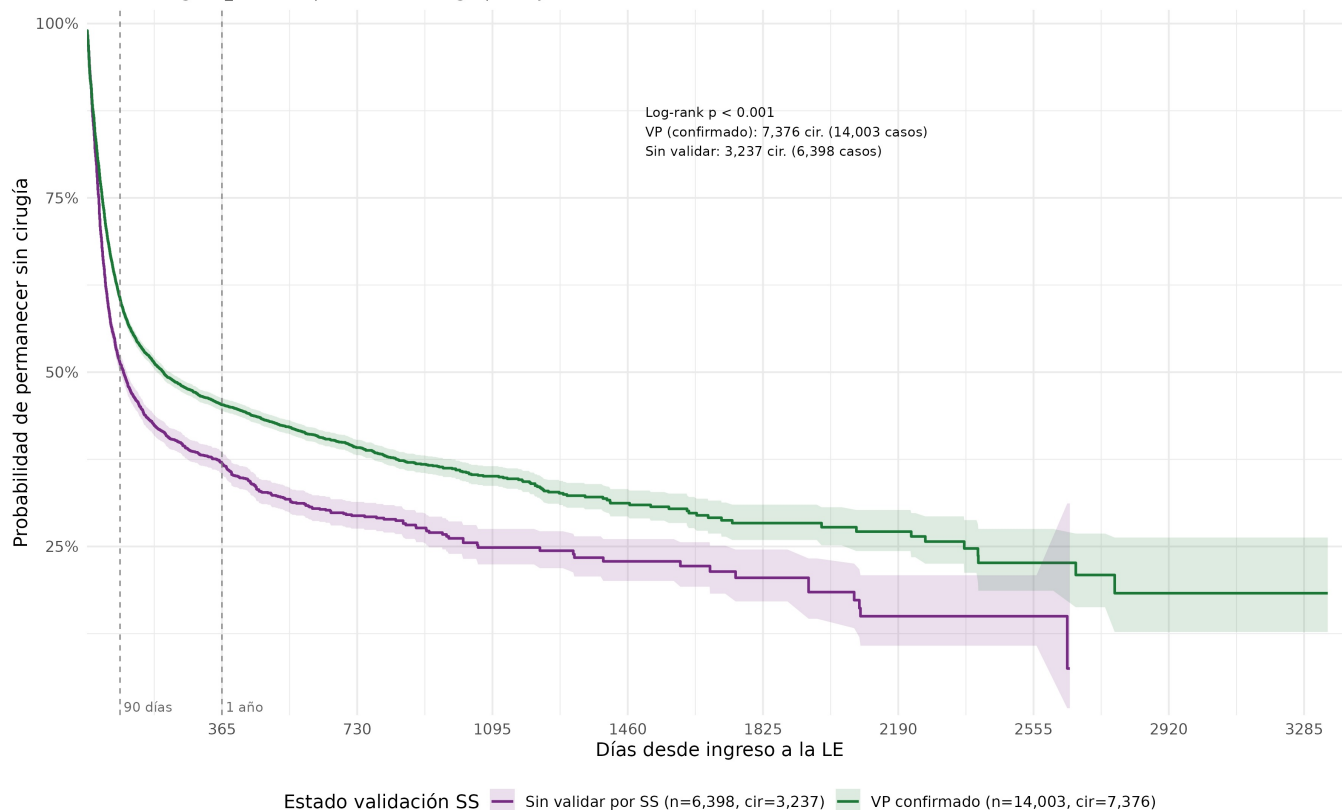
13.5 KM: VP confirmado vs Sin validar

Tiempo a cirugía: VP vs Sin validar (sospechosos)			
Medianas Kaplan-Meier			
Grupo	Mediana KM	IC 95%	N° cirugías

Sin validar	97 días	87 - 106	3.295
VP (confirmado)	206 días	190 - 224	7.518
Log-rank test: $p < 0.001$			

Acceso a cirugía — VP confirmado vs Sin validar por SS

Evento: cirugía (c_salida=1) | Censura: sin cirugía | Excluye FP (descartados)



Kaplan-Meier | Sospechosos (excl. FP) | n = 20,401 | Fecha ref: 2026-03-20

Curvas KM de acceso a cirugía: VP confirmado (verde) vs Sin validar (púrpura). Contraintuitivamente, los sin validar acceden a cirugía más rápido pero tienen peor supervivencia (ver §D4).

13.6 Análisis VP vs Sin validar

Resultado llamativo: Los sospechosos sin validar acceden a cirugía **MÁS RÁPIDO** (mediana 97 días) que los VP confirmados (206 días), pero — como se observó en la Dimensión 5 — tienen peor supervivencia. Este patrón admite múltiples explicaciones posibles:

- **Mayor urgencia clínica:** Los sin validar podrían incluir casos más graves o urgentes que se operan rápidamente, pero cuya complejidad también se asocia a peor pronóstico.
- **Demoras administrativas y madurez operativa:** La validación por SS y el circuito oncológico completo agregan tiempo (diagnóstico → estadificación → comité oncológico → cirugía), pero algunos SS aún están en proceso de consolidar estos flujos.
- **Complejidad diagnóstica:** Casos donde es difícil determinar la etapa oncológica o confirmar el diagnóstico pueden quedar sin validar y al mismo tiempo representar patologías más complejas.
- **Cirugía sin protocolo oncológico:** Los sin validar que llegan a cirugía podrían hacerlo por el circuito quirúrgico general, sin la estadificación y planificación que ofrece el protocolo oncológico.

La realidad probablemente es una combinación de estos factores. **No se cuenta con los datos necesarios para establecer causalidad** (estadio tumoral, tipo de cáncer, comorbilidades, motivo de la no validación), pero el patrón es consistente: **la velocidad de intervención sin integración al circuito oncológico no se traduce en mejor pronóstico**, sugiriendo que la calidad y coordinación del proceso clínico son determinantes críticos del resultado.

13.7 Mortalidad precoz post-cirugía: ¿La velocidad refleja urgencia?

Para profundizar en este hallazgo contraintuitivo, se analiza la mortalidad en ventanas temporales post-quirúrgicas. Si los sin validar que se operan rápido lo hacen por urgencia clínica (y no por eficiencia del sistema), esperaríamos mayor mortalidad precoz post-cirugía.

Mortalidad post-cirugía por ventana temporal

Sospechosos operados (c_salida = 1): Sin validar vs VP				
Ventana post-cx	Sin validar (n=3.295)		VP (n=7.518)	
	Defunciones	Tasa (×1.000)	Defunciones	Tasa (×1.000)
≤30 días	19	5.8	35	4.7
≤60 días	42	12.7	75	10.0
≤90 días	61	18.5	103	13.7
Total	144	43.7	290	38.6

Los sin validar muestran mayor mortalidad precoz en todas las ventanas temporales, con un **gradiente que se acentúa en los primeros 90 días** (tasa 18.5 vs 13.7 por 1.000). Esto es consistente con la hipótesis de que la cirugía rápida en sin validar refleja urgencia clínica, no eficiencia del sistema.

Días a cirugía según desenlace			
¿Los que mueren se operaron más rápido?			
Grupo	Desenlace	n	Mediana días a Cx
VP (confirmado)	Fallecido	290	23
VP (confirmado)	Vivo	7.228	46
Sin validar	Fallecido	144	25
Sin validar	Vivo	3.151	33

En **ambos grupos**, los pacientes que fallecen tuvieron cirugía más rápida que los que sobreviven. Esto confirma que la velocidad de acceso a cirugía no es necesariamente indicador de eficiencia, sino de **gravedad clínica**: los casos más urgentes llegan antes a pabellón, pero su pronóstico es peor precisamente por la severidad de su condición.

Nota

Hallazgo clave: Los sin validar tienen un 47.3% de cirugías en menos de 30 días (vs 38.1% en VP), mayor mortalidad precoz post-quirúrgica, y los que fallecen se operaron más rápido que los sobrevivientes. Estos tres elementos configuran un perfil de **urgencia/gravedad extrema**, no de acceso eficiente.

Sobre la antigüedad en LE: La mediana de antigüedad en lista de espera de los operados es de 236 días para sin validar vs 240 días para VP (Wilcoxon p = < 0.001). Existe diferencia estadísticamente significativa en antigüedad. Notablemente, la dirección del efecto no apoya la hipótesis de que los sin validar se operan más rápido por llevar menos tiempo en el sistema (registro más tardío): los VP tienen mayor antigüedad en LE, lo que es esperable dado que el proceso de validación requiere tiempo adicional. Esto no descarta otros sesgos temporales, pero indica que la diferencia en tiempos a cirugía refleja fundamentalmente urgencia clínica.

14 Síntesis y Conclusiones

14.1 Sobre la eficacia del modelo

Hallazgos sobre eficacia del modelo	
Hallazgo	Evidencia cuantitativa
El modelo actúa como marcador de riesgo clínico temprano	Mortalidad 7.5x mayor en sospechosos (33.1 vs 4.4 por 1.000; p < 0.001). No solo clasifica texto: identifica riesgo clínico real.
La detección se asocia a mayor acceso a tratamiento	VP acceden a cirugía 4.8x más rápido que FN (206 vs 986 días; p < 0.001), +17pp en acceso quirúrgico. Diferencia asociada a la detección, no solo a la performance del modelo.
Sistema de triage de alto recall con costo controlado	Concentra revisión en 3.8% de la LE (~27x reducción). NNR de ~1.4: por cada 1.4 casos revisados, se confirma 1 oncológico.
Mejora sostenida en el tiempo	VPP crece de 71.5% → 81.1% en 6 meses: círculo virtuoso entre validación SS y desempeño.

El modelo no es una herramienta de clasificación textual: es un **marcador de riesgo clínico temprano** que identifica pacientes con desenlaces significativamente peores y cuya detección se asocia a diferencias medibles en el acceso a tratamiento. La mejora sostenida del VPP evidencia un círculo virtuoso donde la validación por SS retroalimenta la calidad del sistema.

En términos más amplios, este reporte documenta cómo un sistema de priorización algorítmica se asocia a diferencias en trayectorias clínicas. El modelo no mejora el diagnóstico ni el tratamiento en sí mismo — **optimiza el orden en que el sistema de salud responde:** identifica qué pacientes necesitan atención prioritaria. Y esa reordenación, combinada con la acción clínica de los Servicios de Salud, se asocia a diferencias medibles en acceso a cirugía, tiempos de resolución y supervivencia.

La evidencia revela además una conclusión estructural: **el cuello de botella del sistema no es detectar el cáncer, sino procesarlo a tiempo.** El 23.2% de sospechosos sin validar presenta peor supervivencia y accede a cirugía más rápido pero con peor resultado. El análisis de indicadores proxy ([\\$D4b](#)) sugiere que esta brecha podría reflejar tanto una falla de proceso (el sistema no alcanza a validar) como la gravedad intrínseca de los casos (la enfermedad progresa más rápido que el ciclo administrativo). En ambas lecturas, la conclusión operativa es la misma: el sistema debe cerrar más rápido la distancia entre detección y acción clínica.

14.2 Sobre la validación por Servicios de Salud

Hallazgos sobre validación SS	
Hallazgo	Evidencia cuantitativa
La validación se asocia a mejor pronóstico	Sospechosos sin validar tienen peor supervivencia que VP (log-rank p = 0.0027; HR estratificado = 1.24, IC 95%: 1.05 — 1.48). Efecto consistente en 12 de 19 SS analizados.
Brecha crítica de cobertura	23.2% de sospechosos (6.379 pacientes) no han sido validados. Mediana de tiempo a muerte: 160 días (SV) vs 162 días (VP).
Interpretación dual: protección vs causalidad reversa	Los sin validar egresan en 36 días (mediana) vs 50 días los VP. El 77.4% de sin validar egresó en <90 días. La falta de validación podría ser consecuencia de la velocidad de la enfermedad, no solo causa del peor pronóstico (ver [\$D4b](#d4b)).
La cirugía rápida es proxy de gravedad, no de eficiencia	En ambos grupos, los pacientes que fallecen se operaron más rápido que los sobrevivientes. La mortalidad precoz post-quirúrgica (≤90 días) es mayor en sin validar.

La validación por Servicios de Salud se asocia a diferencias clínicas significativas: los sospechosos no validados tienen peor supervivencia (HR = 1.24, p = 0.0123). Sin embargo, como se exploró en la [sección D4b](#), esta asociación admite interpretaciones complementarias. La velocidad de intervención sin integración al circuito oncológico no se traduce en mejor pronóstico, pero esto podría reflejar tanto la ausencia de protocolo oncológico como la mayor gravedad intrínseca de los casos que no alcanzan a ser validados.

14.3 Discusión: la velocidad como signo de gravedad

📌 Importante

La cirugía rápida no es cirugía temprana: es cirugía tardía en la historia de la enfermedad.

Los hallazgos de la [\\$D6](#) sobre mortalidad precoz post-quirúrgica no son un dato secundario: constituyen la pieza que completa el patrón observacional detrás de los resultados globales del reporte. La evidencia converge en tres niveles:

- Evidencia de impacto:** El modelo identifica riesgo clínico real. Los sospechosos tienen 7.5x más mortalidad que los no sospechosos ([\\$D3](#)), y los VP confirmados tienen mejor supervivencia que los sospechosos sin validar ([\\$D4](#)).
- Brecha de proceso:** El 23.2% de sospechosos sin validar configura una zona ciega del sistema. Como se exploró en [\\$D4b](#), parte de esta brecha podría reflejar pacientes cuya trayectoria clínica es más rápida que el ciclo de validación. En ambos casos — sea por falla administrativa o por velocidad de la enfermedad — estos pacientes no acceden al circuito oncológico estructurado.
- La consecuencia:** Los pacientes no validados terminan en un flujo de **urgencia reactiva** que, aunque es rápido en días, es ineficaz en términos de supervivencia precoz. No es que la cirugía rápida sea mala — es que la cirugía rápida en este grupo refleja pacientes que ya llegan en peor estado clínico, operados por urgencia y sin el protocolo oncológico que mejora el pronóstico.

Lo que lo confirma es el análisis de mortalidad precoz post-cirugía por ventanas temporales (≤30, ≤60, ≤90 días) y la comparación de días a cirugía por desenlace: los pacientes que fallecen son precisamente los que se operaron más rápido, y los sin validar muestran mayor mortalidad precoz en todas las ventanas. El sistema, cuando actúa de forma coordinada, tiende a nivelar. El problema no es cómo se opera, sino **quién logra llegar a operarse y en qué condiciones**. Este patrón es consistente con confusión por indicación, donde la variable “tiempo a cirugía” captura severidad, no eficiencia.

El modelo no cambia la biología del paciente ni el resultado de la cirugía. Lo que los datos sugieren es que se asocia a diferencias en **la oportunidad y el contexto en que ese paciente entra al sistema**. Cuando eso funciona — detección, validación, protocolo oncológico — los tiempos son más largos pero los resultados son mejores. Cuando falla, el sistema igual reacciona (opera rápido), pero ya en un escenario más adverso.

14.4 Limitaciones del análisis

Las siguientes limitaciones son inherentes al diseño observacional de este estudio. Para cada una se presenta la contra-evidencia disponible que permite evaluar en qué medida podría explicar los resultados observados.

1. **Sesgo de información (immortal time bias):** Los VP han sido identificados y validados, lo que implica un proceso diagnóstico previo. Los sin validar no tienen este filtro, por lo que las comparaciones KM podrían estar afectadas por diferencias en el momento de entrada al estudio. Contra-evidencia: Las curvas KM se separan desde los primeros meses de seguimiento, no solo a largo plazo. Si el efecto fuera puramente de tiempo inmortal, esperaríamos separación tardía.
2. **Lead-time bias (sesgo de adelanto diagnóstico):** La detección más temprana por el modelo podría inflar artificialmente los tiempos de supervivencia observados en los VP, al adelantar el punto de inicio del seguimiento sin necesariamente cambiar el desenlace final. Contra-evidencia: Si fuera solo adelanto diagnóstico, esperaríamos tiempos similares de acceso a cirugía entre VP y FN; sin embargo, los VP acceden 4.8x más rápido ([\\$D6](#)), sugiriendo una diferencia real en la trayectoria de atención, no solo un desplazamiento temporal del diagnóstico.
3. **Confounders no controlados:** La mortalidad de los sin validar podría reflejar otros factores (acceso geográfico, comorbilidades, nivel socioeconómico) no controlados en este análisis observacional. Contra-evidencia: El gradiente de impacto es consistente a través de múltiples dimensiones independientes — mortalidad ([\\$D3](#)), cirugía ([\\$D6](#)), resolución ([\\$D5](#)) — lo que hace improbable que un confounder único explique todos los efectos simultáneamente.
4. **Sesgo temporal:** Casos más antiguos tienen mayor tiempo de seguimiento y mayor probabilidad de evento. La composición temporal de cada grupo podría influir en las estimaciones KM.
5. **Cobertura de validación variable:** Del universo sin arrastre (328.066 casos), 78.122 han sido validados por SS (23.8%). Las métricas se calculan sobre este subconjunto, que podría no ser representativo del universo total.
6. **Baja cobertura reciente:** Enero 2026 (51.8%) y febrero 2026 (17.1%) tienen cobertura de validación muy baja, haciendo que las métricas de esos meses sean poco confiables y sesgadas a la baja.
7. **Definición de oncológico:** Se excluyen los Antecedentes de la definición de positivo. Pacientes con historia oncológica que podrían requerir seguimiento activo no se cuentan como VP en este análisis.
8. **Causalidad vs correlación:** Este es un estudio observacional retrospectivo sin línea base histórica. Si bien el diseño no permite demostrar causalidad formalmente, la consistencia, magnitud, plausibilidad clínica y gradiente dosis-respuesta de los efectos observados configuran un patrón compatible con un efecto operativo del sistema de detección y validación, aunque la demostración causal formal excede el alcance de este diseño.
9. **Sin ajuste por gravedad:** No se controla por estadio tumoral, tipo de cáncer o comorbilidades. Las diferencias observadas podrían reflejar parcialmente diferencias en case-mix entre grupos.
10. **Causalidad reversa en validación:** Como se exploró en [\\$D4b](#), la asociación entre falta de validación y peor pronóstico podría reflejar parcialmente que los pacientes más graves progresan antes de que el ciclo de validación los alcance. No se dispone de un campo `fecha_validacion` en la BD, por lo que los análisis son indirectos (proxies).

Estos sesgos, individualmente o en conjunto, no explican la totalidad de los efectos observados, pero tampoco pueden descartarse. La magnitud, consistencia y coherencia de los resultados a través de múltiples dimensiones independientes sugieren un efecto real del sistema de detección y validación, aunque la demostración causal formal excede el alcance de este diseño observacional.

14.5 Recomendaciones

La evidencia presentada en este reporte indica que el principal desafío del sistema no radica únicamente en la detección de pacientes con riesgo oncológico — donde el modelo muestra sensibilidad de 95% y especificidad de 95.4% — sino en la capacidad de traducir esa detección en una respuesta clínica oportuna y consistente. Con 6.379 sospechosos aún sin validar (23.2%), una mortalidad 7.5 veces mayor en sospechosos, y evidencia de que los casos más graves progresan antes de que el sistema los alcance ([\\$D4b](#)), las recomendaciones se orientan a cerrar la brecha entre detección algorítmica y acción clínica.

A nivel de sistema:

- **Acelerar el ciclo de validación de sospechosos:** La evidencia muestra que los sospechosos sin validar presentan peor supervivencia ($HR = 1.24$, $p = 0.0123$). Sea porque la validación facilita el acceso al circuito oncológico o porque los casos más graves progresan antes de ser validados, la implicación es la misma: el ciclo de validación debe ser más rápido que la enfermedad. Se recomienda establecer plazos definidos y monitoreo activo.
- **Intervenir en los 6.379 sospechosos sin validar:** Esta cohorte de alto riesgo incluye tanto pacientes que el sistema no alcanzó a procesar como potenciales casos de progresión rápida. Priorizar aquellos con señales de trayectoria acelerada (ingreso reciente + cirugía programada o ya realizada).
- **Focalizar en SS con baja sensibilidad:** El SS Chiloé presenta sensibilidad $<70\%$, lo que indica una proporción elevada de falsos

negativos. Requiere acompañamiento operativo para fortalecer sus procesos de validación.

- **Incorporar priorización basada en riesgo clínico en la gestión de listas de espera:** La evidencia muestra que la velocidad de atención no equivale a eficiencia — los pacientes operados más rápido presentan mayor mortalidad precoz ([SD6](#)), y los sospechosos sin validar egresan en mediana 36 días vs 50 los VP. Esto sugiere que la capacidad quirúrgica debería asignarse no solo por antigüedad en lista, sino incorporando señales de riesgo oncológico del modelo para maximizar el beneficio clínico de cada intervención.

Estas recomendaciones tensionan los indicadores actuales de oportunidad quirúrgica: la evidencia de este reporte muestra que la rapidez de acceso no siempre refleja desempeño del sistema, sino gravedad clínica (ver [SD6](#)).

A nivel de infraestructura clínica de datos:

- **Impulsar la codificación estructurada como habilitador de decisiones clínicas:** La codificación CIE-10 no es solo un estándar de registro — es infraestructura para decidir mejor y más rápido. El modelo ya demostró que mejora con mejor data (VPP 71.5% → 81.1%) y que esa mejora se asocia a diferencias observables en trayectorias clínicas. Promover la codificación desde los establecimientos es invertir directamente en la capacidad del sistema para detectar y actuar a tiempo.
- **Avanzar hacia Fase 2 (clasificación en origen):** La clasificación oncológica desde el registro clínico en establecimientos permitirá detección más temprana y reduciría la dependencia del screening centralizado.

A nivel de evidencia:

- **Diseñar estudio prospectivo o cuasi-experimental:** Para cuantificar formalmente el impacto causal del sistema sobre desenlaces clínicos. Se propone una agenda de investigación con cuatro líneas complementarias:
 - **Variabilidad entre SS como cuasi-experimento natural:** La heterogeneidad en adopción y cobertura de validación entre los 29 Servicios de Salud genera un escenario de variación exógena. Comparar desenlaces clínicos entre SS con alta vs baja adopción, controlando por case-mix, permitiría estimar el efecto causal de la validación mediante un diseño de diferencias en diferencias o variables instrumentales.
 - **Landmark analysis:** Medir supervivencia desde la fecha de clasificación algorítmica (no desde el ingreso a LE) para controlar formalmente el immortal time bias y confirmar si las diferencias persisten.
 - **Análisis restringido por cohorte temporal:** Comparar VP vs FN dentro del mismo mes de ingreso para controlar el sesgo temporal y asegurar que los grupos sean comparables en tiempo de seguimiento.
 - **Propensity score matching:** Emparejar casos por edad, sexo, Servicio de Salud, tipo de prestación y tiempo en lista de espera para aislar el efecto de la detección algorítmica de los confounders observables.

Estos diseños permitirán responder una pregunta central para la política pública: **en qué medida la actuación oportuna del sistema se asocia causalmente a mejores pronósticos.**

Sobre la interpretación de la velocidad de atención:

La evidencia presentada en este reporte sugiere que la reducción de tiempos de atención, por sí sola, no garantiza mejores resultados clínicos. Una proporción importante de las atenciones más rápidas corresponde a pacientes con mayor gravedad — posiblemente en etapas más avanzadas de enfermedad o con mayor complejidad clínica —, lo que se traduce en una mayor mortalidad precoz post-cirugía (ver [SD6](#)). La velocidad de resolución no puede interpretarse exclusivamente como un indicador de eficiencia, sino también como un reflejo de urgencia clínica, lo que introduce una limitación relevante en la lectura de los indicadores tradicionales de oportunidad.